IDENTIFYING DECEPTION USING NOVEL TECHNOLOGY-BASED APPROACHES TO UNCOVER CONCEALED INFORMATION

by

Jeffrey Gainer Proudfoot

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT INFORMATION SYSTEMS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2014

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Jeffrey Gainer Proudfoot, titled *Identifying Deception Using Novel Technology-Based Approaches to Uncover Concealed Information* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: April 9, 2014
Jay F. Nunamaker, Jr.

____ Date: April 9, 2014
Judee K. Burgoon

____ Date: April 9, 2014
Joseph S. Valacich

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: April 9, 2014
Dissertation Director: Jay F. Nunamaker, Jr.

____ Date: April 9, 2014
Dissertation Director: Judee K. Burgoon

STATEMENT BY AUTHOR

SIGNED: ______________________________

Jeffrey Gainer Proudfoot

ACKNOWLEDGEMENTS

The Department of Homeland Security's National Center for Border Security and Immigration (BORDERS) provided support and funding for portions of this dissertation. Additionally, funding from the Center for Identification Technology Research (CITeR), a National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC), supported portions of this dissertation.

I acknowledge the exceptional efforts and support extended to me by the members of my dissertation committee: Jay F. Nunamaker Jr., Judee K. Burgoon and Joseph S. Valacich. I am grateful for your unwavering guidance, dedication, and assistance. I also thank Susan A. Brown for her help, guidance, and support.

I also want to acknowledge Dr. Randy Boyle, the faculty member, collaborator, and friend responsible for my decision to pursue a career in academia. I am indebted to you for your mentorship.

Additionally, I want to recognize Aaron Elkins, Nathan Twyman, and Ryan Schuetzler for facilitating in collecting and processing the data analyzed in this dissertation. Finally, I acknowledge my fellow doctoral students in the Center for the Management of Information. It has been a pleasure collaborating with many of you during the last four years. I look forward to more opportunities for collaboration in the future.

DEDICATION

I dedicate this dissertation to my wife, Leah. Without her encouragement, sacrifice, patience, and love, this journey would never have started or finished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Concealing information, one of the many forms of deception, is a pervasive phenomenon as it is present in virtually every facet of interpersonal communication. In some cases, information concealment can have profound implications (e.g., insider threats in organizations, security screening at the border, and criminal interviews). New technologies are under development to aid in identifying concealed information, however, additional research is needed in three key areas to increase the feasibility of using these technologies in real-world credibility assessment contexts. First, research is needed to investigate the accuracy of new credibility assessment technologies relative to existing deception-detection systems. Demonstrating that new technologies meet or exceed detection accuracies of existing systems (e.g., the polygraph) is critical. Second, research is needed to determine if a targetless Concealed Information Test (CIT) is feasible. Existing CIT research supports the presence of main effect differences between persons concealing information and the control group. These behaviors may permit the detection of concealed information without the use of customized sets of stimuli. Eliminating the need to create customized sets of stimuli for each examinee would drastically increase the ease with which an automated system can be used to conduct a CIT. Finally, research is needed to illuminate various elements of the human-computer interaction that occurs during automated credibility assessments. This is a new domain of human-computer interaction as system users in this context are not instigating the interaction, and in many cases, they may be seeking to limit the effectiveness of the system. Before novel systems designed to conduct credibility assessments can be adopted, further research is needed to illuminate how users perceive, respond to, and strategically manage their behaviors when interacting with systems of this nature. This dissertation contains the results of a research

program designed to address each of these areas. First, an experiment was designed to investigate the accuracy rates of two promising noncontact measures of concealed information (oculometrics and vocalics) relative to electrodermal activity (EDA)[1]. Second, an experiment was designed to evaluate the feasibility of using a targetless CIT to elicit main effect differences between concealers and the control group to identify concealed information. And third, a thorough analysis of examinees' general perceptions, self-reported stress and arousal, perceived effort and performance, and use of countermeasures within the context of an automated credibility assessment interview was conducted. This research effort has yielded the following findings. First, eye tracking and vocalics can be used to identify significant differences in the behaviors and physiology of examinees concealing information, however, the accuracy with which truth tellers and information concealers can be classified remains impractical for an applied setting. Second, there are main effect differences between persons concealing information and telling the truth, however, the use of countermeasures may limit the accuracy with which concealers can be identified. Finally, the presence of concealed information and the use of crime-relevant questions alter how examinees perceive and react to a system designed to identify concealed information. The limitations of this research, as well as directions for future research, are discussed.

---

[1] EDA is ubiquitously used to conduct both standard polygraph interviews as well as CITs. In the context of conducting a CIT, EDA is often the only measure collected or interpreted by the examiner.

1 INTRODUCTION

Concealing information, one of the many forms of deception, is a pervasive phenomenon as it is present in virtually every facet of interpersonal communication. While innocuous in many cases, some instances of information concealment can have profound implications.

For example, insider threats, defined as “individuals who were, or previously had been, authorized to use the information systems they eventually employed to perpetrate harm” (Randazzo, Keeney, Kowalski, Cappelli, & Moore, 2004, p. 1) can adversely affect organizations by exploiting their trusted status, resulting in substantial financial losses, and in some cases, irreparable damage to the organization. Some organizations utilize pre-employment credibility assessment interviews in addition to other measures in an attempt to reduce insider threats; however, insider threat attacks remain “one of the most dangerous threats organizations face today” (Baracaldo & Joshi, 2012, p. 167). Novel technology-based methods of identifying concealed information[2] have the potential to identify and mitigate the prevalence and impact of insider threat incidents.

Border security and passenger credibility assessments are operations that could also benefit from novel technology-based methods of identifying concealed information. Millions of travelers cross international borders annually (Derrick, Elkins, Burgoon, Nunamaker Jr., & Zeng, 2010). One responsibility of officers tasked with conducting credibility assessment interviews is to identify travelers concealing information about their identity, immigration or visa status, or their intentions upon gaining entry to a foreign country. However, empirical evidence suggests that humans are unreliable detectors of deceit (Bond & DePaulo, 2006), regardless of professional

---

[2] Methods not yet validated or subjected to thorough empirical investigation.

status (Burgoon, Buller, Ebesu, & Rockwell, 1994; Hartwig, Granhag, Strömwall, & Vrij, 2004). The sheer volume of individuals crossing the border, coupled with the difficulty of identifying concealed information, presents a formidable challenge for security personnel tasked with facilitating commerce while ensuring national security.

Criminal interviews could also greatly benefit from novel technology-based methods of identifying concealed information. Two of the most commonly-used interviewing techniques for conducting veracity assessments are the Comparison Question Test (CQT)[3] and the Behavioral Analysis Interview (BAI) (Vrij, 2008). Despite wide use for both event-related and screening applications, the CQT is frequently criticized for its lack of scientific grounding and questionable rates of accuracy (Aftergood, 2000; Honts, Raskin, & Kircher, 2002; Meijer & Verschuere, 2010; 2003; Saxe, Dougherty, & Cross, 1985). Similarly, while some research supports the validity of the BAI, other work investigating its validity has yielded less-compelling results (Blair & Kooi, 2004; Ekman, O'Sullivan, & Frank, 1999; Horvath, Blair, & Buckley, 2008; Horvath, Jayne, & Buckley, 1994; Vrij, Fisher, Mann, & Leal, 2006).

In response to these gaps and countless others, a number of topics have been investigated to improve the accuracy with which humans and technologies can identify deception and concealed information. Some key aspects of this research effort include the estimation of deception detection accuracy rates and the development of automated tools to facilitate in identifying deception and concealed information (Bond & DePaulo, 2006; Boyle, Clements, & Proudfoot, 2012; Buller & Burgoon, 1996; Burgoon, Blair, & Strom, 2008; Burgoon, Fang, & Twitchell, 2010; Burgoon & Nunamaker Jr., 2004; Burgoon & Qin, 2006; Burgoon, 1996; Burgoon, Derrick, et al.,

---

[3] The CQT is the standard interviewing technique used to conduct polygraph examinations.

2008; DePaulo et al., 2003; Derrick et al., 2010; Hartwig & Bond Jr., 2011; Nunamaker Jr., Burgoon, et al., 2012; Nunamaker Jr., Derrick, Elkins, Burgoon, & Patton, 2011).

While a rich corpus of deception research exists, no "silver bullet" cue or technique has been identified to validly and reliably identify deception and concealed information across contexts, cultures, communication modalities, and communicators. Furthermore, the polygraph, the technology most commonly used to identify physiological indicants of deception and concealed information, is highly limited as it is a one-investigator to one-subject methodology. The current polygraph platform is not easily scalable to a large number of investigations as it requires an extensive time commitment, intricate sensor calibration, customized interview questions, multiple administrations per examinee, and chart scoring conducted by a skilled examiner. There exists a tremendous opportunity to develop a more accurate, rapid, and highly-scalable information system for conducting credibility assessments; a platform that harnesses the power of automated interviewing and data collection, fusion of data from disparate noncontact sensors[4], and real-time unbiased data analysis.

Researchers in the National Center for Border Security and Immigration (BORDERS), a Department of Homeland Security Center of Excellence, are utilizing a design science methodology (Hevner & Chatterjee, 2011; Hevner, March, Park, & Ram, 2004; Nunamaker Jr., Twyman, & Giboney, 2013) to develop such a system (Derrick, 2011; Nunamaker Jr. et al., 2011; Twyman, 2012). This system has been named the Automated Virtual Agent for Truth Assessment in Real time, or AVATAR (Nunamaker Jr., Elkins, Twyman, & Derrick, 2012). A key area of research in the BORDERS Center is identifying the most effective interviewing methods that can be

[4] Behaviors that do not require contact with the examinee to collect data; e.g., oculometrics and vocalics.

used by the AVATAR to conduct credibility assessments. The Concealed Information Test (CIT)[5] is an interviewing technique extensively researched due to its simplicity, its inherent objectivity, and an extensive foundation of literature pointing to its validity[6] (Twyman, 2012; Verschuere, Ben-Shakhar, & Meijer, 2011).

Electrodermal activity (EDA) is the core autonomic measure used to collect physiological data during CITs (Krapohl, McCloughlan, & Senter, 2009). EDA requires direct contact with the examinee as electrodes must be placed on the fingers or palms of the examinee, thus, it is not feasible for an automated credibility assessment context. Due to the limitations of collecting EDA, noncontact sensors have become a recent area of investigation as they would render an automated CIT feasible for use (Elkins, Derrick, & Gariup, 2012; Elkins, 2011; Nunamaker Jr., Elkins, et al., 2012; Proudfoot, Twyman, & Burgoon, 2012, 2013). Vocalic and oculometric sensors, sensors that capture various measurements of speech and eye movements, respectively, have demonstrated promising results when used to conduct an automated CIT (Derrick, Moffit, & Nunamaker Jr., 2011; Elkins et al., 2012; Proudfoot et al., 2013; Twyman, 2012). A preliminary investigation of the robustness of noncontact sensors to countermeasures also reported promising results (Twyman, Schuetzler, Proudfoot, & Elkins, 2013). What remains to be investigated is the accuracy and potential utility of using noncontact sensors relative to EDA when employing the CIT. The results of a laboratory experiment designed to investigate the relative accuracy rates of oculometrics, vocalics, and EDA will be discussed in Chapter 4 of this dissertation.

---

[5] Also referred to as the Guilty Knowledge Test (GKT), the peak of tension test, or the stimulation test (Ansley, 1992)

[6] A more complete overview of relevant CIT literature is discussed in Chapter 2 of this dissertation.

An additional challenge hindering the use of the CIT in an automated credibility assessment paradigm is the need to identify and implement customized target items for each examinee. Target items are stimuli that should elicit physiological differences between those concealing information and those not concealing information. However, the findings of two CIT studies suggest that individuals concealing information may exhibit main effect physiological and behavioral differences relative to the control group (Lubow & Fein, 1996; Proudfoot et al., 2013). These findings suggest that a CIT may not require the use of target items to identify concealed information. While main effect differences have been identified in these two prior studies, CIT researchers have yet to directly investigate the feasibility of using a targetless CIT. The results of a laboratory experiment designed to measure the presence of these main effect differences during a targetless CIT, as well as the duration of these differences, will be discussed in Chapter 6 of this dissertation.

Finally, a key aspect of credibility assessment interviews is the human-computer interaction that occurs. In most contexts, humans trigger or instigate their own interaction with a system to benefit from any number of capabilities that it can provide. Automated credibility assessments are a context in which humans are likely not seeking the interaction, especially if they have information to conceal. Chapter 7 of this dissertation will present the findings of research attempting to illuminate differences in the perceptions, emotions, and frequency with which examinees attempt to use countermeasures contingent on the presence of (1) concealed information and (2) the presence of concealed information and interview questions that are crime-relevant.

In short, this research can make a significant contribution to the development and adoption of automated credibility assessment information systems leveraging the CIT interviewing method as it will provide insight on: (1) the current dependence on contact sensors to conduct a CIT, (2) the need to identify and incorporate target items into a CIT, and (3) the human-computer-interaction that occurs during automated credibility assessment interviews.

2 THE CONCEALED INFORMATION TEST (CIT)

The CIT is an empirically-validated interviewing technique (Verschuere, Ben-Shakhar, et al., 2011) that has potential for use in automated credibility assessment interviews. This section introduces the CIT interviewing technique, discusses the orienting response as the underlying mechanism measured during a CIT, describes the design and procedures used to conduct a traditional CIT, reviews traditional sensors used to collect physiological data during a CIT, discusses a common method used to score EDA data collected during a CIT, and concludes with a discussion outlining the advantages and disadvantages of the CIT.

2.1 Introduction

Methods of interpreting psychophysiological measurements to identify deception can be categorized into two main approaches: identifying deception and identifying recognition (Krapohl et al., 2009). Krapohl and colleagues (2009) created a taxonomy of the major psychophysiological deception detection (PDD) methods. A variety of approaches constitute the taxonomy, including the prevalently used deception-based CQT (Vrij, 2008)) and the recognition-based CIT. A re-creation of this taxonomy is provided in Figure 1.

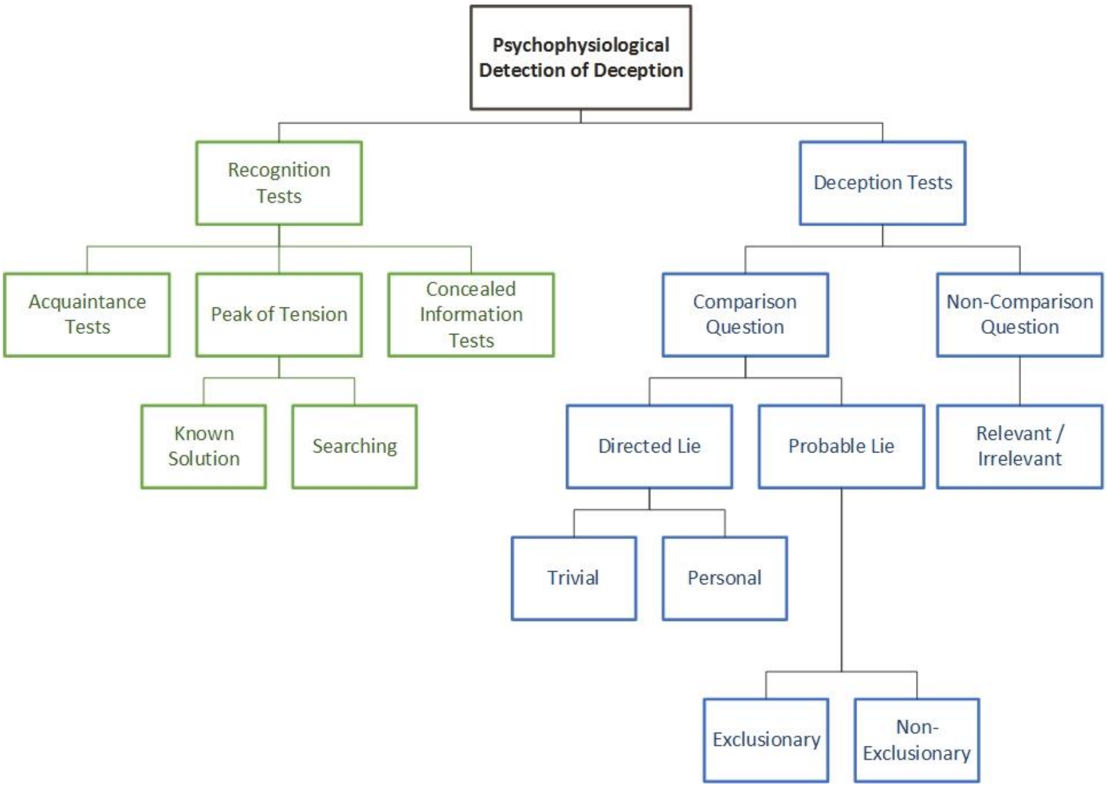


Figure 1 - Taxonomy of PDD Methods (Krapohl et al., 2009)

The CIT was formally developed by David Lykken (1959) based on interviewing concepts proposed decades earlier (Crane, 1915; Krapohl & Velez, 2001; Münsterberg, 1908; Twyman, 2012). The CIT was formulated with the intention of measuring variations in an examinee's physiology attributable to the recognition of familiar or personally significant stimuli. It can be used by criminal examiners as a lone technique or as a supplementary interview to augment[7] more prevalently used interviewing techniques (e.g., the Comparison Question Test (CQT)) (Krapohl et al., 2009)[8].

[7] If used in conjunction with the CQT, it is common practice to administer the CIT first (Krapohl, 2010).

[8] While the primary application of the CIT is conducting criminal interviews, it can also be used in a variety of clinical applications, including detecting memory traces and identifying malingering (Allen, 2011).

2.2 The Orienting Response / Reflex

For the CIT to possess discriminatory power, a guilty examinee must "…show some involuntary physiological response (e.g., Galvanic Skin Response[9] (GSR)) to stimuli related to remembered details of a crime" (Lykken, 1959, p. 385). The physiological phenomenon serving as the basis for the CIT is the *orienting response / reflex* (Krapohl et al., 2009; O'Gorman, 1979; Siddle, Kyriacou, Heron, & Mathews, 1979; Sokolov, 1963b, 1966; Verschuere, Crombez, De Clercq, & Koster, 2004). It is an involuntary response that is manifested when attention fixates to a novel or personally significant stimulus (Sokolov, 1963a). Pavlov (1947, p. 27) defined the orienting reflex as "…this condition which brings about the immediate response in man and animals to the slightest changes in the world around them, so that they immediately orientate their appropriate receptor-organs in accordance with the perceptible quality in the agent bringing about the change, making full investigation of it."

A wealth of scientific research has been conducted concerning the types of physiological effects that occur in conjunction with the orienting response (Gamer, Bauermann, Stoeter, & Vossel, 2007; Gamer, Verschuere, Crombez, & Vossel, 2008; Lykken, 1974) as well as the measurements that can be used to identify it. Traditional measures used for the CIT replicate those used for the CQT, and include: EDA, phasic heart rate (pHR), respiration line length (RLL), and finger pulse waveform length (FPWL) (Ambach, Bursch, Stark, & Vaitl, 2010; M. T. Bradley & Ainsworth, 1984; M. T. Bradley & Janisse, 1981; Elaad & Ben-Shakhar, 1989; Gamer et al., 2007, 2008; Lykken, 1959; Podlesny & Raskin, 1978; Suzuki, Nakayama, & Furedy, 2004;

[9] Galvanic Skin Response (GSR) is now commonly referred to as Electrodermal Activity (EDA).

Verschuere et al., 2004). Matsuda and colleagues (2012) propose the use of new measures, including: reaction times, facial responses, electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and event-related potentials (ERP) as a means of augmenting and improving the aggregate accuracy rates of existing measures[10] (Allen, Iacono, & Danielson, 1992; Allen & Iacono, 1997; Ben-Shakhar & Meijer, 2012; Ekman, 2001; Nose, Murai, & Taira, 2009; Verschuere & De Houwer, 2011).

2.3 The CIT Interview

The CIT interview is fairly standardized[11] and easy to administer (Krapohl et al., 2009; Krapohl & Velez, 2001). The general approach is to present an examinee with stimuli[12] whilst measuring their physiology. Stimuli are comprised of test items from two categories: *target items* and *nontarget items*[13].

2.3.1 Target and Nontarget Items

Target items should elicit an orienting response in examinees attempting to conceal information. Target items are interspersed with equally-plausible nontarget items (Patrick, 2011). Nontarget items are important because they serve as a baseline to make comparisons between the physiological responses that occur when target

---

[10] Neuroimaging techniques have been an area of recent investigation, and while they may not be conducive to criminal interviewing or automated screening, understanding deception and the concealment of information using these technologies can yield valuable insights. Neuroscience researchers using electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) to identify deception and concealed information stress the importance of better understanding the cognitive processes associated with deception and recognition as the most prevalently used physiological indicators (e.g., EDA) are simply peripheral manifestations of phenomena that are originating in the brain (Kozel, Padgett, & George, 2004; Langleben et al., 2005; Verschuere, Ben-Shakhar, & Meijer, 2011).

[11] The standard CIT can be modified depending on a number of factors, including the context, the sensors used, time constraints associated with a specific testing scenario, etc. (Krapohl, 2010).

[12] Stimuli are most often presented audibly but they can also be presented visually (Krapohl, McCloughlan, & Senter, 2009).

[13] Target items and nontarget items can also be referred to as key and control items or critical and noncritical items, respectively.

items and nontarget items are presented. On this topic Patrick (2011, p. 3) stated that "…to ensure that the observed ("dependent") effect is attributable to the experimental ("independent") manipulation, one must establish a comparison condition that mirrors the experimental condition in all respects aside from the manipulation of interest. With this principle in mind, the CIT was composed of items in multiple choice format, with alternative choices for each item (one of them crime-relevant, the other extraneous) formulated to appear equally plausible to an innocent examinee."

2.3.2 Foil Construction

One target item is placed randomly in a group of nontarget items; this group is referred to as a *foil*[14]. A foil is comprised of several test items, although the use of one target and five nontarget items has been advocated[15] (Krapohl et al., 2009). A CIT is comprised of several foils due to the inverse relationship between the number of foils and the statistical probability that indications of concealed information could have occurred due to chance. It has been recommended to include a sufficient number of foils in a CIT to yield less than a 1% chance that an examinee has been wrongly classified as concealing information (Krapohl, 2010). Target and nontarget items can be displayed in intervals of 12-15 seconds, but additional time (e.g., 25 seconds) can be allotted per stimulus depending on the measures collected (Krapohl & Velez, 2001; Krapohl, 2010).

CIT foils are comprised of target items dealing with different elements of a crime or the event of interest in question (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003). For example, in the case of an investigation concerning the robbery

---

[14] Foils are also referred to as blocks (Matsuda, Nittono, & Allen, 2012).

[15] Structuring a foil with 1 target item and 5 nontarget items results in an easier calculation of error estimates as one nontarget item is not considered for analysis, resulting in a foil of only 5 items for consideration (Krapohl, 2010).

of a car, foils can be created to test the examinee on his or her knowledge of the color, make, model, and interior color of the car. In a traditional CIT, in which stimuli are presented audibly, the examiner would state each foil item, and the examinee would then verbally repeat each item after the examiner. In a visual CIT, the stimulus is presented on a screen and the examinee responds by stating either 'yes' or 'no' concerning their familiarity with that information[16].

A more formal example of a CIT foil is as follows: "If you were the perpetrator of this crime, you will know where the criminal gained access to the house. Where was the place of entry? Was it the: (1) front entrance? (2) kitchen door? (3) bathroom window? (4) balcony? (5) room on the second floor?" (Nakayama, 2002, p. 50). Best practices in concealed information testing dictate that data for the first item in each foil should be discarded and excluded from analysis; this is due to the commonality of orienting responses occurring in response to the first item in each foil, even if the examinee has no recognition or prior knowledge of that stimulus (Lykken, 1960). The target item is randomly interspersed with nontarget items in the foil, but it is never the first item in the foil (Lykken, 1960).

2.3.3 CIT Administration

While each stimulus is presented to the examinee, and he or she is responding, physiological responses are measured using a variety of sensors; these data are recorded, and upon concluding the CIT, the examiner then interprets these responses and looks for anomalous activity associated with the presentation of target items (relative to the responses that occur in conjunction with nontarget items). On the

[16] There may be an inherent risk in requiring the examinee to respond by saying only "yes" or "no". Krapohl (2010) points out that this method of responding may cause dissociation between the examinee and the target item, resulting in muted physiological variations and a possible false-negative classification.

occasion that a CIT examination has to be redone[17], target and nontarget items should be randomized within their foils to reduce habituation or any form of anticipatory effect (Krapohl, 2010).

2.3.4 Pretest Practices

For optimal results, a brief interview is administered prior to conducting a CIT (Krapohl et al., 2009). During this interview, the examiner should explain the processes associated with conducting a CIT and answer any questions that the examinee has concerning the test. The examinee must then complete relevant paperwork, including a form designed to elicit information about the crime or incident of interest of which the examinee may have knowledge[18]. Target and nontarget items can be reviewed with the examinee prior to the test to ensure that he or she understands all stimuli, can pronounce all stimuli, and cannot distinguish the target item from the accompanying nontarget items in each foil (Krapohl et al., 2009). A preliminary identification of crime-relevant information is a crucial step in the CIT process, as withholding information about the crime or incident of interest at this time may ultimately implicate an innocent person later on as being a potential suspect. Information gleaned during the interview is then reviewed by the examiner and verbally committed to by the examinee.

2.3.5 The Searching / Probing CIT: An Alternative Approach

The traditional CIT is grounded in the assumption that criminal investigators have sufficient knowledge of the crime to identify salient target items. Variations in examinees' physiological responses to target and nontarget items are then measured

---

[17] This can occur for any number of reasons, including (1) exterior distractions or noise that could trigger erroneous orienting responses or (2) detecting the use of countermeasures.

[18] Due to the propensity for media outlets to release information about criminal events, it is very common for innocent persons to possess detailed knowledge of these events, thus hampering the utility of traditional CIT use in the field (M. T. Bradley, Barefoot, & Arsenault, 2011; Matsuda et al., 2012).

and interpreted during the test to identify concealed knowledge. This approach is referred to as the Known Solutions CIT; it is the most prevalently used method of conducting a CIT in Japan (Osugi, 2011).

If criminal investigators do not have sufficient information about a crime to conduct a Known Solutions CIT, a technique referred to as the Searching CIT[19] can be utilized (M. T. Bradley, Barefoot, & Arsenault, 2011; MacLaren, 2001; Matsuda et al., 2012; Osugi, 2011; Patrick, 2011). For this alternative approach, examiners must identify gaps in their knowledge of a crime that require additional information, then formulate foils of feasible alternatives related to those gaps that can be used to conduct a Searching CIT. For example, if investigators have determined that a suspect is distributing narcotics, but investigators are uncertain as to the location of his or her drug stash, they could conduct a Searching CIT to gain more information concerning the location of the narcotics. Investigators would first identify the most feasible locations where the narcotics could be located, then ask the suspect each of the alternatives. For example, “Are the narcotics hidden in your (a) house, (b) car, (c) cabin, (d) with family elsewhere, or (e) buried?” If the physiology of the examinee indicates recognition to any of these responses, investigators can create a more specific Searching CIT to find out more information (e.g., the room in the house in which the narcotics are located). The Searching CIT adds an additional level of complexity for analysis as the examiner must not only identify recognition, but identify the stimulus that triggered that recognition (Matsuda et al., 2012).

MacLaren (2001) described an anecdote in which the United States Customs Service effectively used a Searching CIT in 1988 resulting in a large seizure of narcotics. Due to the rarity of CIT use in the United States, the anecdote in its entirety

[19] The Searching CIT is also referred to as the Probing CIT (Osugi, 2011).

is worth including; it is as follows: "…after 2 days of unsuccessfully searching a ship believed to be transporting illegal narcotics, they [U.S. Customs] interrogated crew members about the location of the drugs with the aid of a GKT. The location of a secret compartment within the walls of the ship was identified using these means, and 4,800 pounds of cocaine were confiscated. The drugs were valued at $350 million, and it was the largest drug seizure in American history at the time…" (MacLaren, 2001, p. 675). This anecdote demonstrates the utility that a Searching CIT can provide if investigators are limited in the information that is available to them.

2.4 Data Collection

EDA is the measure predominantly used for data collection during a CIT; however, other PDD measures traditionally used for standard polygraph interviews can also be used (Krapohl et al., 2009; Krapohl & Velez, 2001). Alternative PDD measures include phasic heart rate (pHR), respiration line length (RLL), and finger pulse waveform length (FPWL) (Gamer, 2011). Research supporting the use of these additional sensors suggests that some show indications of promise, but require additional research to solidify their utility, while others have received meager or no support for use in a CIT context (Krapohl, 2010). A discussion of each of these sensors is presented in the following sections.

2.4.1 EDA

The use of EDA to measure changes in phasic skin conductance has long been associated with identifying autonomic physiological cues of deception. Traditionally, research has investigated skin conductance as a means of PDD within the context of the standard CQT-based polygraph examination. Its utility in that context was recognized and was then adapted for use in the CIT (Lykken, 1959).

In a criminal interviewing paradigm, changes in skin conductance are measured by placing electrodes on the fingers or palms of the examinee; however, in other contexts, EDA can also be accurately measured by placing electrodes on the forehead or feet (Boucsein, 2012). Electrodes placed on the skin form a circuit as one electrode emits a current while the second electrode measures variations in that current. Current variations are largely attributable to changes in the level of sweat present on the stratum corneum of the skin. To improve electrode measurements, an electrolyte solution is often applied in the form of a paste or gel (Grey & Smith, 1984). To identify concealed information, an examiner will look for phasic responses, represented by larger trough-to-peak responses, that occur in conjunction with the presentation of target items (Christie, 1981). This occurs due to "a coactivation of the sympathetic and the vagal branch of the autonomic nervous system and it is at least in part related to the orienting response" (Gamer, 2011, p. 27). Phasic responses are inherently temporary and normalize after a short period of time until returning to tonic levels.

The initial study evaluating the use of EDA to conduct a CIT utilized a number of mock crime scenarios. Some participants committed multiple crimes, others committed only a single crime, and a final group committed no crime at all. All participants were randomly presented with target items and equally plausible control items. Based on the dependent measure EDA, members of the control group were classified with 100% accuracy, while concealers were classified with 88%[20] accuracy (Lykken, 1959). A follow-up study conducted by Lykken confirmed the robustness of the GKT approach to countermeasures (Lykken, 1960).

---

[20] It is interesting to note that the seminal empirical evaluation of the CIT resulted in an 88% rate of accuracy in identifying persons concealing information. This same accuracy rate was reported decades later in a meta-analysis of CIT accuracy rates.

While mock crime scenarios are prevalent in deception and guilty-knowledge research, Gamer (2011) identified several experimental paradigms in which EDA has been used as a dependent measure to test a variety of stimuli, including: information available in field settings (Davidson, 1968; Elaad, Ginton, & Jungman, 1992; Elaad, 1990), cards (Ben-Shakhar, 1994), numbers (Horvath, 1978, 1979), code words (Waid, Orne, Cook, & Orne, 1978; Waid & Orne, 1980), and autobiographical information (Ben-Shakhar, Lieblich, & Kugelmass, 1975; Elaad, 1994).

2.4.2 Other PDD Sensors

In addition to EDA, PDD measures traditionally used for CQT-based polygraph examinations can be used for data collection during a CIT (Ambach et al., 2010). These measures include phasic heart rate (pHR), respiration line length (RLL), and finger pulse waveform length (FPWL).

2.4.2.1 Phasic Heart Rate (pHR)

Variations in pHR during a CIT mimic those found present in EDA as an examinee should exhibit phasic differences in response to target stimuli (Abercrombie, Chambers, Greischar, & Monticelli, 2008). In the case of pHR, phasic changes are represented by a deceleration[21] in heart rate relative to tonic pHR (Graham, 1979; Matsuda et al., 2012). pHR is traditionally measured using a blood pressure cuff, but an ECG or laser doppler vibrometer (LDV) can also be used (Derrick et al., 2010; Ryan Jr., Pavlidis, Rohrbaugh, Marchak, & Kozel, 2003). Empirical research suggests that the classification of concealed knowledge through the recognition of heart rate decelerations is a valid approach (Adachi & Suzuki, 1991; M. T. Bradley & Ainsworth, 1984; M. T. Bradley & Janisse, 1981; Podlesny &

[21] As opposed to a phasic increase in EDA.

Raskin, 1978). However, the effect sizes are small (Krapohl et al., 2009) and EDA remains the more robust and thus widely-used measure (Verschuere et al., 2004).

2.4.2.2 Respiration Line Length (RLL)

RLL is used to interpret rate and depth of breathing; it has been investigated as a supplementary measure to use in conjunction with EDA during a CIT (Ambach et al., 2010). It is used to identify concealed information through the interpretation of respiratory amplitudes, cycle times, and suppression (Elaad & Ben-Shakhar, 2006; Matsuda et al., 2012). RLL is measured by placing pneumographs, traditionally used for CQT-based polygraph examinations, on the upper and lower abdomen of the examinee. When interpreting RLL data, examiners look for respiration volume decreases as a sign of concealed knowledge (Gamer et al., 2008; Kurohara, Kensuke, Hiromi, & Akio, 2001). Despite the possible utility of using this PDD measure, it should be used with caution as respiratory activity can be muddled with "noise" from (a) the examinee vocally responding to each question, (b) intentional manipulations in respiratory activity associated with countermeasures[22], or (c) a combination of the two (Krapohl et al., 2009).

2.4.2.3 Finger Pulse Waveform Length (FPWL)

FPWL is a relatively new measure used for PDD. It is measured using a plethysmograph sensor placed on the middle finger on the non-dominant hand (Ambach et al., 2010). FPWL data are a representation of variations in pulse activity associated with changes in vasoconstriction[23]. Concealed knowledge is indicated by a decrease in pulse rate and pulse amplitude (Elaad & Ben-Shakhar, 2006; Matsuda et

[22] It has also been suggested that RLL be used as a means of identifying countermeasures in conjunction with EDA data, thus, it may be worthwhile to collect RLL as a counter-countermeasure (Krapohl, 2010).

[23] Narrowing of blood cells.

al., 2012). The validity of FPWL has been investigated in a number of empirical studies (Elaad & Ben-Shakhar, 2006; Podlesny, Raskin, & Barland, 1976; Podlesny & Raskin, 1978; Vandenbosch, Verschuere, Crombez, & Clerq, 2009); the findings suggest small but significant effects (Krapohl et al., 2009).

2.5 Scoring

The Lykken scoring method (Lykken, 1959) is the most widely-used and heavily-researched approach for interpreting EDA[24] data collected during a CIT. This is due to its simplicity and practicality for use in the field. For each foil of stimuli, the largest trough-to-peek amplitude and the second-largest trough-to-peak amplitude are identified and awarded two points and one point, respectively. The remaining amplitudes in each group receive no points (Lykken, 1959). The points for each foil are then aggregated and compared to a predetermined cutoff point. In the case of a CIT comprised of five foils, a point total greater than or equal to five results in a classification of Recognition Indicated (RI)[25] (Krapohl & Velez, 2001). In this example, if the point total is less than five, a classification of No Recognition Indicated is given (NRI). If complications during the examination (e.g., movement, outside distractions, coughing, use of countermeasures, etc.) occur, a classification of No Opinion (NO) can be rendered.

Refer to Figure 2 for an example of a foil of EDA data scored using the Lykken scoring method (1959). The green vertical shading signifies the first stimulus presented in the foil. The three gray vertical bars of shading represent the second, third, and fourth stimuli presented in the foil; these are nontarget items. Finally, the

[24] EDA is the most prevalently collected PDD measure in a CIT, thus, it is the only scoring method discussed in this section.

[25] The CIT is generally discussed from the viewpoint that it is a means of identifying the perpetrator of a crime; however, it can also be used to identify witnesses, accomplices, and other persons who can provide valuable information to authorities (Matsuda et al., 2012).

red vertical shading indicates that the fifth and final stimulus is a target item. It is clear in this example that the largest trough-to-peak amplitude occurred in conjunction with the target item, thus, it was awarded two points. The second-largest trough-to-peak amplitude occurred while the second item in the foil was displayed; accordingly, it was awarded a single point.

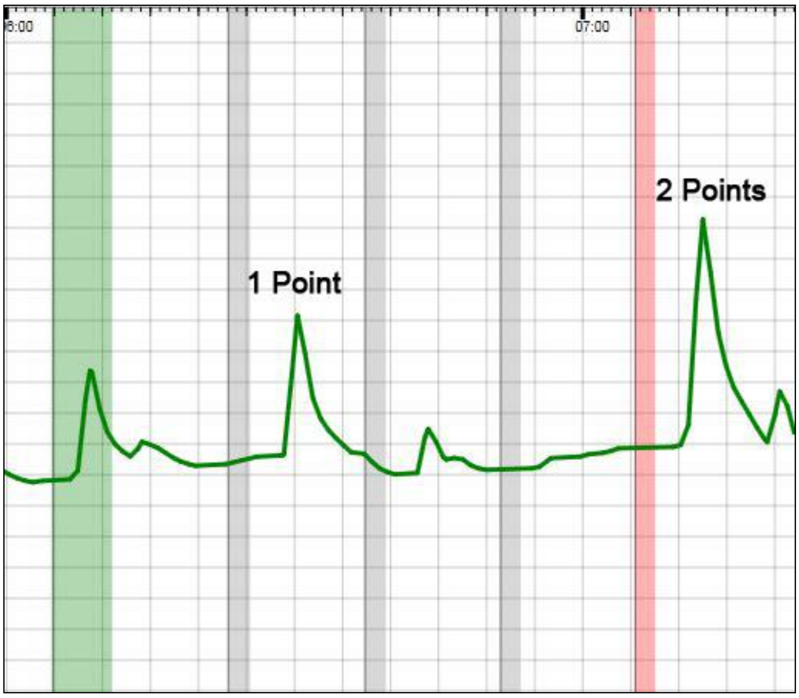


Figure 2 - Example of Lykken (1959) Scoring Method

A criticism of the Lykken scoring method is that it does not utilize all of the information present in the chart (Matsuda et al., 2012; Meijer, Verschuere, & Ben-Shakhar, 2011). For example, in Figure 2, it is clear that the fifth stimulus in the foil has the largest trough-to-peak amplitude; however, the Lykken method does not consider the relative difference between the largest amplitude and the four accompanying stimuli in that foil. Thus, an amplitude only slightly larger than the other amplitudes in the foil is the same as an amplitude five times larger than any other amplitude in the foil. This is valuable information that is lost using this method;

however, amplitude differences can add value when using a more robust, quantitatively-based scoring method[26].

2.6 Evaluating the CIT

A wealth of empirical research supports the CIT as a valid interviewing technique (Ben-Shakhar & Elaad, 2003; Elaad, 1990; Gamer, 2011; Krapohl et al., 2009; Krapohl & Velez, 2001; MacLaren, 2001; Yokoi, Okazaki, Kiriu, Kuramochi, & Ohama, 2001). A report investigating the validity of polygraph tests found the CIT to have an estimated accuracy rate of 88% (Krapohl, 2010). Despite its empirical support and relatively high rate of accuracy, it is scarcely used in the field for criminal investigations (Suzuki et al., 2004). Few countries observe widespread use of this interviewing technique (Elaad, 1990; Hira & Furumitsu, 2002; Nakayama, 2002; Yokoi et al., 2001), including Israel and Japan.

The most commonly utilized interviewing techniques are the CQT and the BAI (Vrij, 2008). Despite wide use for both event-related and screening applications[27], the CQT is frequently criticized for its lack of scientific grounding and questionable rates of accuracy (Aftergood, 2000; Ben-Shakhar, Bar-Hillel, & Kremnitzer, 2002; Fiedler, Schmid, & Stahl, 2002; Honts et al., 2002; Iacono & Lykken, 2002; Iacono, 2000; Lykken, 1991, 1998; Meijer & Verschuere, 2010; 2003; Saxe & Ben-Shakhar, 1999; Saxe et al., 1985; Vrij, 2008)[28]. Podlesny (1993) conducted a review of FBI investigations in which the CQT was administered to

---

[26] Refer to work by Matsuda and colleagues (2012) for a review of several alternative scoring methods/approaches.

[27] An event-related CQT could be conducted as part of a criminal investigation (e.g., a murder trial) whereas a screening CQT could be administered to a pool of applicants completing the interview process for a government agency.

[28] Please refer to these publications for additional insight concerning the design and use of the CQT as the emphasis of this research is the CIT.

determine if the CIT could have been effectively used as a part of each investigation. His findings suggest that roughly 13% of these cases could have benefitted from the administration of a CIT (Krapohl & Velez, 2001; Podlesny, 1993). Findings concerning the validity of the BAI have also been called into question (Blair & Kooi, 2004; Ekman et al., 1999; Horvath et al., 2008, 1994; Vrij et al., 2006)[29].

The graphic represented in Figure 3 was created by Twyman (2012) to compare basic characteristics of the CIT, CQT, and BAI, including: time requirements, common criteria for assessment, examiner skill level requirements, and reported usage. It is clear from Figure 3 that the CIT has a number of advantages over the CQT and BAI, including a less demanding time requirement, a high level of validity, and a very low skill-level requirement for the examiner (Twyman, 2012).

Inter-viewing Technique	Time Required to Conduct Interview	Scientific Consensus on Validity	Most Common Criterion for Assessment	Inter-viewer Skill Level Required	Practitioner Usage
CIT	2-15 minutes*	High Validity	Presence of elevated orienting response following onset of relevant stimulus	Very Low	Limited to Japan and some use in Israel (Nakayama, 2002; Vrij, 2008)
CQT	2-4 hours**	Low Validity	Presence of elevated psychophysiological response during relevant question(s)	High	Widespread use in North America, Asia, and Europe (Vrij, 2008)
BAI	15-45 minutes***	Uncertain; Nuanced	Expert analysis of verbal and non-verbal behavior during interview	High	Used in the United States including some business applications; also some international use (John E. Reid & Associates, 2011)

*Exact time is a function of how many questions are used (usually between 3 and 6).
**Estimated from a subjective review of polygraph examiner practitioner promotional material.
***Lower bound estimate based on amount of time required to minimally ask and respond to all BAI questions. Upper bound estimate reflects potential for follow up questions.

Figure 3 - Twyman's (2012) Comparison of CIT, CQT, and BAI

[29] Please refer to these publications for additional insight concerning the design and use of the BAI as the emphasis of this research is the CIT.

In addition to these core attributes, there are a number of advantages and disadvantages of the CIT that should be considered.

2.6.1 Advantages of the CIT

Advantages of the CIT include: the ability to calculate error estimates thus protecting innocent examinees, the reliance of the test on recognition rather than emotions associated with deception, the adaptability of the CIT, and the brevity of the testing process. Each of these advantages is discussed in more detail in the following sections.

2.6.1.1 Calculation of Error Estimates / Protecting the Innocent

The format of the CIT permits the ability to calculate the statistical likelihood that a given examinee received a certain score given a specific number of foils (Krapohl & Velez, 2001). Krapohl (2010) created a table enumerating the probability scores that an examinee is likely unaware of crime-relevant details for CITs constituted of between two and eight foils[30]. A re-creation[31] of this table is provided in Figure 4.

Number of Tests \ Score	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	12.0	4.0												
3	28.0	12.8	3.2	0.8										
4	43.8	24.6	10.1	3.7	0.8	0.2								
5	57.7	37.5	19.7	9.2	3.2	1.0	0.2	0.0						
6	68.9	49.9	30.9	16.9	7.7	3.1	1.0	0.3	0.0	0.0				
7	77.6	60.9	42.3	26.3	14.2	6.8	2.7	1.0	0.3	0.1	0.0	0.0		
8	84.1	70.2	53.1	36.4	22.2	12.2	5.8	2.5	0.9	0.3	0.1	0.0	0.0	0.0

Figure 4 - Probability Table for CITs (Krapohl, 2010)

[30] Assuming that each foil is comprised of six items, with one of the items discarded as it is the first item in the foil.

[31] A color gradient has been added to this table to improve interpretability at a glance. Cells shaded in green represent higher probabilities that the examinee is likely naïve of crime-relevant information, whereas cells shaded in red represent instances in which the examinee is more likely to be concealing information.

It can be inferred from this table that an examinee with a score of three, upon completing a CIT comprised of eight foils, has a high likelihood (84.1%) of being truly naïve of crime-relevant information and should be classified as NRI. Contra wise, an examinee with a total score of eight, after completing a CIT comprised of 4 foils, has a very low probability (.2%) of being truly naïve, and should be classified as RI. The ability to predict error estimates for each CIT provides a means of reducing the number of false positives, thus, the CIT provides a level of protection for innocent examinees. This is a considerable strength relative to the frequency with which standard CQT-based polygraph interviews result in false-positive classifications (Krapohl & Velez, 2001).

2.6.1.2 Recognition vs. Emotion

A key distinction between the CQT and the CIT is the method by which they are used to identify deception. The CQT is a direct attempt to identify deception as examinees are directly asked about their involvement in a crime (e.g., “Did you murder John Doe?”). Posing such a question to an examinee will elicit a variety of concurrent emotional responses that can make the identification of deception difficult. A guilty examinee will likely encounter feelings of stress and anxiety when posed with such a question as he or she is afraid of being convicted, while an innocent examinee will likely encounter similar feelings due to the fear of being wrongly accused. The CIT eliminates this conundrum by focusing solely on the examinee’s recognition of crime-relevant stimuli. Innocent examinees having no knowledge of a crime should feel indifferent to the stimuli items constituting each foil. Furthermore, if an innocent examinee is experiencing a general state of stress or arousal, their physiology will likely remain constant across foil items and can thus serve as a baseline. Consequently, the fear of being wrongly accused is drastically reduced in a CIT, resulting in a much clearer data set for the examiner to interpret.

2.6.1.3 Adaptability of the CIT

The guidelines used to construct and administer CITs are malleable contingent upon a number of factors, including the context, the availability of salient target items, the sensors available for data collection, and the preferences of the examiner. This renders the potential utility of the CIT much higher than if the test had to be conducted using a stringent protocol. Examiners can tailor the number of foils, the number of items in each foil, the measures used for data collection, the medium by which stimuli are presented to the examinee[32], etc. This is a distinction from the CQT as polygraph examinations used for criminal investigations must be conducted in accordance with specific regulations and guidelines established by the American Polygraph Association (APA) (Krapohl & Velez, 2001). The inherent flexibility of the CIT should solidify it as a viable alternative for use in a wide assortment of criminal interviewing contexts and applications, including automated credibility assessment interviews.

2.6.1.4 Test Duration

The actual administration of the CIT requires a brief period of time; however, it is similar to the CQT in that they both require a pre-test interview. The CQT pre-test interview requires the examiner to complete a number of tasks, including the creation of control questions[33] that should induce an average person to respond

---

[32] Audibly or visually.

[33] E.g., “Have you ever lied to someone who trusted you?”

deceptively[34]. This interview must continue until the examiner is confident that the finalized question set will prompt guilty examinees to be most concerned about relevant questions, while innocent examinees will be most concerned about control questions. A point of distinction between the CQT and CIT is that the preparation necessary to conduct a CIT[35] can occur prior to the arrival of the examinee, thus, the actual CIT pretest interview and test can be completed in a short period of time. Figure 5 contains estimates of average CIT[36] durations based on the number of foils and the number of stimuli per foil.

	Number of Stimuli per Foil					
Number of Foils	3	4	5	6	7	8
1	0.8	1.0	1.3	1.5	1.8	2.0
2	1.5	2.0	2.5	3.0	3.5	4.0
3	2.3	3.0	3.8	4.5	5.3	6.0
4	3.0	4.0	5.0	6.0	7.0	8.0
5	3.8	5.0	6.3	7.5	8.8	10.0
6	4.5	6.0	7.5	9.0	10.5	12.0
7	5.3	7.0	8.8	10.5	12.3	14.0
8	6.0	8.0	10.0	12.0	14.0	16.0
9	6.8	9.0	11.3	13.5	15.8	18.0
10	7.5	10.0	12.5	15.0	17.5	20.0

This table was constructed assuming 15-second intervals between stimuli

Figure 5 - Estimations of CIT Durations (in minutes)

Based on the values in Figure 5, a CIT comprised of five foils, containing six stimuli per foil, would result in an overall test duration of 7 minutes and 30 seconds.

---

[34] A fundamental assumption of the CQT is that the interview will be constructed with control questions triggering an innocent examinee to experience heightened arousal as compared with relevant questions.

[35] E.g., visiting the crime scene to identify salient target items, pairing target items with equally plausible nontarget items, etc.

[36] Estimates do not include the pretest interview.

Increasing the number of foils to eight would increase the overall test duration to ten minutes, thus, CITs designed in accordance with best practices require a minimal time expenditure.

2.6.2 Disadvantages of the CIT

Disadvantages of the CIT include: the importance and difficulty of selecting stimuli, countermeasures, and obstacles preventing adoption and use of the CIT by criminal examiners. Each of these disadvantages is discussed in more detail in the following sections.

2.6.2.1 Stimuli Selection

Stimuli selection is arguably the most critical aspect of administering a CIT. Creating a CIT that can elicit sufficient physiological responses to multiple target items requires time and careful consideration on the part of the examiner. Failing to include salient target items can result in a false negative outcome from the test; however, including target items that persons other than the perpetrator have knowledge of can result in false positives. To construct a CIT with discriminatory power, the following guidelines should be considered:

1. Target items are based on central features of the crime which the perpetrator can likely recall (Nahari & Ben-Shakhar, 2011).
2. There exists an adequate number of salient target items to construct a robust CIT (Krapohl et al., 2009).
3. Target items considered for use are researched to ensure that they are not accessible in the public domain (M. T. Bradley et al., 2011; Matsuda et al., 2012).

4. Nontarget items are equally plausible and prevent an innocent examinee from identifying the target item (Patrick, 2011).
5. A sufficient number of foils are used to reduce the probability that an innocent examinee is wrongly accused (Krapohl, 2010).

The probability of finding salient target items increases if investigators can visually inspect the scene of the crime to identify details that would likely be recalled by the perpetrator (Matsuda et al., 2012). The selection of target items is often complicated by the prevalence of crime-relevant information distributed by media outlets. In Japan, examiners conduct research to account for crime-relevant information that has been released into the public domain (Matsuda et al., 2012); information identified during the examiner's search is withheld from use.

An additional complication associated with identifying target items is the frequency with which criminals are intoxicated or under the influence of illegal substances when committing a crime (Krapohl & Velez, 2001). Such circumstances can drastically reduce the accuracy with which an examiner can identify target items that will prove salient. Based on the criteria enumerated previously, and the additional complications that can inhibit the selection of stimuli for CIT use, it is evident that this process has numerous pitfalls.

2.6.2.2 Countermeasures

When developing methods and technologies to aid in the detection of deception and concealed information, it is important to consider the potential use and effectiveness of persons employing countermeasures[37] to mitigate the accuracy of

---

[37] By definition, psychological characteristics are not considered countermeasures; however, it is worth noting that some psychological traits may inhibit the effectiveness of the CIT (e.g., antisocial behavior and psychopathy) (Verschuere, Crombez, Koster, & De Clercq, 2007; Verschuere, 2011).

those methods and technologies. Empirical investigations of the CQT indicate that its accuracy can be reduced when mental and physical countermeasures are employed (Honts & Kircher, 1994; Honts, Raskin, & Kircher, 1987; Twyman, Schuetzler, et al., 2013). Shortly after the initial test of the CIT (Lykken, 1959), Lykken conducted a study to evaluate the robustness of this approach to countermeasures (Lykken, 1960). While his initial findings indicated that countermeasures did *not* have a profound effect on the ability of the CIT to identify concealed information, more recent research has reported that countermeasures *can* effectively be used (Ben-Shakhar & Dolev, 1996; Ben-Shakhar, 2011, 2011; Elaad & Ben-Shakhar, 1991; Matsuda et al., 2012). The CIT is especially vulnerable to countermeasures as it is standard practice to only collect EDA data during the interview. As the CQT traditionally utilizes an array of sensors, the level of difficulty required to mitigate each of these sensors requires a greater level of training than simply attempting to diminish the accuracy of EDA alone. Krapohl (2010) recommends the collection of respiratory data in addition to EDA during a CIT as it can serve as a means of identifying the attempted use of countermeasures during the test. In Japan, CIT examiners utilize multiple sensors in an effort to identify mental and physical countermeasures (Matsuda et al., 2012).

2.6.2.3 Adoption and Use

A considerable disadvantage plaguing the CIT is a culture of criminal examiners fixated on the use of deception-detection techniques rather than utilizing recognition-detection methods (Krapohl, 2010). This is a troubling trend considering the thorough review of literature presented in this chapter supporting both the validity of the CIT as well as the relative ease with which it can be conducted. Podlesny's (1993) finding that only 13% of FBI investigations could have utilized a CIT is important to acknowledge; however, the successful adoption and use of the CIT requires a paradigm shift in how criminal investigations, criminal interviews, and media operations are conducted. Japan has successfully implemented the CIT as a viable interviewing technique (Matsuda et al., 2012) as roughly 5,000 CITs are conducted there annually (Osugi, 2011). Other countries and law enforcement agencies have an opportunity to mimic the success of the CIT in Japan and take advantage of an interviewing technique that is theoretically and empirically supported, easy to administer, and has controls built in to protect innocent examinees. The development of innovative adaptations of the CIT which draw on novel sensors, and can thus be used in new contexts, may also serve as a catalyst for CIT use in a traditional criminal-interviewing context (Matsuda et al., 2012).

2.7 Conclusion

The CIT is a recognition-based criminal interviewing technique underutilized by the vast majority of criminal examiners and law enforcement agencies worldwide. This underutilization occurs despite the presence of sound theoretical support and extensive empirical testing affirming its validity. The standard measure predominantly used for data collection is EDA; however, other PDD measures often used for standard polygraph examinations can be used in conjunction with EDA. albeit with marginal gains in discriminatory power. The CIT inherently protects innocent examinees and its structure permits quantitative estimates of error probabilities. The CIT can be conducted in a short period of time; however, a considerable amount of preparation and due diligence is required to construct a sufficient number of foils, to identify salient target items, and to include equally-plausible nontarget items which render the test effective. Due to its adaptability and feasibility for automation, and a growing interest in identifying new sensors that can be used for data collection, the

CIT has the potential to be modified for use in nontraditional interviewing paradigms, including automated credibility assessment interviews.

3 NOVEL APPROACHES TO DATA COLLECTION IN THE CIT

3.1 Introduction

The CIT has the potential to contribute to criminal investigations in a number of capacities; however, it is underutilized by examiners (Krapohl et al., 2009), plagued with the distribution of crime-relevant information by the media (M. T. Bradley et al., 2011; Matsuda et al., 2012), and limited in its applicability due to the manner in which criminal investigations are conducted (Podlesny, 1993). Creating a climate in which the CIT can be widely utilized requires paradigm shifts in how the test itself is perceived, how media outlets report on crime, and how criminal investigations are managed. While these changes are possible, and may occur in the future, they are unlikely to occur rapidly or simultaneously. In the short term, harnessing the power of the CIT may require adaptation through the development of novel sensor technologies[38] and novel test-administration methods. Additionally, identifying new circumstances in which a system capable of conducting adaptations of the CIT is paramount.

One operation which could benefit from a system capable of conducting adaptations of the CIT is border security[39]. Personnel conducting screening operations along the border are tasked with making rapid decisions about the credibility and intent of border crossers while contending with large volumes of traffic. Technologies have been developed to aid in the screening of vehicles, commercial traffic, cargo,

[38] The use of 'novel' in this chapter does not suggest that this dissertation is proposing oculometrics and vocalics as being a completely novel approach. Prior work that has utilized these sensors for a CIT is discussed later in this chapter. The purpose of Chapter 4, and the accompanying study presented in Chapter 6, is to further evaluate and validate these two "new" data collection approaches for use in a CIT context. As oculometrics and vocalics have not been adopted or used by practitioners, they are still referred to as being novel in this dissertation.

[39] Other proposed applications of such a system include bank loan applications and audits (Derrick, 2011).

and personal belongings, but systems designed to augment or automate human veracity assessments have yet to be deployed in this context. When interviewed about ways in which screening processes could be optimized, representatives of several government agencies[40] responded that "…they would like to see technology fielded that will aid in the detection of hostile intent, criminal backgrounds, and deception when interviewing people at ports of entry and when apprehending suspects in the field" (Derrick et al., 2010, pp. 41–42).

A system designed for use in ports of entry or in the field would require the utilization of an automated credibility assessment system capable of collecting data without contacting the examinee. The following sections present relevant literature on two noncontact sensors that have been researched for use in conducting a CIT. A summary of this literature, followed by a research question, is presented at the end of this chapter. An empirical study exploring the research question proposed in this chapter is presented in Chapter 4.

3.2 Novel Approaches to Data Collection in the CIT

An extensive foundation of research pointing to the validity of the CIT exists (Ben-Shakhar & Elaad, 2003; Elaad, 1990; Gamer, 2011; Krapohl et al., 2009; Krapohl & Velez, 2001; MacLaren, 2001; Yokoi et al., 2001); however, this research is based on the use of traditional CIT sensors which are not conducive for automation due to the dynamic nature of sensor configuration. For example, to measure EDA, an examiner must affix two electrodes to the fingers or palm of the examinee, then conduct a brief data collection to verify that the electrodes are yielding baseline readings that appear to be accurate and reliable. During this process, the examiner

[40] The representatives referred to in this section were from the following agencies: U.S. Customs and Border Protection (CBP), Immigration and Customs Enforcement (ICE), and the Federal Bureau of Investigation (FBI).

may also examine the condition of the examinee's fingers,[41] and apply electrolyte gel or paste to improve the connection of the electrodes on the surface of the skin (Grey & Smith, 1984). These tasks are associated with configuring only *one* traditional PDD measure used for CITs; the processes associated with configuring and testing *multiple* PDD sensors is not feasible for an automated interviewing process designed to *reduce* the responsibilities of human screeners and examiners.

A number of unconventional data-collection approaches have been proposed and investigated for CIT use (Matsuda et al., 2012), including: reaction times, facial responses, electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and event-related potentials (ERP). These measures could be used as surrogates of traditional PDD measures, or used to augment and improve upon the accuracy rates of traditional PDD measures (Allen et al., 1992; Allen & Iacono, 1997; Ambach, Stark, Peper, & Vaitl, 2008; Ben-Shakhar & Meijer, 2012; Ekman, 2001; Gronau, Ben-Shakhar, & Cohen, 2005; Langleben et al., 2002, 2005; Nose et al., 2009; Seymour & Kerlin, 2008; Seymour, Seifert, Shafto, & Mosmann, 2000; Verschuere, Crombez, Degrootte, & Rosseel, 2010; Verschuere & De Houwer, 2011); however, for the purpose of automated screening, they introduce new limitations. For example, collecting and interpreting neural activity requires the time-consuming process of carefully configuring an EEG cap on the head of the examinee. Measuring variations in blood oxygenation by interpreting fMRI readings requires the use of expensive and specialized equipment. If advances in technology allow these new approaches to be used without extensive configuration processes, expensive equipment, or the need to contact the examinee, they should be considered for an

[41] This is done to identify abrasions or skin conditions (e.g., exceptional dryness) which may inhibit accurate or consistent data collection.

automated noncontact screening system; however, in their current state, they are not feasible for use in such a context.

Two novel approaches used to collect data for concealed information testing are oculometrics and vocalics (Burgoon, Derrick, et al., 2008; Derrick et al., 2010, 2011; Elkins et al., 2012; Elkins, 2011; Nunamaker Jr., Elkins, et al., 2012; Twyman, Burgoon, Elkins, & Proudfoot, 2013; Twyman, Schuetzler, et al., 2013; Twyman, 2012). These methods do not require contact with the examinee and can be used to measure physiological and behavioral phenomena linked with the act of deceiving or concealing information. These approaches have the potential to render the CIT practical for use in an automated credibility assessment context. The following sections contain a literature review supporting the practicality and utility of using oculometrics and vocalics for data collection during a CIT.

3.2.1 Oculometrics

Oculometric measures, including: pupil dilation, eye-gaze fixation points, eye-gaze fixation durations, eye-gaze fixation patterns, and saccades[42] contain a wealth of information concerning the cognitive and physiological states of a person. One application of interpreting oculometric behaviors includes the identification of recognition (Althoff & Cohen, 1999; Ellis, Shepherd, & Davies, 1979; Ryan Jr. et al., 2003; Stacey, Walker, & Underwood, 2005). A prevalently-researched oculometric approach used to identify recognition with faces is the interpretation of fixation patterns (Althoff & Cohen, 1999; Ellis et al., 1979). A review of key research efforts in this domain is presented in the following section; however, inconsistencies and contradictory results are present in this literature. In response to these inconsistencies,

---

[42] A rapid movement of eye gaze in one direction.

literature investigating a CIT-based approach is presented as an alternative means of using oculometrics to identify recognition.

3.2.1.1 Using Eye-Gaze Fixation Patterns to Identify Recognition

Ellis and colleagues (1979) conducted three experiments to investigate the presence of eye-gaze fixation patterns associated with viewing familiar and unfamiliar faces. In these experiments, participants viewed a corpus of photos while their eye fixations were measured. Researchers worked to identify patterns present in the data collected while participants viewed whole faces, internal facial features (eyes, nose, and mouth), and external facial features (remaining features of the face not counted as being internal features) to see if any fixation patterns accompanied familiarity. The first experiment presented participants with faces of famous individuals. Results from this experiment indicated that participants predominantly fixated on internal facial features while viewing these photos. In the second experiment, participants viewed a corpus of unfamiliar faces; no fixation patterns were found. Finally, the third experiment presented participants with famous faces mixed with other famous faces. A tendency to fixate on the internal features of familiar faces was again present. The results of *this* work suggest that internal facial features are utilized more than other facial features in the context of face recognition.

Althoff and Cohen (1999) also investigated the presence of fixation patterns when viewing famous and non-famous faces. Althoff and Cohen (1999) reported a tendency for participants to fixate eye gaze more frequently on the interior facial features of *unfamiliar* individuals, orthogonal to the results of Ellis et al. (1979). In addition to this finding, the authors provide evidence of an eye-movement-based memory effect, an effect that "…can be seen as a change in the nature of processing of novel versus repeated items, with implications for other effects of prior exposure such as those seen in examples of repetition priming" (Althoff & Cohen, 1999, p. 997). Hannula and colleagues also reaffirm the notion that oculometric behavior can be used to identify memories of previous experiences (Hannula et al., 2010).

Stacey et al. (2005) conducted a number of experiments evaluating the reliability of using eye-gaze fixation points as a means of identifying familiarity. The results of this work further call into question the feasibility of using fixation points on facial features as a reliable means of identifying familiarity. Based on the results of two experiments, Stacey and colleagues (2005) found that internal facial features are frequently used for face recognition; however, significant differences in viewing either internal or external features for familiar versus unfamiliar faces was not present. A third experiment presented subjects with multiple images simultaneously. The results of the third experiment indicate that there exists an increased duration of fixation on internal facial features when participants evaluate familiar faces, but this only occurred when participants viewed multiple images simultaneously (Stacey et al., 2005).

Ryan and colleagues (2007) studied the effect of previous exposure to images on eye-movement scanning patterns during subsequent viewings of those images. During the first experiment, participants were presented with multi-face displays of non-famous individuals, some of which contained novel faces while others contained a target[43] face. The duration of each fixation point was measured and used for analysis. Researchers were able to reliably classify the images as being target faces with high rates of accuracy. The second experiment required subjects to choose a familiar face, and as in experiment one, participants were presented with a variety of images simultaneously, including repeat famous, repeat non-famous, and novel-

[43] A face that the participant had been exposed to previously.

famous faces. High accuracy rates were reported for classifying familiar faces based on fixation point durations. The final experiment in this study investigated eye fixations when participants were instructed to look away from the familiar face. The results indicate that participants indeed averted eye gaze from the familiar face; however, when presented with a familiar face not previously shown during the experiment, subjects fixated longer on that face. These findings suggest that the memory of a familiar face, and the act of verifying that it is indeed a familiar face, is obligatory, and will initially override control over eye behavior, at least during the initial recognition of that image.

3.2.1.2 Using A CIT Approach to Identify Recognition

Schwedes and Wentura (2011) extended the work of Ryan and colleagues (2007) by examining the utility of measuring fixation durations[44] within the context of administering a concealed information test. In this study, participants were exposed to three different conditions of face-image configurations in an attempt to distinguish between recognition and preparing to select an object[45]. The findings of this work indicate that (1) participants exhibited longer durations on familiar faces when tasked with identifying familiar faces but concealing their recognition of those faces, and (2) when tasked with selecting familiar faces, participants exhibited longer fixation durations than when tasked with concealing their knowledge of a recognized face (Schwedes & Wentura, 2011).

Derrick and colleagues (Derrick et al., 2011) leveraged a novel adaptation of the CIT in an attempt to measure recognition. Their approach entailed the manipulation of an object in a photo that only one condition had any knowledge of to

---

[44] Fixation durations on familiar faces had been found to be longer than fixation durations on unfamiliar faces (Ryan, Hannula, & Cohen, 2007).

[45] Referred to as *response intention* (Ryan et al., 2007; Schwedes & Wentura, 2011).

determine if participants that were familiar with the object would view the photo differently than participants with no prior knowledge of the object in the photo. More specifically, participants in the "guilty" condition constructed an inert improvised explosive device (IED) and packed it in a bag. Figure 6 contains an image of the IED used in this experiment.

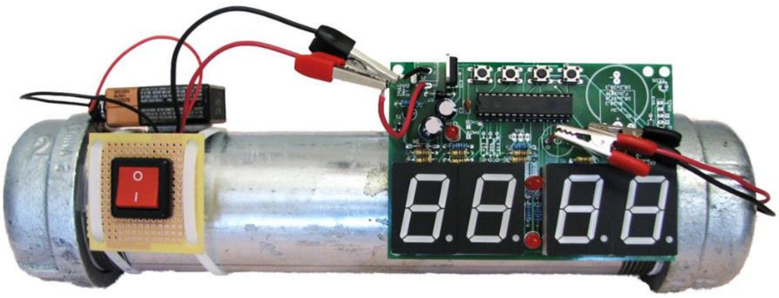


Figure 6 - IED Used in Derrick Bomb Study

The control group packed only households items in a bag and had no knowledge of the IED. Upon packing the bag, participants in each group passed through a simulated screening environment during which a number of images were displayed. One image was of the IED but a key component of the IED had been removed in the photo[46]. An analysis of oculometric data found that participants in the "guilty" group fixated their eye gaze on the location in the photo where the missing component should have been while participants in the control group fixated on a different area of the photo. Based on oculometric differences between conditions, researchers were able to discriminate between guilty and innocent participants with a high rate of accuracy.

[46] In the manipulated photo, the yellow piece of plastic housing the red trigger mechanism, battery, and battery hook-ups was removed leaving the entire left half of the pipe unobstructed.

The findings of (Derrick et al., 2011) are a promising proof-of-concept; however, this study has a number of limitations that render this approach infeasible for use in a field setting. For example, presenting a *manipulated* photo of an IED in an actual screening paradigm is unrealistic as it would be impossible to anticipate the type of bomb and bomb components that a smuggler would attempt to bring through a checkpoint. Furthermore, this approach would require not only anticipating the exact type of bomb that would be smuggled through a checkpoint, but it would also require preemptively manipulating a photo of that bomb before presenting it to examinees. Due to these limitations, Twyman (2012) conducted an experiment to investigate the impact of abstraction on the target item. This experiment utilized a similar experimental task[47] but differed in the type of stimuli presented to examinees. In this study, words[48] were displayed in groups of four on a slide, with one word in each foil of slides designated as the target item (Twyman, 2012). Refer to Figure 7 for an example of a slide presented to participants in this experiment (including the eye-gaze fixations of one participant).

---

[47] Participants packed a bag and passed through a simulated screening checkpoint.

[48] As opposed to the lone objects displayed during the Derrick et al. (2011) study.

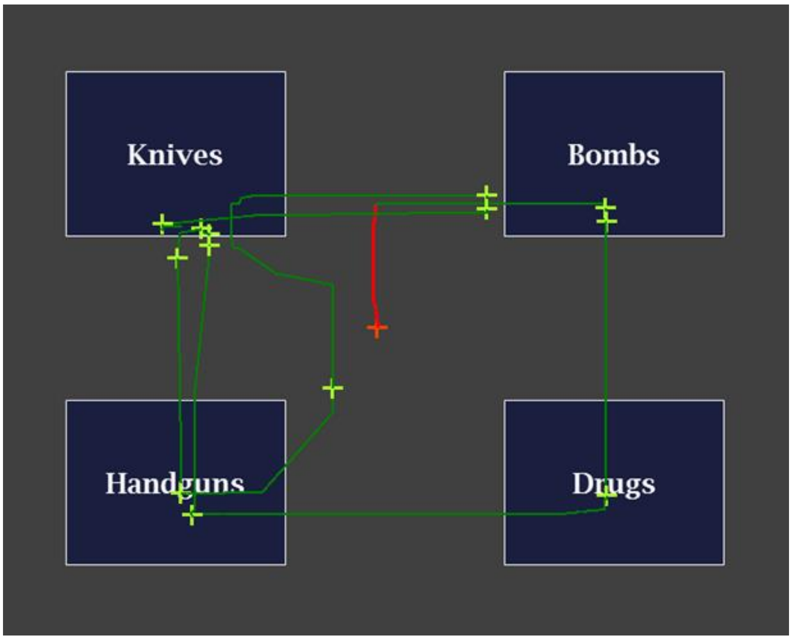


Figure 7 - Example Slide used by Twyman (2012)

The results of this study indicate that guilty participants had a greater propensity to demonstrate an initial saccade toward the critical item. Upon viewing the target item, guilty participants averted eye gaze from the target item and then fixated on the center of the screen. It is possible that this tendency to avert eye gaze from the target item and view a benign point on the screen is linked to the fight-or-flight response[49] (Cannon, 1929). It may also be an element of behavioral control (DePaulo, Kirkendol, Tang, & O'Brien, 1988).

3.2.1.3 Using Pupil Dilation to Identify Recognition

Pupil responses have also been found to be a reliable indicator of familiarity. Early evaluations of pupil dilation report that it occurs in conjunction with both short-term memory (STM) (Beatty & Kahneman, 1966) and long-term memory (LTM) (Janisse, 1977) retrieval tasks (Heaver, 2011). Additional support for these findings

---

[49] Interviewees are unable to physically escape the aversive stimulus in front of them by leaving the room, thus, they may resort to averting their eyes away from the threatening stimulus as a means of "escaping" it.

was presented by Gardner and colleagues (1975) as they found evidence of pupil dilation being linked with cognitive processes associated with the storage and retrieval of memory (Gardner et al., 1975). The tendency for pupils to dilate in response to a repeated exposure to a stimulus is often referred to as the Pupil Old/New Effect (PONE) (Heaver, 2011). Maw and Pomplun (2004) tested the PONE using a corpus of famous and non-famous faces and found that participants demonstrated increased pupil dilation when viewing images of famous faces (Maw & Pomplun, 2004).

Autonomic responses associated with the orienting response have also been found to result in increased pupil dilation (Goldwater, 1972; Nieuwenhuis, De Geus, & Aston-Jones, 2011; Nunnally, Knott, Duchnowski, & Parker, 1967). This finding suggests that pupil dilation can be used as a valid measure of detecting concealed information (M. T. Bradley & Janisse, 1981; Janisse & Bradley, 1980; Lubow & Fein, 1996), thus, the observation of target items in a CIT should result in increased pupil dilation. Research has also been conducted to determine if a correlation exists between increases in EDA, the standard PPD measure used in CITs, and pupil dilation during orienting responses. This research indicates that the orienting response is a common mechanism triggering simultaneous increases in both pupil dilation and EDA (M. M. Bradley, Miccoli, Escrig, & Lang, 2008).

3.2.2 Vocalics

Research investigating the use of vocalic measures to identify deception has yielded promising results. A number of studies have reported that vocal pitch can increase due to increases in stress and arousal (Apple, Streeter, & Krauss, 1979; Bachorowski & Owren, 1995; DePaulo et al., 2003; Elkins & Stone, 2011; Rockwell, Buller, & Burgoon, 1997; Zuckerman, DePaulo, & Rosenthal, 1981). Related literature supports a relationship between speaking deceptively and the presence of disfluencies and hesitations in speech (Buller & Burgoon, 1996; Rockwell et al., 1997) which may be attributable to the cognitive load/taxation associated with formulating a lie. Finally, vocal quality, defined as the harmonic-to-noise ratio (Boersma, 1993), has been found to decrease during deceptive speech (Elkins et al., 2012).

3.2.2.1 Using Vocal Features to Identify Deception

Despite the promising results present in the previous section, attempts at leveraging the interpretation of vocalic indicators of deception in commercial technologies has proven unsuccessful. The Psychological Stress Evaluator (PSE) was an early attempt at developing a tool to detect arousal indicative of deception using vocal cues. It measured variations in vocal frequencies associated with truthful and deceptive responses. Law enforcement organizations were quick to adopt the PSE for use in conducting criminal interviews; however, a number of studies have reported that PSE accuracy rates are close to chance (Brenner, Branscomb, & Schwartz, 1979; Horvath, 1978, 1979). The Computer Voice Stress Analyzer (CVSA) was a more recent attempt at interpreting changes in the voice to identify arousal (Verschuere, Ben-Shakhar, et al., 2011). The CVSA measured microtremors in the voice (Lippold, 1970, 1971). Similar to the PSE, a number of empirical investigations attempting to validate this tool have found the accuracy of the CVSA to be close to chance (Brown, Senter, & Ryan Jr, 2003; Cestaro & Dollins, 1994; Hollien, Harnsberger, Martin, & Hollien, 2008; Nachshon & Feldman, 1980).

Several tools have been developed since the introduction of the PSE and CVSA, although reports indicate that the majority of these tools operate at accuracy rates no better than their predecessors (Elkins, 2011; Harnsberger, Hollien, Martin, &

Hollien, 2009; Sommers, Brown, Senter, & Ryan, 2002). However, one study conducted both laboratory and field studies to investigate the accuracy rates of the TrusterPro tool, and reported classification accuracies of 80% and 90%, respectively (Gamer, Rill, Vossel, & Gödert, 2006; Van Damme, 2001). Elkins (2011) postulates that among other challenges, the inability to reliably use microtremors as a basis for deception detection may be attributable to technological deficiencies in measuring such low frequencies coupled with an unclear understanding of how deception moderates vocal features. Additional research is needed to ascertain the viability of leveraging tools designed to identify deception and concealed information using vocal features as the lion's share of the work in this area has yielded poor accuracy rates.

3.2.2.2 Using Vocal Features to Identify Concealed Information

While a majority of literature investigating vocalics within a deception-detection context has focused on a traditional truth-versus-deception paradigm, some work has been done in a CIT context. One study was conducted to evaluate the accuracy of the PSE in a CIT context; however, the findings of this study are inconclusive (Verschuere, Ben-Shakhar, et al., 2011)

Elkins and colleagues (2012) demonstrated that vocalic measures of deception can be used in a CIT interviewing context. A deception experiment was conducted using EU border guards as participants. Border guards were issued visas; some visas contained accurate information while others received visas that had been tampered with. Each participant completed an automated screening interview during which a number of questions were asked and an image of the visa was presented on a screen. Researchers monitored the responses of each participant using real-time analyses and found that individuals concealing information consistently exhibited a drop in vocal quality when making statements about the accuracy of their visa information.

Researchers used this information to discriminate between concealers and truth tellers and were able to classify each group with a high rate of accuracy.

3.3 Summary / Research Question 1

Extensive work has evaluated oculometrics as a means of identifying recognition. The use of internal versus external features to identify recognition remains inconclusive; however, an eye-movement-based memory effect has been identified, making possible the discrimination of novel and familiar stimuli using eye-gaze fixation patterns. Additional work has compared fixation duration differences when multiple face images are presented simultaneously. This work indicates that there exists a tendency to fixate on familiar stimuli longer, a byproduct of the initial recognition followed by a period of confirmation. Recent research has investigated fixation durations and initial saccades as a means of conducting an adaptation of the CIT. This work indicates that (1) fixation durations are longer on familiar faces, (2) fixation durations are longer on photo manipulations, and (3) individuals concealing information often demonstrate countermeasures to avoid detection (e.g., an initial saccade to target items followed by an aversion of eye gaze from target items). Additionally, pupil dilation has been found to increase as a result of memory retrieval, perceiving familiar stimuli, and the orienting response.

A number of vocalic measures have proven diagnostic in identifying deception, including variations in: pitch, disfluencies, hesitations, and vocal quality. However, early attempts at creating commercialized systems that leverage these cues have yielded poor results. Recently, vocalic measures have been used to identify deception and concealed information in a CIT paradigm; the findings from this work are promising.

Prior work supports the utility of using oculometrics and vocalics to identify deception and concealed information in a CIT interviewing paradigm; however, before these approaches can be considered feasible for adoption and use, empirical support pointing to their validity is needed. An experiment directly comparing the accuracy rates of these two novel approaches to EDA has not been conducted. The following research question is proposed:

RQ1: Can an automated system leveraging noncontact sensors match or exceed the accuracy of a standard EDA-based CIT?

The following chapter reports the findings of a laboratory study designed to investigate RQ1.

4 STUDY 1: A COMPARISON OF EDA, OCULOMETRICS, AND VOCALICS IN THE CONCEALED INFORMATION TEST

4.1 Introduction

The purpose of this study is to evaluate the feasibility of using two nontraditional, noncontact sensors to supplement or augment EDA as a means of identifying concealed knowledge. This was accomplished by conducting a laboratory experiment in which oculometric, vocalic and electrodermal data were collected simultaneously during automated CIT interviews. The discriminatory power of each sensor was compared to identify the utility of augmenting EDA with noncontact sensors for a traditional CIT, or conducting automated CITs using noncontact sensors.

4.2 Literature Synopsis and Hypotheses

The orienting response, investigated by Lykken (1959), occurs when conscious attention fixates on a novel or personally significant stimulus (Sokolov, 1963a). Furthermore, the spotlight theory of attention (Posner, 1980) posits that it is possible to perceive stimuli peripherally before fixating attention to a stimulus of personal significance or interest. Based on this theory, persons concealing information should be able to peripherally perceive familiar stimuli in a visual CIT and will experience an initial orienting response saccade of eye gaze fixation to that familiar stimulus. In an experimental paradigm similar to the one leveraged in this study, Twyman (2012) found that guilty participants' initial saccades were more likely to be oriented toward the target item directly after the onset of the stimuli[50]. Based on research supporting the presence of the orienting response, the spotlight theory of

[50] This visual orienting behavior was significant only during the second of two screenings.

attention, and the work conducted by Twyman (2012), the following hypothesis is proposed.

H1: Participants familiar with target items will direct eye gaze to target items upon stimulus onset whereas participants unfamiliar with target items will not direct eye gaze to target items upon stimulus onset.

Twyman (2012) found that participants concealing information attempted to avert eye gaze from the target item to avoid suspicion. Furthermore, they fixated longer on a neutral point of the screen (the center). This behavior may be attributable to the defensive response or "fight-or-flight" behavior (Cannon, 1929) which is comprised of two core components: the perception of a threat followed by actions taken by the organism to respond to the threat. It is worth noting that research testing the use of a CIT to identify familiarity with faces, without the context of a screening paradigm, found that participants fixated on familiar faces longer; there was no tendency to avert eye gaze from the images of interest (Ryan et al., 2007; Schwedes & Wentura, 2011). As the current study is utilizing an automated credibility assessment paradigm, it is hypothesized that participants presented with target items about which they must lie will experience a defensive response by averting eye gaze from target items.

H2: Participants familiar with target items will avert eye gaze from target items whereas participants unfamiliar with target items will not avert eye gaze from target items.

Research suggests that variations in pupil dilation can be used in conjunction with eye-gaze fixation patterns to identify concealed information. One of the underlying mechanisms contributing to this response is the association between memory retrieval and variations in pupillary dilation, as recognizing a previously perceived stimulus requires the utilization of cognitive processes governing memory retrieval (Gardner et al., 1975). However, a direct bond has been found between increases in pupil dilation and the orienting response, meaning that pupil dilation is an autonomic manifestation of the response already measured in the CIT (Goldwater, 1972; Nieuwenhuis et al., 2011; Nunnally et al., 1967). The validity of this measure has been solidified with a number of empirical investigations (M. T. Bradley & Janisse, 1981; Janisse & Bradley, 1980; Lubow & Fein, 1996).

H3: Participants familiar with target items will exhibit increased pupil dilation upon viewing target items whereas participants unfamiliar with target items will not exhibit increased pupil dilation upon viewing target items.

There exist measurable differences in vocalizations between truth tellers and deceivers. First, stress and arousal can result in increases in vocal pitch (Apple et al., 1979; Bachorowski & Owren, 1995; DePaulo et al., 2003; Elkins & Stone, 2011; Rockwell et al., 1997; Zuckerman et al., 1981). Additionally, deceivers often experience disfluencies and hesitations in responding as they are exerting cognitive effort to fabricate lies (Buller & Burgoon, 1996; Rockwell et al., 1997). Finally, deceivers have been shown to exhibit utterances of lower vocal quality as compared with truth tellers (Elkins et al., 2012).

H4: Participants familiar with target items will exhibit higher vocal pitch when presented with target items whereas participants unfamiliar with target items will not exhibit higher vocal pitch when presented with target items.

H5: Participants familiar with target items will exhibit longer response latencies when presented with target items whereas participants unfamiliar with target items will not exhibit longer response latencies when presented with target items.

H6: Participants familiar with target items will exhibit lower vocal quality when presented with target items whereas participants unfamiliar with target items will not exhibit lower vocal quality when presented with target items.

4.3 Deception Experiment

To test the hypotheses presented in the previous section, as well as the accuracy of the proposed sensors, a deception experiment was designed. Participants were randomly assigned to the manipulation group[51] or the control group. The experiment tasked participants with packing a bag of household items and passing through a screening checkpoint. A CIT was administered during the screening interview. The experimental task was designed with the intention of fostering a sense of realism to increase ecological validity. The task was patterned after relevant research in this domain (Derrick et al., 2010, 2011; Elkins et al., 2012; Nunamaker Jr. et al., 2011; Nunamaker Jr., Elkins, et al., 2012; Nunamaker Jr., Burgoon, et al., 2012).

---

[51] From this point forward, members of the manipulation group will be referred to as concealers.

Concealers were tasked with packing a banned item, in addition to household items, and smuggling it through a simulated screening environment. Concealers also learned the faces of the individuals that they were tasked with delivering the banned item to, as well as the name of the criminal organization that they were working for. This information (faces, banned item, and criminal organization) served as the target items used in the CIT. While passing through the simulated screening, participants were presented with sets of slides; each slide was comprised of four faces, four banned items, or the names of four criminal organizations. Each face, banned-item, or criminal-organization image was centered in one of the four quadrants of the slide. While viewing these images, and responding "Yes" or "No" concerning their familiarity with these items, oculometric, vocalic, and electrodermal data were captured. Concealers lied about their familiarity with target items in order to avoid detection. Both conditions were incentivized to appear innocent. A more detailed description of the experimental task is presented in the following sections.

4.4 Methodology

4.4.1 Participants

A combination of undergraduate and graduate students (N = 79) enrolled in business courses at the University of Arizona were recruited to participate in this study. Key demographics are as a follows: mean age of 23.7 years old; 21.9% were female, and 36.7% of participants were non-native English speakers. Of the 79 individuals participating in the experiment, 5 did not complete the task due to confessing or complications with data collection. This resulted in a data set comprised of 74 cases.

4.4.2 Design and Procedure

Researchers checked in participants at a research center located in the bottom floor of the business building. Upon arrival, preliminary instructions were given and each participant read and signed a consent form. Next, participants were randomly assigned to one of two conditions: concealers or the control group. At this point, experiment facilitators ushered participants to a separate room in which they completed a brief survey[52] and received further written instructions.

Members of the control group were informed that they would be passing through a simulated screening interview similar to one that could potentially be encountered by travelers flagged for secondary screening while passing through a border crossing or hub of public transportation. In an effort to facilitate a sense of realism, each participant in this condition packed a bag of benign "household" items provided for them in the room. Each participant then checked back with the facilitator to receive further instruction. At this point, the participant was ushered into a second room in which the interview took place.

Concealers also received instruction that they would be passing through a simulated screening interview; however, their motives were much more sinister. The instructions contained a scenario describing each participant's involvement in a drug cartel conducting smuggling operations in the region. They were tasked with bringing contraband through the simulated screening and then delivering it to one of three cartel affiliates waiting at a nearby location. To successfully make the drop-off, each participant had the responsibility of memorizing the three faces of the affiliates to ensure that they could later identify the recipients of the contraband. To facilitate this memorization process, a brief exercise was conducted in which each participant listed

[52] The pre-survey was used to collect basic demographic information.

the facial features that they would later use to identify each recipient. Refer to Figure 8 for an example of this task.

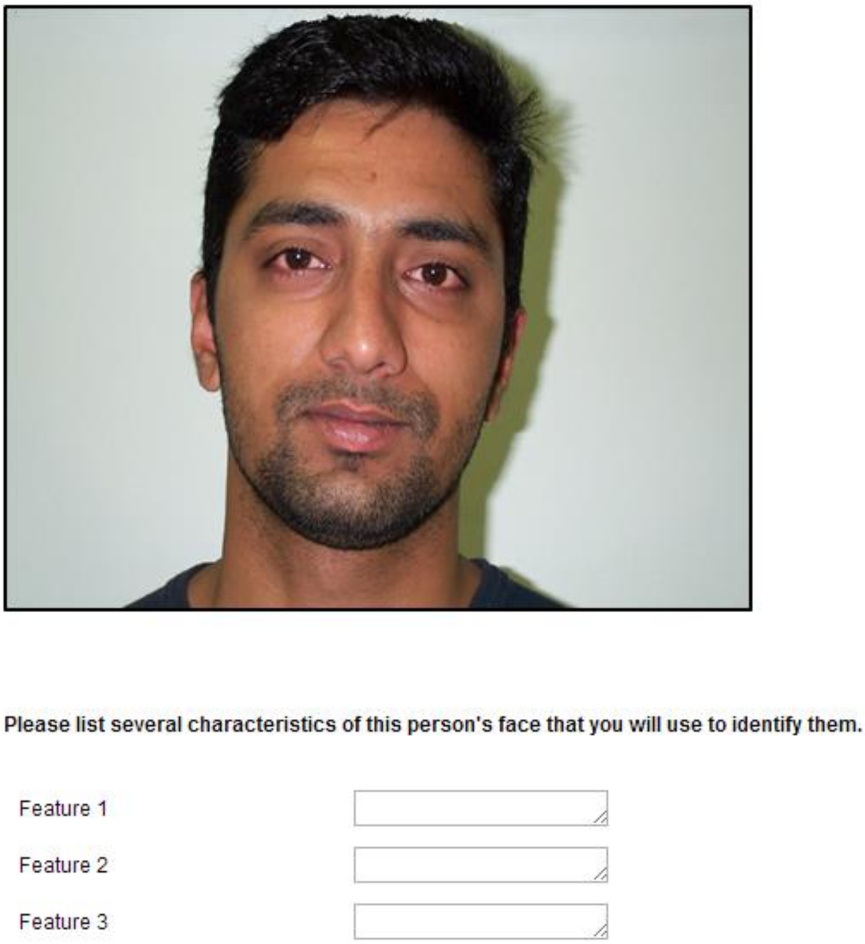
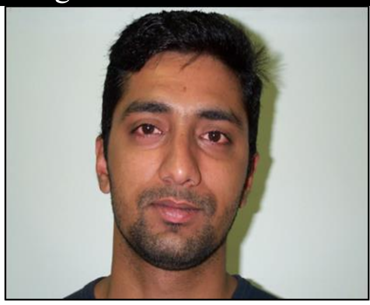


Figure 8 - Target Item Familiarization Task

When this task was completed, participants packed a bag of benign items, then concealed a fake box of ammunition in the bag. At this point, the participant was ushered into a second room in which the interview took place. It is important to note that the five target items appearing in the CIT during the interview process were comprised of the information provided to participants in this condition, specifically: face images of the three affiliates/drop-off recipients, the name of the criminal organization, and the box of ammunition. Refer to Table 1 for a list of each target item, the foil that each target item was assigned to, as well as the accompanying photo that was presented during the automated screening.

Table 1 - CIT Target Items

Foil Number	Target Item	Image
Foil 1	Face 1	
Foil 2	Face 2	
Foil 3	Face 3	
Foil 4	Ammunition	
Foil 5	Sinaloa Cartel	

Members of both conditions understood that the nature of the interview was credibility assessment. Participants were informed that a financial incentive would be distributed to examinees appearing innocent during the interview[53]. This feature was incorporated into the experiment to encourage participants to be motivated to successfully complete the task. Additionally, a warning was given indicating that participants appearing guilty would be subjected to a more time-consuming interviewing process after the initial screening.

Upon entering the simulated screening, each participant was asked to sit down at a small table adjacent to the examiner[54]. The examiner then briefly explained that both contact and noncontact sensors would be used during the screening to measure cues of deception and concealed information. After attaching the electrodermal sensors, the examiner viewed real-time EDA activity to verify that sensors were yielding normal readings. Finally, the examiner removed a cover on the eye tracker allowing the Automated Screening Kiosk (ASK) to begin the interaction.

Phase one of the interview consisted of a calibration process for the eye tracker. This process facilitates the accuracy with which data can be collected and helps standardize the measurement of participants of differing heights, positions, etc. Phase two entailed the administration of a visual CIT. During the CIT, the human examiner in the room did not present the stimuli or ask for a response. The ASK

---

[53] Only top-performing interviewees received a monetary reward; top-performing interviewees were defined as interviewees appearing to be least likely to be concealing information. This was operationalized using EDA Lykken scores. Participants falling into the top 15% of lowest scores received $40.

[54] Prior research contended that examiners conducting criminal interviews should not be seated within the field of view of the examinee; however, this practice was suggested at a time when polygraph devices were analog (Abrams, 1989; Matte, 1996; Reid & Inbau, 1977). Analog polygraph devices are comprised of moving parts that are audible and could thus be distracting to the examinee. A majority of polygraph devices in use today are digital and do not inherently consist of any features that could result in such distractions; as such, it is recommended that the examiner sit in close proximity to the examinee as a means of better identifying the use of countermeasures (Krapohl et al., 2009).

system utilized an embodied conversational agent (ECA)[55] to ask each of the three question types needed to complete the CIT. Refer to Table 2 for a list of each of the three types of images with the accompanying CIT question and the domain of possible responses[56] from the examinee. Refer to Figure 9 for a screenshot of one of the pre-rendered animations used to ask questions during the CIT.

Table 2 - Questions Presented by ASK System for each Foil

CIT Foil	Question	Response
Faces	"The following individuals are wanted by local authorities. Are you familiar with any of these people?"	Yes/No
Banned Items	"The following items are not allowed beyond this point. Are you carrying any of these items?"	Yes/No
Criminal Organizations	"The following criminal organizations are operating in the area. Are you involved with any of these organizations?"	Yes/No

[55] The ECA was created using a corpus of pre-rendered animations.

[56] Recall from Chapter 2 that there may be an inherent risk in requiring the examinee to respond by saying only "yes" or "no". Krapohl (2010) points out that this method of responding may cause dissociation between the examinee and the target item, resulting in muted physiological variations and a possible false-negative classification.

Figure 9 - Embodied Conversational Agent

A preformatted corpora of images was automatically accessed by the system and displayed systematically for a standardized duration of 12 seconds per slide[57]. The ASK system utilized the same corpus of face, banned item, and criminal organization images for each participant; however, the foils, slides, and image locations were randomized[58]. A pilot test was conducted using a large set of images to determine which images should be selected for use as target and nontarget items. For more information about pilot testing, refer to Appendix A. Figure 10, Figure 11,

[57] A common duration of stimulus presentation is 7000ms (e.g., (Schwedes & Wentura, 2011)); however, a longer duration was used in this experiment due to the life cycle of the electrodermal response. A sufficient amount of time must be allowed to accommodate the electrodermal response delay, the response itself, and normalization back to tonic levels.

[58] Slide order and image location were randomized to eliminate location-based effects; however, face-image groupings did not follow this schema. In a traditional CIT, individual stimuli are presented serially with the guideline that nontarget items must be equally plausible to target items. In this adaptation of the CIT, each slide serves as an item in the foil; however, each slide contains four images, thus, each slide can be viewed as a “micro” foil. Randomly grouping face images on a slide could result in heterogeneity due to variations in skin tone, ethnicity, attractiveness, facial features, clothing, hair color, etc. To mitigate these effects, careful consideration was made when grouping these images in sets of four.

and Figure 12 are examples of face, banned-item, and criminal-organization slides used during the experiment.

Figure 10 - Faces

Figure 11 - Banned Items

Figure 12 - Criminal Organizations

Prior to the presentation of each slide, a fixation cross was displayed for 500ms to standardize the location of eye-gaze fixations prior to stimulus onset. Refer to Figure 13 for an image depicting the fixation cross.

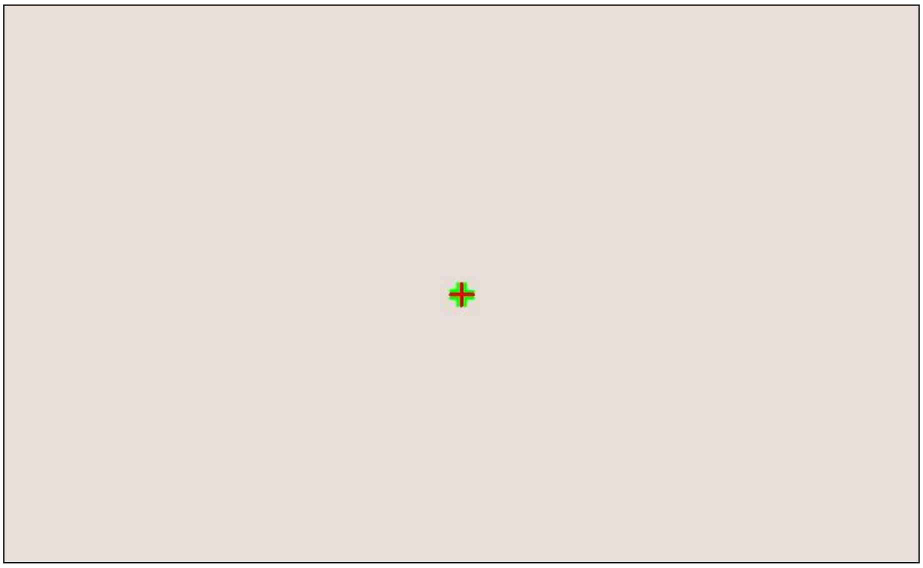


Figure 13 - Fixation Cross

Each foil was comprised of five slides total, resulting in only four remaining slides considered for analysis; one of these four slides contained the target item. The placement of the target in each foil was randomly assigned. Refer to Figure 14 for a visualization of the protocol that was used to display the slides in each foil. The first slide in the sequence is comprised of four green rectangles[59] indicating that the data from this slide will not be considered for analysis. The remaining four slides, represented by images made up of blue rectangles, are considered for analysis. The red rectangle represents the randomly assigned target item for this foil.

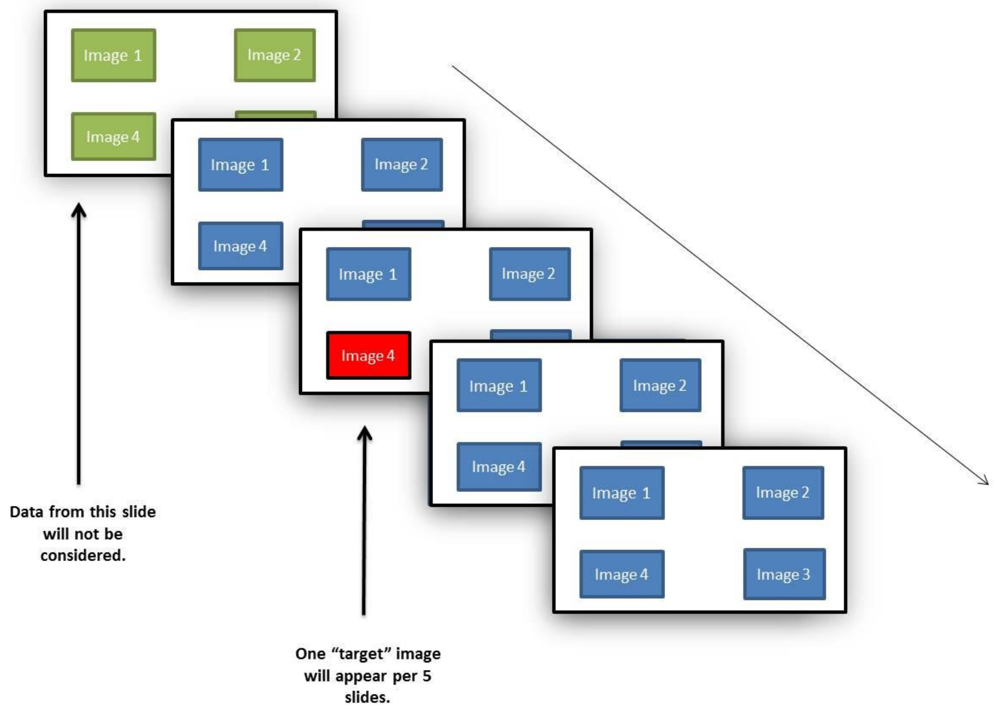


Figure 14 - Visualization of Slides Presented During a Foil

The ASK system managed the sensor array during the experimental task to ensure that data were collected at the appropriate times. The human examiner operated the software to collect EDA measurements and annotate charts whenever

[59] The green and blue rectangles in this figure represent the locations where face, banned-item, or criminal-organization images were located on each slide.

there was an external disturbance or movement of any kind by the examinee (e.g., a loud noise outside, the examinee shifting positions, the examinee coughing or sneezing, etc.). A webcam was used to collect a video feed of the interview as a means of later identifying artifacts in the data that could reduce accuracy. Figure 15 contains an image of the screening interview setup.

Figure 15 - Screening Interview Setup

Upon completing the automated screening, concealers were released from their smuggling task and directed to a room to complete post-measures designed to verify that they had no prior knowledge of target items before arriving for the experiment. Post-measures were also used to ensure that smuggling participants in fact remembered target items. Additionally, post-measures were designed to glean qualitative information about each participant's experience in the interview and solicit information concerning their knowledge or experience with criminal interviews. Members of the control group also completed post-measures.

4.4.3 Instrumentation

Oculometric data were compiled by an EyeTech™ Digital Systems VT2 infrared eye tracker. It was fixated under a 19-inch monitor on which the ECA and stimuli were presented to each participant. Data were recorded at a rate of 30 samples per second. Figure 16 contains an image of the EyeTech VT2 unit.

Figure 16 - EyeTech VT2

A noise-cancelling Andrea™ array microphone was used to capture the utterances of participants. Responses were recorded in WAV-Format (16-bit, 11.025 kHz (converted from 48khz)). Vocalic features from the audio data were captured and analyzed using algorithms developed by researchers in the BORDERS center. Figure 17 contains an image of the Andrea array microphone.

Figure 17 - Andrea Array Microphone

Electrodermal data were captured using a Stoelting CPSpro device. EDA was measured using a constant voltage system (0.5 V) using two electrodes placed on the palmar surfaces of the medial phalanx of the second and fourth fingers of the left hand (Gamer et al., 2006). The charts were analyzed and scored manually by the investigator and one additional coder (reliabilities will be reported in the following section of the paper) according to a scoring methodology proposed by Lykken (Lykken, 1959). Figure 18 contains an image of the CPSpro unit.

Figure 18 - CPSpro Unit

Data were collected from each sensor with the intention of making a classification of Recognition Indicated or No Recognition Indicated. The analysis of this data, and subsequent results, are discussed in the following section.

4.5 Analysis and Results

4.5.1 Manipulation Checks

A manipulation check was conducted after the screening interview to ensure that concealers could recall target items. While completing post-measures, concealers were tasked with reviewing the complete set of face images used during the experiment and identifying the three recipients of the contraband. Roughly 80% of concealers could recall all three faces, 12% could recall only 2, and 3% identified only a single face correctly. Concealers were also tasked with reviewing a list of criminal organizations and selecting the cartel that they were working for. Only 60% of concealers selected the correct criminal organization. Due to the salience of concealing the banned item in a bag of benign items, it was assumed that concealers could easily identify this item, thus a manipulation check for the banned item was not conducted.

It is evident from the manipulation check on criminal organizations that a large portion of concealers were unable to remember the organization for which they were tasked with smuggling the banned item. Incorporating data from these participants into a statistical model could impede the possibility of finding significant results. Furthermore, a manipulation check for the banned item was not incorporated in the post-measures, thus, it is unknown what portion of concealers recognized or remembered exactly what the banned item was. Hypothesis tests were conducted using the entire data set; however, supplementary analyses were conducted on oculometric data using only data comprised of the three foils of face images. Analyses on this restricted data set are provided in Appendix C.

4.5.2 Electrodermal Activity

Hypotheses concerning EDA were not specified in this study as there exists a wealth of research supporting the ability to measure the orienting response using EDA. However, EDA was collected for each participant to allow comparisons to be made between this measure and the oculometric and vocalic measures that hypotheses H1-H6 were specified to test. Processes used to interpret EDA data are discussed in the following sections. Hypothesis test results for oculometrics and vocalics will be presented in sections 4.5.3 and 4.5.4, respectively.

4.5.2.1 EDA Scoring

Electrodermal data were analyzed using the scoring method proposed by Lykken (Krapohl et al., 2009; Lykken, 1959) as it is the scoring method[60] reportedly used most frequently for CIT research and CIT criminal interviews (Krapohl et al., 2009). Charts were scored by multiple coders to increase the reliability of the scores and to eliminate any possible biases introduced by a single coder. To ensure objectivity, coders were unaware of the assigned condition for each EDA chart. Cronbach's α (Cronbach, 1951) and Rosenthal's effective reliability (Rosenthal, 1982) were calculated to determine inter-rater reliability. Refer to Table 3 for reliability coefficients.

Table 3 - Inter-Rater Reliabilities for Lykken Scoring of EDA Charts

Reliability Analysis	Coefficient
Cronbach's α	0.788
Effective Reliability	0.805

According to a commonly accepted standard for interpreting reliability scores (Kline, 1999), the results in Table 3 fall at the threshold between 'Acceptable' and 'Good'. These scores also exceed the recommendation of Nunnally (1978) for reliability scores to exceed 0.7.

4.5.2.2 EDA Score Distribution

To visualize the distribution of scores for each condition, a scatterplot[61] was created. Keeping in mind that scores greater than or equal to five were classified as concealing knowledge, it can be inferred from Figure 19 that a majority of true

[60] This scoring method was discussed in detail in section 2.5 of this dissertation.

[61] A number of figures and tables in this section have concealers labeled as 'Deceptive' and members of the control group listed as 'Truthful'.

negatives were properly classified, whereas about one third of true positives were deemed innocent.

Figure 19 - Scatterplot of Lykken Scores

A boxplot was also created to compare the minimum and maximum values of scores for each condition, as well as the upper/lower quartiles and the median. Refer to Figure 20 to view a representation of the boxplot.

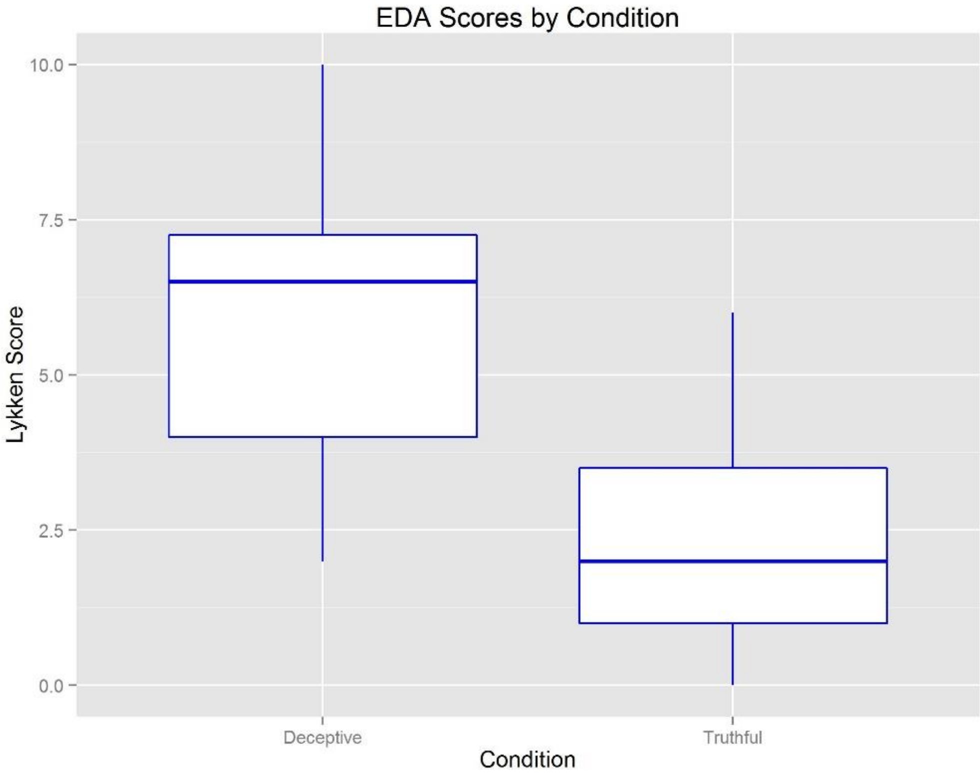


Figure 20 - Box Plot of Lykken Scores

The box plot in Figure 20 indicates that there is virtually no overlap of scores between the central quartiles of each condition distribution. It is also apparent that the distribution for concealers is skewed towards higher scores whereas the distribution for the control group is slightly skewed towards lower scores. Means and standard deviations of Lykken scores for each condition are reported in Table 4.

Table 4 - Means and Standard Deviations of Lykken Scores (by Condition)

	Mean	Standard Deviation
Deceptive Condition	6.000	2.169
Truthful Condition	2.256	1.681

The mean value of Lykken scores for concealers is 6.0, a value close to the cutoff point of 5.0, the point at or above which a participant is classified as concealing information. A mean of 2.256 for members of the control group is more than double the distance from the cutoff point relative to the mean of scores for concealers. A standard deviation of 2.169 for concealers indicates that the values of Lykken scores for concealers are widely dispersed relative to the densely-grouped control-group scores yielding a standard deviation of 1.681.

4.5.2.3 Classification Accuracy

To allow comparisons to be made between sensors, overall accuracy, sensitivity, and specificity were calculated. To calculate accuracy, classifications based on Lykken scores were compared with the condition of each participant. While inter-rater reliability was previously identified as being in the ‘acceptable’ to ‘good’ range, classification discrepancies between the two coders were present. Conflicting classifications between coders were discussed and reconciled. The resulting EDA classification accuracy rate after reconciliation was 77%. The sensitivity of a test represents its true positive rate, whereas its specificity refers to its true negative rate (Mossman & Somoza, 1991; Somoza & Mossman, 1991). Based on inferences made from the plots alone, it was anticipated that the sensitivity of EDA-based classification would be low, whereas the specificity would be high. Analysis of the data confirmed this assumption. A majority of concealers (21 of 35) were correctly identified for a sensitivity of .6. The vast majority of control group members (36 of 39) were correctly identified for a specificity of .923.

4.5.2.4 ROC Curve from EDA Data

A receiver operating characteristic (ROC) curve was created to visually represent the performance of the binary classification system assigning individuals to be either concealers or members of the control group based on their Lykken score. ROC curves are used as a means of measuring test accuracy by “providing a succinct representation of diagnostic performance across a test's entire range of possible cutoffs" (Mossman & Somoza, 1991, p. 330). Fawcett (Fawcett, 2004, p. 1) describes them as "a technique for visualizing, organizing, and selecting classifiers based on their performance."

An ROC chart is comprised of a y-axis plotting sensitivity (the true positive rate), an x-axis plotting 1 minus specificity (the false positive rate), and a straight diagonal line, originating from point (0, 0), cutting the plotting area in half; this line represents a classifier operating at "chance" levels. An accurate classifier would result in a curved line rising from the bottom-left of the plotting area with a steep slope, then passing through the top-left portion of the chart prior to traversing to the right side of the plotting area. An inaccurate classifier would result in a ROC curve closely following the "chance" diagonal. A more quantitative evaluation of a ROC curve is the area under the curve (AUC) measurement, which calculates the actual surface area below the ROC curve in the chart (A. P. Bradley, 1997; Hanley & McNeil, 1982; Mossman & Somoza, 1991). An AUC close to 1.0 indicates that the binary classifier is accurate; AUC values close to 0.5 indicate that a classifier is operating at "chance" levels. Refer to Figure 21 for the ROC curve estimating the predictive power of Lykken scores.

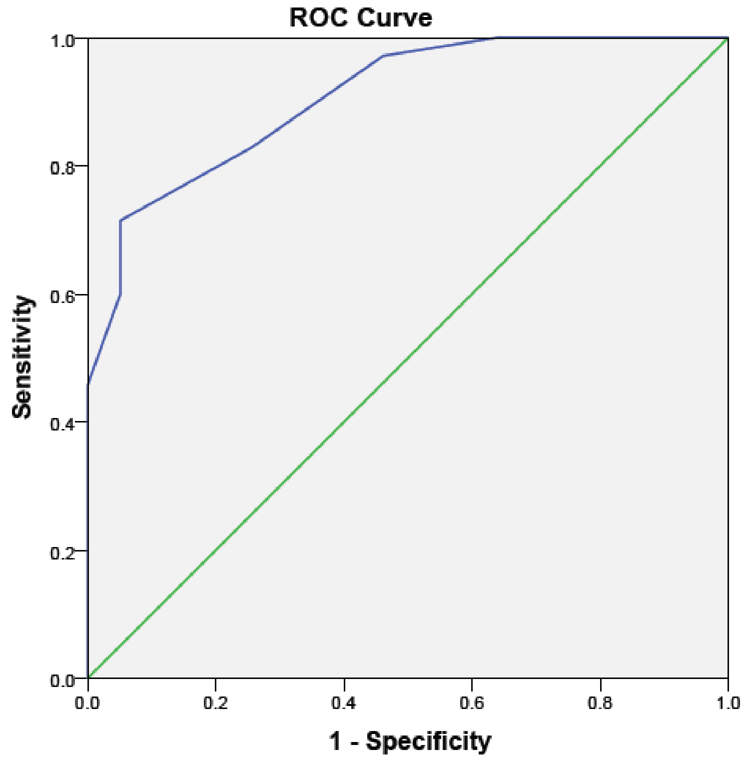


Figure 21 - ROC Curve Estimating Predictive Power of Lykken Scores for EDA

The AUC coefficient for this ROC curve is 0.906, indicating that a binary classification system constructed using the EDA data compiled for this study is considered accurate.

4.5.3 Oculometrics

4.5.3.1 Orienting Response to Target Items

H1 was tested with a logistic regression model to determine if an oculometric orienting reflex occurred in concealers when target items were presented by ASK. This type of analysis was used due to the dichotomous nature of the dependent variable. The identification of an orienting response to target items was operationalized by measuring the initial eye movement of the examinee for a distance of 20 pixels from the origin point of each examinee's initial fixation on the screen. The image closest to the eye gaze of the examinee after 20 pixels of eye movement was determined to be the first item of interest on the screen. This analysis provided evidence that H1 was not supported as there was not a significant effect. The results of this analysis are presented in Table 5.

Table 5 - Results of Logistic Regression Model for Orienting Response

Fixed Effects	β	β Standard Error
Intercept	-1.308***	0.187
Concealed Information	-0.432 (n.s.)	0.290

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

4.5.3.2 Oculometric Threat Avoidance: Eye-Gaze Aversion from Target Items

H2 was tested with a multilevel regression model (n = 7062) to identify the presence of strategic efforts by concealers to avert their eye gaze from target items. The results of this model ($t(7058) = -2.41$, $p < .05$) indicate that H2 is supported due to a reduction in the amount of time that concealers viewed target items relative to participants in the control group (3.9%). The remaining fixed effects in the model were not supported; however, in the case of target items alone, a lack of significant results is critical. This finding suggests that target items did not possess any inherent attributes that caused members of the control group or concealers to view them differently relative to nontarget items. A summary of the results from this model are provided in Table 6.

Table 6 - Oculometric Threat Avoidance (Eye-Gaze Aversion)

Fixed Effects	β	β Standard Error	t Value
Intercept	0.266***	0.010	25.189
Concealed Information	-0.008 (n.s.)	0.014	-0.562
Target Image	-0.003 (n.s.)	0.011	-0.236
Position of Target Item within the Foil	0.001 (n.s.)	0.001	0.995
Concealed Information and Target Image	-0.039*	0.016	-2.406

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

4.5.3.3 Pupil Dilation: Differences between Slides

H3 was tested with a multilevel regression model (n = 1850) to determine if concealers exhibited increased pupil dilation when target items were presented. The only significant effect in this model is the position of the target item within the foil ($t(1846) = -2.96$, $p < .01$), meaning that when the target item appeared earlier in the foil, pupil dilation was measured to decrease by about 0.017 mm or 0.68%. Concealers did not exhibit an overall pupil dilation increase when viewing an image containing a target image as compared with control images. The other fixed effects were not significant. As a measure of control, it is worth noting that the target image alone did not inherently elicit any significant changes in pupil dilation. Table 7 summarizes the results of this model.

Table 7 - Pupil Dilation (Differences between Slides)

Fixed Effects	β	β Standard Error	t Value
Intercept	2.489***	0.081	30.544
Concealed Information	-0.067 (n.s.)	0.116	-0.573
Target Image	0.025 (n.s.)	0.027	0.918
Position of Target Item within the Foil	-0.017**	0.006	-2.964
Concealed Information and Target Image	-0.004 (n.s.)	0.039	-0.102

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

4.5.4 Vocalics

4.5.4.1 Pitch

H4 was tested using a repeated measures ANOVA to determine if concealers exhibited increased pitch when presented with target items. Contrary to H4, there is no significant difference in pitch for the interaction effect of Condition and Item Type, $F(1,72) = .046$, $p = .83$. However, a significant effect is present for the main effect Foil, $F(1,72) = 3.40$, $p = .009$. To further explore this finding, descriptive statistics were calculated for each foil. The results are summarized in Table 8.

Table 8 - Descriptive Statistics for Pitch across Foils (Hz)

Foil	Mean Pitch	SD	Median	Min	Max	Range	Skew	Kurtosis
Foil 1	166.20	47.43	157.09	97.33	355.56	258.23	0.77	0.05
Foil 2	164.76	46.98	155.49	91.95	322.2	230.25	0.63	-0.46
Foil 3	162.68	45.07	153.95	92.48	293.46	200.98	0.54	-0.72
Foil 4	159.48	43.28	152.43	88.89	284.73	185.85	0.48	-0.83
Foil 5	159.95	42.14	154.49	92.02	310.41	218.39	0.52	-0.5

It is evident from Table 8 that the mean pitch of respondents trended downward as the interview progressed[62]. Figure 22 was created to compare the downward trend in both concealers and the control group, and to visualize temporal effects on differences in mean pitch.

[62] It is worth noting that the labels of Foil 1 through Foil 5 simply represent the progression of foils through the screening interview. Foil labels are *not* representative of specific foils of images.

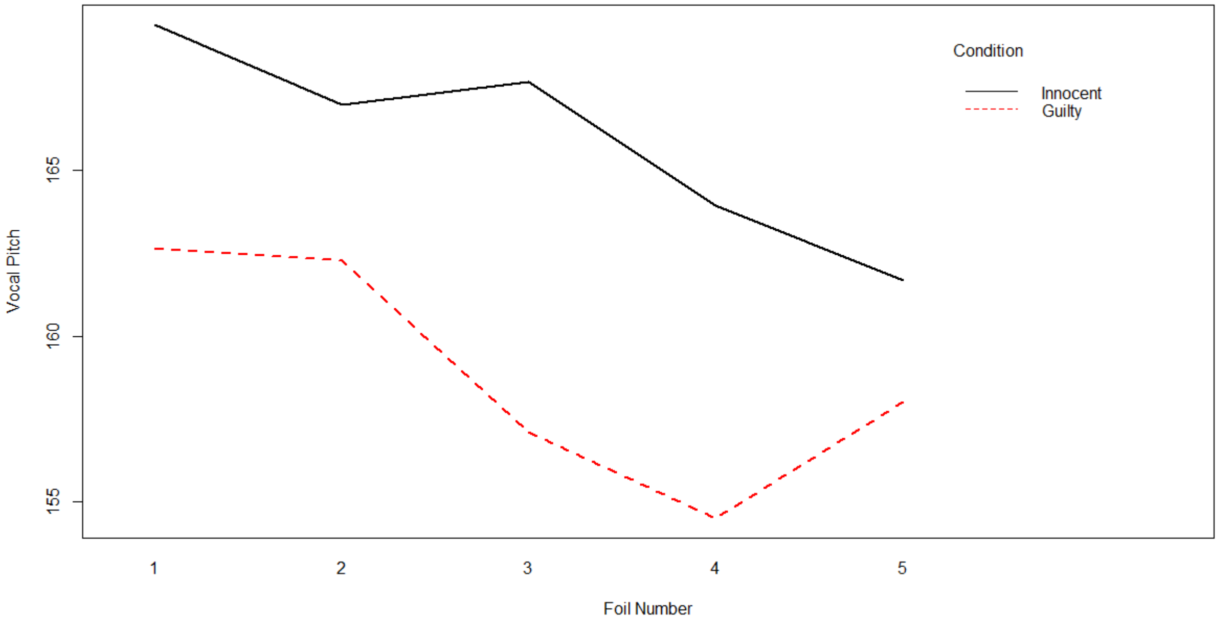


Figure 22 - Temporal Effects on Mean Pitch (Hz) by Condition

It is clear from Figure 22 that participants in both conditions demonstrated a downward trend in pitch as the screening interview progressed. This is likely due to participants habituating to the interviewing process.

4.5.4.2 Response Latency

H5 was tested using a repeated measures ANOVA to determine if concealers exhibited increased response latencies when presented with target items. The results indicate that H5 is supported as there is a significant main effect for response latency by condition, $F(1,72) = 4.485$, $p = .038$. To further explore this finding, descriptive statistics were calculated for response latency by condition. The mean response latency for concealers was 1.84 seconds, while participants in the control group demonstrated a mean response latency of 1.48 seconds. The results are summarized in Table 9.

Table 9 - Descriptive Statistics for Response Latency by Condition

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Concealing Information	1.84	1.68	1.32	0.11	10.12	10.02	2.11	6.68
Innocent	1.48	1.14	1.3	0.11	10.12	10.02	2.22	10.21

Refer to Figure 23 for a visualization of mean differences in response latency (in seconds) by condition.

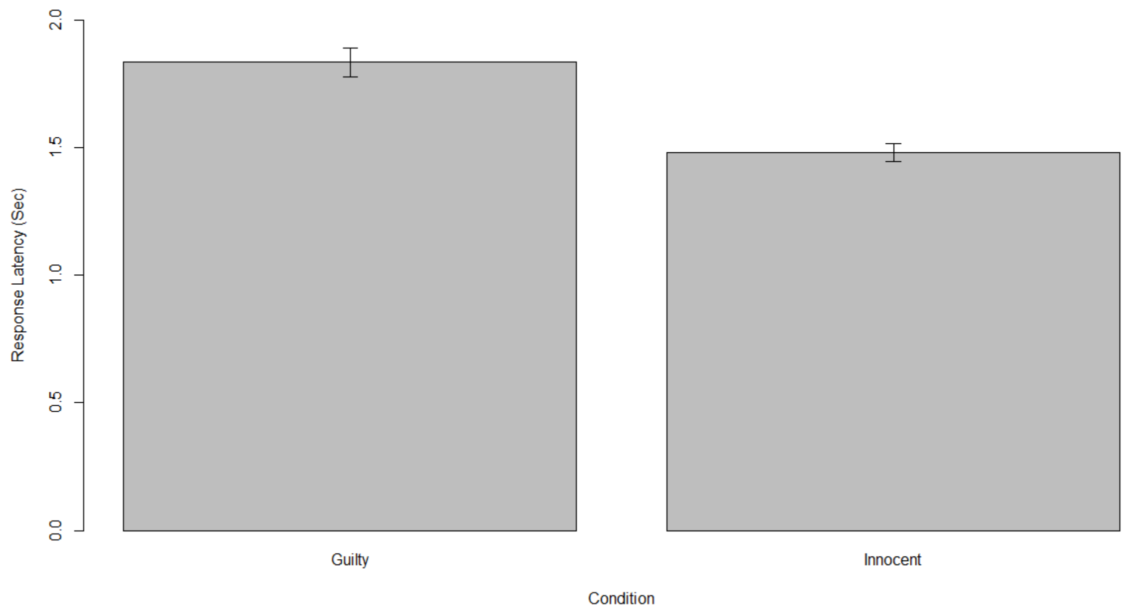


Figure 23 - Mean Differences in Response Latency (in seconds) by Condition

4.5.4.3 Vocal Quality

H6 was tested using a repeated measures ANOVA to determine if concealers exhibited reduced vocal quality when presented with target items. One record was not considered for analysis due to missing data. Partially supporting H6, there is a significant difference in vocal quality for the interaction effect of Condition and Item Type, $F(4,284) = 4.18$, $p = .045$; however, concealers demonstrated higher vocal quality when target items were present. To further explore this finding, descriptive statistics were calculated for target items and non-target items for both conditions. The results are summarized in Table 10.

Table 10 - Descriptive Statistics for Vocal Quality by Condition / Item Type (dB)

	Median	Mean	SE Mean	Variance	SD
Concealed Information and Target Item	10.06	10.26	0.25	11.13	3.34
Concealed Information and Non-Target Item	9.91	9.97	0.13	11.76	3.43
Innocent and Target Item	10.96	10.80	0.26	13.47	3.67
Innocent and Non-Target Item	10.93	10.86	0.13	13.90	3.73

It is evident from the mean vocal quality values in Table 10 that there is a difference of 0.06 dB in vocal quality between target and non-target items for members of the control group. However, there is a larger difference of 0.29 dB between target and non-target items for concealers. Vocal quality *increases* for concealers when key items are present as compared with the presence of non-target items only. Refer to Figure 24 for a visualization of the interaction effect of condition and item type on vocal quality.

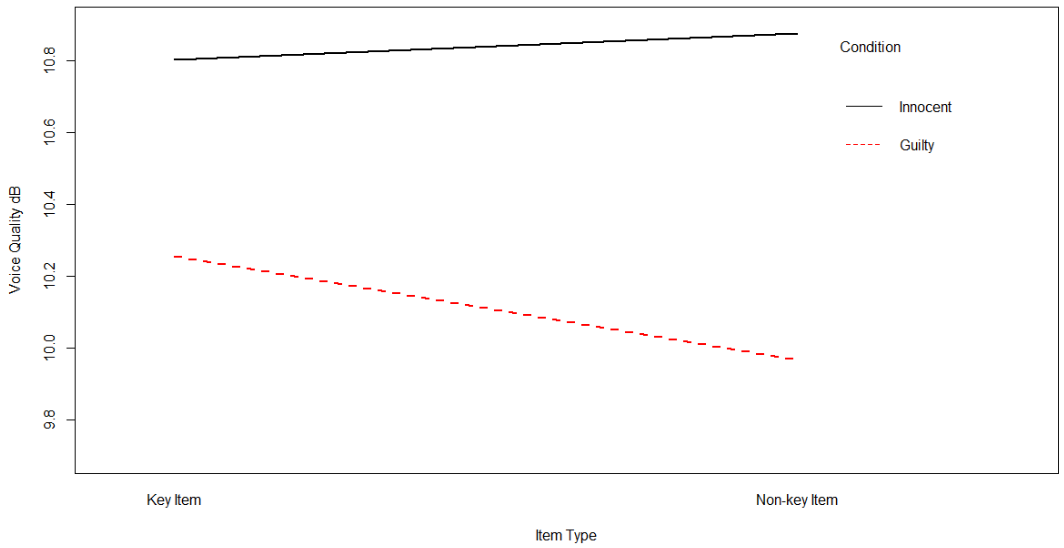


Figure 24 - Interaction Effect of Condition and Item Type on Vocal Quality

4.5.5 Summary of Hypothesis Results

The following tables contain a summary of statistics and outcomes concerning the hypotheses presented and tested in the current research. Table 11 contains this information for the oculometric hypotheses tested and Table 12 contains this information for the vocalic hypotheses tested.

Table 11 - Hypothesis Test Results for Oculometric Analyses

Hypotheses		β	β S.E.	Outcome
H1	Participants familiar with target items will direct eye gaze to target items upon stimulus onset whereas participants unfamiliar with target items will not direct eye gaze to target items upon stimulus onset.	-0.432	0.290	Reject
Hypotheses		**t**	**P**	**Outcome**
H2	Participants familiar with target items will avert eye gaze from target items whereas participants unfamiliar with target items will not avert eye gaze from target items.	-2.406	< .05	Accept
H3	Participants familiar with target items will exhibit increased pupil dilation upon viewing target items whereas participants unfamiliar with target items will not exhibit increased pupil dilation upon viewing target items.	-0.102	n.s.	Reject

Table 12 - Hypothesis Test Results for Vocalic Analyses

Hypotheses		F	P	Outcome
H4	Participants familiar with target items will exhibit higher vocal pitch when presented with target items whereas participants unfamiliar with target items will not exhibit higher vocal pitch when presented with target items.	0.0457	n.s.	Reject
H5	Participants familiar with target items will exhibit longer response latencies when presented with target items whereas participants unfamiliar with target items will not exhibit longer response latencies when presented with target items.	4.485	< .05	Accept
H6	Participants familiar with target items will exhibit lower vocal quality when presented with target items whereas participants unfamiliar with target items will not exhibit lower vocal quality when presented with target items.	4.175	< .05	Partial Support

4.5.6 Exploratory Oculometric Analyses

Two exploratory analyses were conducted using oculometric data. A description of these analyses is provided in the following sections. As with the hypothesis tests, supplementary analyses were conducted on exploratory oculometric data using only data comprised of the three foils of face images. Analyses reporting on the restricted data set are provided in Appendix D.

4.5.6.1 Oculometric Threat Avoidance: Viewing the Center of the Screen

An exploratory analysis was completed to identify any additional oculometric threat avoidance countermeasures that concealers were strategically employing in an effort to avoid detection. A multilevel regression model (n = 1850) was specified to determine if concealers were using the center of the screen as a neutral location to view in an effort to avoid looking at any of the images, as one of them could be the location of a target item. This analysis revealed that the center of the screen was indeed viewed longer by concealers than members of the control group for an average of 6.1% longer ($t(1846) = 2.24$, $p < .05$). This effect occurred when target items were both present and absent from the screen, meaning a main effect difference occurred between groups. Table 13 summarizes the results of this model.

Table 13 - Oculometric Threat Avoidance (Fixation on Center of Screen)

Fixed Effects	β	β Standard Error	t Value
Intercept	0.055**	0.019	2.826
Concealed Information	0.061*	0.027	2.239
Target Image	0.004 (n.s.)	0.008	0.500
Position of Target Item within the Foil	0.002 (n.s.)	0.002	1.048
Concealed Information and Target Image	0.018 (n.s.)	0.011	1.625

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

4.5.6.2 Pupil Dilation: Differences within slides

While the interaction effect of condition and target item did not yield a significant effect, a supplemental analysis was conducted evaluating changes in pupil dilation between target and non-target items on the same slide. A multilevel regression model (n = 7062) was specified to identify changes in pupil dilation associated with viewing target and nontarget items within a foil item. The fixed effect measuring the position of the target item within the foil was removed as it is not relevant to this analysis. Participant condition, the presence of a target item, and the presence of concealed knowledge and a target item were treated as fixed effects. The model indicates that there was a significant interaction effect between a participant concealing information and the presence of a target item ($t(7058) = 2.22$, $p < .05$), indicating that when a group of four images was presented simultaneously, one of which was a target item, a concealer would exhibit an increase in pupil dilation (0.092 mm or 3.64%) while viewing the target item. As a measure of control, it is worth noting that there were no significant changes in pupil dilation based on condition alone or the presence of a target image. Table 14 summarizes the results of this model.

Table 14 - Pupil Dilation (Differences within Slides)

Fixed Effects	β	β Standard Error	t Value
Intercept	2.523***	0.082	30.503
Concealed Information	-0.071 (n.s.)	0.119	-0.597
Target Image	0.009 (n.s.)	0.029	0.325
Concealed Information and Target Image	0.092*	0.042	2.222

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

4.5.7 Classification Accuracy Comparison

The overarching objective of this research was to simultaneously collect data using the standard sensor (EDA), in addition to two noncontact sensors (oculometrics and vocalics), during the administration of a CIT, with the intention of comparing the classification accuracies of each sensor. As reported previously in this paper, leveraging EDA data and scoring charts using the Lykken method resulted in a classification accuracy of 77% and 78% (using the complete and restricted data sets, respectively). Visualizations of EDA data (refer to Figure 19) indicated that this relatively high level of accuracy was attributable to an extremely high rate of true negatives. However, there was a high rate of false negatives as roughly one third of scores associated with guilty participants fell below the threshold for classifying an individual as guilty. Possible explanations for the high rate of false negatives are discussed in the following sections. Classification accuracies for oculometrics and vocalics were not promising despite utilizing a variety of machine-learning classification approaches; even the best-performing classifier yielded poor accuracy rates. The highest classification accuracy rate resulting from this analysis was 65% using vocalic features.

4.6 Discussion

EDA was found to be an accurate method of classifying concealers and members of the control group; however, the relatively high accuracy rate was inflated with a high classification of true negatives while roughly two thirds of true positives were correctly identified.

H1 was not supported by the data, indicating that concealers did not demonstrate an orienting response to target items when they were present. Twyman (2012) found this response to occur when participants were simultaneously viewing four images containing text. This was attributed to the spotlight theory of attention (Posner, 1980) which states that information is peripherally processed before attention shifts from the current focus. It is possible that face images are too rich in detail for peripheral processing to identify a familiar versus non-familiar stimulus while text may be a feasible stimulus to be processed peripherally.

H2 was supported as concealers averted eye gaze from target items while members of the control group did not. This finding supports the notion of oculometric threat avoidance as concealers attempted to avoid viewing an aversive stimulus as a strategic means to avoid detection.

H3 was not supported by the data, indicating that concealers did not exhibit increased pupil dilation upon viewing slides containing target items. One explanation for this finding is the duration of the slides being displayed on the screen (12 seconds). Differences in pupil dilation may have only occurred for a fraction of the time that target images were displayed, resulting in a non-significant difference in behavior between conditions for the *total* duration that each slide was displayed.

Supplemental oculometric analyses found that concealers fixated longer on the center of the screen irrespective of the presence of target items. When a target item is present, this behavior can be classified as another manifestation of oculometric threat avoidance as guilty participants avoided viewing the aversive stimulus and chose the center of the screen as a point of safety. However, an explanation for this behavior to occur when target items are *not* present is investigated in a follow up study presented in Chapter 6 of this dissertation.

Concealers also exhibited an increase in pupil dilation when viewing the target item as compared with the control items on the same slide (as opposed to the hypothesis predicting increased pupil dilation for slides containing target items versus slides not containing target items). This more granular finding indirectly supports H3, and suggests that a lack of support for H3 may be attributable to the method of analysis and not the absence of the hypothesized behavior. This increased pupil dilation for the duration of the interview is likely attributable to an increased overall level of arousal in participants concealing information. This phenomenon is also investigated in the follow up study presented in Chapter 6 of this dissertation

Vocalic analyses found that H4 was not supported, indicating that there was no significant difference in pitch between conditions when a target item was present. This may be attributed to a lack of arousal or stress as concealers may not have felt a sufficient level of risk or jeopardy while being interviewed. Additional insight on this topic is provided in the limitations section. Changes in pitch may also occur over longer temporal periods, mitigating the ability to measure more granular variations in pitch associated with the presence of target and non-target items. Supplemental vocalic analyses indicated that there are temporal effects associated with pitch as the mean pitch of concealers and members of the control group trended downward over time. This may be attributable to participants habituating to the screening environment and the protocol of the screening interview.

H5 was supported by the data as concealers exhibited longer response latencies than members of the control group. The tendency for guilty participants to exhibit longer response latencies is a manifestation of increased cognitive effort associated with fabricating the lie and/or employing countermeasures to appear innocent.

H6 was partially supported as concealers exhibited an *increase* in vocal quality when presented with target items as compared with nontarget items[63]. This increase in vocal quality may be indicative of at least two phenomena: (1) concealers experienced an orienting response to the target item resulting in an increase in vocal quality, and (2) in an attempt to be more convincing about their innocence, concealers may speak more clearly/loudly to assert their lack of familiarity with target items. This assertion of innocence resulted in increased amplitude and thus increased vocal quality. This finding parallels work conducted by Burgoon which found that "higher

[63] H6 posited that guilty participants would demonstrate lower vocal quality for target items.

motivation was associated with more complete, clear, and direct verbal messages…" (Burgoon & Floyd, 2000). The findings reported by Burgoon were significant for participants who first told the truth and then deceived, matching the CIT paradigm in which the first foil item does not contain a target item, meaning that participants in this experiment always told the truth first.

4.7 Implications

This study makes several contributions to research in a number of domains, including the: deception detection, concealed information testing, automated-screening technologies, oculometrics, and vocalics research areas.

First, some of the results reported in this paper support the findings of previous work. Prior work evaluating the use of oculometrics to conduct a CIT found that participants tended to fixate *longer* on familiar faces. This study supports the findings of Twyman (2012) as concealers monitored in a screening paradigm will avert eye gaze from target items and fixate longer on the center of the screen. This finding provides further support that individuals tend to visually avoid stimuli they find aversive, even if it is novel or personally significant.

The finding that concealers demonstrated increased pupil dilation when directly viewing target items, as compared with control images on the same slide, supports findings from a recent FRONTEX workshop conducted in Apeldoorn, Netherlands, during which concealers pretended to be soccer hooligans attempting to gain access to a match from which they were banned. This finding suggests that H3 may not be the most effective approach for detecting differences in pupil dilation, and that this alternative approach may be the most reliable method.

It is reported in the vocalics section that both concealers and members of the control group exhibited a downward trend in pitch over the course of the interview.

This finding has profound implications for the mental state of examinees as an interview progresses, as well as the format and structure of interview questions and how the ordering of questions may influence the types of responses that are present. The finding that concealers exhibited an increase in vocal quality for target items is related to findings reported by Burgoon (2000). Understanding whether or not people are more assertive while lying, or if they are more hesitant and uncertain, is a critical and ongoing question in deception literature.

In short, some of the oculometric and vocalic cues hypothesized to demonstrate significant changes in behavior were supported. These cues should be further investigated and considered for technologies used to identify deception and concealed information in an automated credibility assessment context. Subsequently, a number of the vocalic and oculometric cues hypothesized to demonstrate significant differences in behavior were not supported. These cues either require further evaluation or they may not be reliable indicators of concealed information and deception. As such, future work can confirm these findings or focus elsewhere in an attempt to identify cues that *can* contribute to a system designed to identify deception and concealed information.

From the standpoint of discriminatory power, either the effect sizes resulting from this study could not be used as robust classifiers or the classification methods that were tested were not optimal for classification using this type of data. Differences between the various cues were marginal, thus, attempting to leverage only these cues in an automated-screening context would yield an unacceptable level of false positives. However, similar CIT research investigating oculometrics and vocalics to identify concealed information have reported higher rates of accuracy. It is possible that inherent limitations in this study reduced the discriminatory power of oculometric and vocalic data, or that classification methods utilized by other researchers are more conducive to the creation of accurate classification models using this type of data. Despite these limitations, this research provides a contribution to existing work as a number of oculometric and vocalic measures found in prior work to be useful in identifying concealed information were either replicated or identified as needing additional investigation. Additionally, this study also confirms the ability of an automated system conducting a visual CIT to elicit the orienting response, as the classification accuracy of using EDA data was 77%. Future research can further evaluate these cues individually, or in concert with other sensors or technologies, in an effort to develop a robust framework for identifying deception and concealed information.

5 A NOVEL CIT METHOD

5.1 Introduction

The standard method used to conduct a CIT is the Known Solutions CIT. A Known Solutions CIT requires the examiner to specify a set of customized target items for the CIT based on crime-relevant information. In Chapter 2 of this dissertation, a number of guidelines[64] were presented which should be considered when selecting target and nontarget items. These guidelines help to ensure the salience, plausibility, and thus discriminatory power of CIT stimuli. However, there is a CIT methodology that can be employed without adhering to all of these guidelines.

Chapter 2 includes an overview of the Searching CIT. In this adaptation of the Known Solutions CIT, target items are *not* known by the examiner. Rather, the examiner formulates foils of stimuli that *may* contain details of the crime. Upon administering a Searching CIT, the examiner reviews variations in the examinee's physiology in an effort to identify stimuli that may be crime-relevant. This variation on the CIT adds a level of complexity relative to the Known Solutions CIT as *searching* for target items is more challenging than identifying physiological responses *associated with* known target items (Matsuda et al., 2012). While the Searching CIT has been used effectively in the field (MacLaren, 2001), it may be possible to identify concealed knowledge without the use of, or attempt to identify, crime-relevant information. This novel approach would utilize the presence or

[64] Stimuli Selection Guidelines: Target items are based on central features of the crime which the perpetrator can likely recall (Nahari & Ben-Shakhar, 2011). There exists an adequate number of salient target items to construct a robust CIT (Krapohl et al., 2009). Target items considered for use are researched to ensure that they are not accessible in the public domain (M. T. Bradley et al., 2011; Matsuda et al., 2012). Nontarget items are equally plausible and prevent an innocent examinee from identifying the target item (Patrick, 2011). A sufficient number of foils are used to reduce the probability that an innocent examinee is wrongly accused (Krapohl, 2010).

absence of strategic and nonstrategic cues of deception to identify concealed information.

5.2 Strategic Cues of Deception and Concealed Information

A primary objective of deception researchers is to identify a cue, or cues, that can be used to accurately discriminate between truth tellers and deceivers. Some cues have exhibited promising results; however, these findings are based on laboratory research, meaning that the generalizability of these findings is limited across contexts, cultures, communicators, and communication modalities. Several meta-analyses of deception research have been conducted in an attempt to identify cues of deception that persist across laboratory studies; identifying a cue possessing discriminatory power across several studies would demonstrate generalizability and thus indications of promise for use in less-controlled environments (e.g., actual criminal investigations).

DePaulo and colleagues (2003) reviewed 158 cues from an extensive corpus of deception studies and found that “many behaviors showed no discernible links, or only weak links, to deceit” (2003, p. 74). Hartwig and Bond also conducted a meta-analysis of deception research and concluded that the validity of any behavioral cue or cues of deception has yet to be established (Hartwig & Bond Jr., 2011). While these meta-analyses provide evidence that a robust cue of deception has yet to be identified, it is possible that the act of deceiving or concealing information may be identifiable by monitoring strategic and nonstrategic behaviors that deceivers or concealers display in the attempt to *appear innocent*.

5.2.1 Interpersonal Deception Theory (IDT)

The lion’s share of deception research has examined communication dyads in which one participant is tasked with deceiving the other. These interactions are typically brief, lack realism, lack any sense of jeopardy if the deceiver is caught, and are thus limited in their generalizability (Miller & Stiff, 1993). To instigate deception researchers to conduct experiments in more interactive, dynamic, and thus generalizable contexts, Buller and Burgoon developed Interpersonal Deception Theory (IDT) (Buller & Burgoon, 1996). IDT is comprised of 18 propositions that can be used as a framework to conduct more realistic deception research. A key aspect of IDT is the contention that deceivers act strategically to avoid detection while leaking nonstrategic cues and behaviors due to arousal and an inability to control all of their behaviors and emotions (Buller & Burgoon, 1994). Deceivers attempt to minimize leakage as much as possible (Zuckerman et al., 1981) by managing self-presentation (DePaulo, 1992) and other behaviors that they associate with deception (Hocking & Leathers, 1980).

Buller and Burgoon state that "Once senders decide to deceive, they must also be concerned about appearing credible, allaying receiver suspicions, minimizing their responsibility for deceit, and avoiding unpleasant consequences if deception is detected" (Buller & Burgoon, 1996, p. 216). Buller and Burgoon (1996) propose that there exist differences in the behavior of the deceiver and the recipient. These differences are attributable to the deceiver attempting to "manage" his or deception by completing each of the aforementioned objectives. Part A of Proposition 3 in IDT states the following: "Compared with truth tellers, deceivers (a) engage in greater strategic activity designed to manage information, behavior, and image and (b) display more nonstrategic arousal cues, negative and dampened affect, noninvolvement, and performance decrements" (Buller & Burgoon, 1996, p. 218). It is important to note that there is a key distinction between the strategic and nonstrategic behaviors cited in Proposition 3. Strategic behaviors are inherently controllable as the deceiver can choose to utilize any number of countermeasures to appear innocent; however, nonstrategic cues can include behaviors that are either controllable or uncontrollable by the deceiver. For example, the strategic behavior of feigning a smile can be controlled by the deceiver; however, cues of arousal (e.g., increased pupil dilation) are autonomic responses that cannot be controlled consciously. It is important to note that as a communicator works harder to appear innocent, performance decrements will increase (Berger, Karol, & Jordan, 1989).

5.2.2 Strategic Behaviors in a CIT Context

The findings of two CIT studies suggest that individuals concealing information may exhibit strategic and nonstrategic cues associated with the act of concealing information (Lubow & Fein, 1996; Proudfoot et al., 2013). The results presented in Chapter 4 of this dissertation indicate that there are main effect differences between individuals concealing information relative to members of the control group. These differences were found utilizing a visual CIT in which stimuli were presented in groups of 4, with a stimulus centered in each of the four quadrants on the screen. Specifically, these differences are as follows:

1. Concealers fixated on the center of the screen longer than members of the control group.
2. Concealers exhibited increased pupil dilation for the duration of the interview relative to members of the control group.
3. Concealers experienced longer vocal response latencies relative to members of the control group[65].

---

[65] Researchers investigating differences in response times between deceivers and truth tellers have claimed that this measure should not be used for diagnostic purposes as it can easily be manipulated by the interviewee (Farwell & Donchin, 1991) (as with respiration in the CQT). However, a study

The tendency for concealers to fixate on the center of the screen is a strategic behavior employed to avoid viewing target items as they are aversive in nature (Twyman, 2012). The tendency for concealers to exhibit longer response latencies can be attributed to the cognitive taxation associated with acting strategically to avoid detection[66]. The persisting increase in pupil dilation can be attributed to an overall increased level of arousal in participants concealing information due to the anxiety associated with being detected (Ekman & Friesen, 1969, 1974; Zuckerman et al., 1981).

Upon recognizing these differences, an immediate consideration was the evaluation of these three measures in a CIT containing *no* target items. A variation of the CIT in which no target items are included appears to be a natural progression from the Searching CIT, which is also conducted without the use of pre-specified target items. The key distinction between this proposed method and the Searching CIT is that the Searching CIT is used to identify recognition and familiarity with information that *may* be crime relevant, while this adaptation of the CIT would be used to identify strategic and nonstrategic cues exhibited by an individual attempting to conceal information.

Upon formulating this novel approach, a review of relevant literature was conducted to determine if this CIT approach had been tested previously. This search yielded the identification of a mock-crime study in which pupillary responses (PR) were compared to EDA in a visual CIT (Lubow & Fein, 1996). A key finding of this study was that participants guilty of committing the theft exhibited increased pupil

---

investigating the influence of countermeasures on vocal response times reported that the use of countermeasures had a marginal impact (Seymour, Seifert, Shafto, & Mosmann, 2000).

[66] For a more thorough explanation of both of these differences, refer to the discussion section in Chapter 4.

dilation for target stimuli, control stimuli, and the presentation of the fixation cross. Lubow and Fein (1996) attribute this main effect difference to "…a generally elevated level of arousal in the guilty as compared to the innocent group" (Lubow & Fein, 1996, p. 175). This finding was replicated by the study presented in Chapter 4. The presence of a persisting increase in pupil dilation for the duration of a CIT is congruent with Part B of Proposition 3 in IDT as it can be classified as a nonstrategic arousal cue.

5.3 Summary / Research Question 2

There exist two methods that can be used to conduct a CIT: the Known Solutions CIT and the Searching CIT. Both of these methods hinge on the identification of physiological variations associated with verified or assumed target items. CIT researchers have established a general set of guidelines that should be used to identify salient target items; however, these guidelines can be difficult to adhere to given the nature of many crimes and criminal investigations. Subsequently, CITs can be infeasible for use or prone to false negative classifications if salient target items cannot be identified. A CIT method that does not require the incorporation or identification of target items may prove useful if alternative methods of identifying concealed information can be identified. IDT posits that deceiving in interpersonal communication is a dynamic process in which deceivers likely exhibit strategic and nonstrategic cues resulting from their deception. Two CIT studies have identified the presence of strategic and nonstrategic cues exhibited by individuals concealing information, thus, it may be possible to use a targetless CIT to identify strategic and nonstrategic cues indicative of concealed information. The following research question is proposed.

RQ2: Can a targetless CIT be used to elicit strategic and nonstrategic cues associated with concealing information?

The following chapter reports the findings of a study designed to investigate RQ2.

6 STUDY 2 PART I: USING A TARGETLESS CIT TO ELICIT STRATEGIC AND NONSTRATEGIC CUES OF CONCEALED INFORMATION

6.1 Introduction

The purpose of this study is to determine if a visual CIT designed *without* target items can be used to elicit strategic and nonstrategic cues associated with concealing information. It was expected that examinees concealing information while completing a CIT *without* target items would still exhibit increased pupil dilation, eye gaze fixation on the center of the screen, and increased vocalic response latencies during a visual CIT. An experiment was designed to compare the oculometric and vocalic behaviors of participants assigned to one of three conditions: the no-targets group, the targets-group, and the control group[67]. In addition to identifying the presence of behavioral and physiological differences between conditions, it was important to determine *how long* differences between the manipulation groups and the control group persisted[68].

6.2 Literature Synopsis and Hypotheses

A wealth of CIT research has been conducted over the course of several decades (Verschuere, Ben-Shakhar, et al., 2011). A key area of this research is evaluating the types of sensors that can be used to accurately and reliably measure the orienting response (Twyman, Burgoon, et al., 2013). Traditional sensors used to conduct CITs were adopted from the standard sensors utilized for polygraph

[67] The 'targets group' and 'no-targets' group will be referred to as the 'manipulation groups' when referenced together as both groups completed the experimental manipulation of concealing information during the screening interview.

[68] Post-measures administered after the screening interview were analyzed to identify differences in the perceptions, self-reported stress and arousal, perceived performance, and countermeasures reported by members of each condition. Post-measures results are presented in Chapter 7.

interviews; EDA serves as the most commonly-used measure for CITs (Krapohl et al., 2009). Recent research has focused on harnessing new sensor technologies to replace or augment traditional sensors. A specific area of interest is identifying sensors that can be used for a CIT without contacting the examinee (Nunamaker Jr., Burgoon, et al., 2012; Proudfoot et al., 2013; Twyman, Burgoon, et al., 2013). The findings presented in Chapter 4 of this dissertation support prior work suggesting that eye tracking and vocalics are two promising noncontact methods that can be used to identify concealed information.

An interesting finding of this research is the identification of main effect differences between concealers and members of a control group. When completing a visual CIT, concealers exhibited increased pupil dilation, a tendency to fixate on the center of the screen, and increased response latency for the duration of the CIT, irrespective of the presence or absence of target items. Deception literature supports the presence of these behaviors, as deceivers tend to exhibit strategic and subsequent nonstrategic cues associated with their deception. These findings motivate the current study designed to determine if main-effect differences are exhibited by concealers when completing a CIT *lacking* target items. It is hypothesized that concealers will exhibit these same main-effect differences when completing a targetless CIT. The following hypotheses are proposed:

H1: Participants concealing information will fixate longer on the center of the screen relative to members of the control group when completing a targetless CIT.

H2: Participants concealing information will exhibit increased pupil dilation relative to members of the control group when completing a targetless CIT.

H3: Participants concealing information will exhibit longer vocalic response latencies relative to members of the control group when completing a targetless CIT.

The behaviors referenced in H1, H2, and H3 are based on findings from CIT research in which target items *were* used. While these behaviors persisted for the duration of the interview despite the presence of target items on only 5 of the 25 slides, a total absence of target items for the entire duration of a CIT may reduce the presence of strategic and nonstrategic behaviors exhibited by concealers over time. It is likely that as concealers recognize that they are not being confronted or tested on the information that they are attempting to conceal, they will gradually reduce efforts to avoid detection. The following hypotheses are proposed:

H4: Oculometric differences between participants concealing information and members of the control group completing a targetless CIT will diminish over time.

H5: Vocalic differences between participants concealing information and members of the control group completing a targetless CIT will diminish over time.

6.3 Deception Experiment

A deception experiment was conducted to test the hypotheses presented in the previous section. This experiment was patterned after the experiment presented in Chapter 4 but with a number of modifications that will be discussed in the following sections. For this study, participants were randomly assigned to one of three possible conditions: the no-targets group, the targets group[69], or the control group. Members of the no-targets group possessed guilty knowledge; however, the CIT that they completed during the experiment did not contain target items. Members of the control group had no guilty knowledge and also completed a CIT lacking target items. Data collected from members of the control group were used to ensure that behaviors exhibited by participants concealing information were attributable to the manipulation. Members of the targets group possessed guilty knowledge and completed a CIT *containing* target items. Data from this condition can be used to identify differences between the no-targets group and the targets group. A detailed description of the methodology used for this study is described in the following sections.

6.4 Methodology

6.4.1 Participants

Undergraduate students (N=116) were recruited from business courses at the University of Arizona to participate in this experiment. Virtually all (98%) of participants were college juniors and seniors. The average age of all participants was 21.2 years. Roughly half (57%) of participants were male and 70% of the participants were U.S citizens. The remainder of the sample was represented by a variety of

---

[69] Members of the targets group completed a CIT containing target items. This group was the focus of the research presented in Chapter 4. However, in this study, they served as a form of control group to make comparisons with participants of the no-targets group.

nationalities, including 12% Mexican, 10% Asian, 1% American Indian, 1% Pacific Islander, and 7% other. A majority (88%) of the sample was comprised of native English speakers. All participants reported having no prior experience in law enforcement, criminal investigations, or credibility assessment interviews. Only one participant reported hearing information about the experiment prior to participating; data for this participant was not considered for analysis.

6.4.2 Design and Procedures

Participants arrived at the experiment location and were randomly assigned to one of three possible conditions: the no-targets group, the targets group, and the control group. Persons willing to participate reviewed and signed consent forms informing them that they would complete an automated screening interview. Consent forms for members of the manipulation groups included information that they would pack sensitive materials and conceal this fact during the interview. Upon completing the consent form, each participant received additional instructions and proceeded to the next phase of the experiment.

At this point, each participant completed a pre-survey and packed a bag with benign items[70]. Members of the manipulation groups completed the pre-survey and packed the bag with benign items; however, they were also tasked with completing a second pre-survey in which they learned about their association with the Sinaloa drug cartel. Members of these two conditions learned that they had to conceal a fake IED[71]

---

[70] The bag was packed with numerous articles of clothing and a camera case.

[71] In the experiment described in Chapter 4, the banned item used was a fake box of ammunition. Ammunition was selected due to the frequency with which smugglers actually traffic guns and ammunition across the border, thus, it was deemed ecologically valid. However, it was determined that packing the IED is a more uncomfortable and stressful task, and likely fosters feelings resembling emotions experienced by someone actually trafficking contraband across the border.

inside of the bag[72] and deliver it to one of three individuals waiting in the atrium of the business building. In order to deliver the IED to one of these individuals, they had to memorize three different faces of criminal associates. To ensure that each participant in the manipulation groups could remember the faces, they had to complete an exercise in which they listed several facial features that they would later use to identify the recipients. The three faces, the IED, and the Sinaloa cartel name served as the five target items used for the visual CIT. Members of all conditions were incentivized to take the task seriously as they were informed that they would receive a monetary reward for avoiding suspicion during the interview.

An image depicting the location where each participant completed the pre-survey and packed the bag is provided in Figure 25. An image of the IED that was packed by members of the manipulation groups is provided in Figure 26.

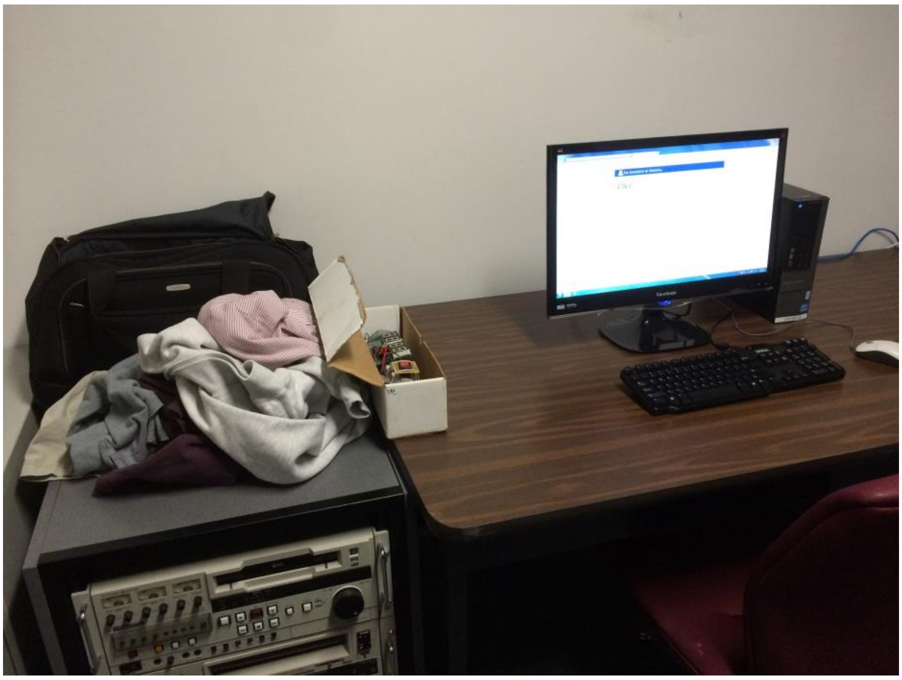


Figure 25 - Location of Pre-Survey and Bag

[72] It was interesting to see how each participant packed the IED inside of the bag. Some participants simply stuck the box containing the IED under a shirt; others carefully wrapped the box in multiple layers of clothing. A handful of participants hid the IED inside of the zipped liner at the bottom of the bag.

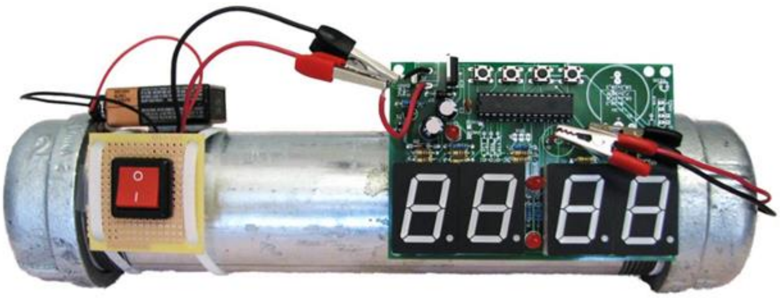


Figure 26 - Fake IED

Upon completing the pre-survey and packing the bag, all participants were directed to the interviewing room. Participants were asked to place the bag on the floor next to an Automated Screening Kiosk (ASK) and step onto a force platform located on the floor in front of the system. The force platform was not used for data collection during this experiment, but it served an important function as it was used to standardize the location where each examinee stood during the interview[73]. The ASK system is comprised of a desktop tower and a retractable arm; a monitor, speakers, two webcams, an eye tracker, a microphone, and LED lights are mounted on the retractable arm[74]. The height of the sensor array was configured based on the height of the examinee by adjusting the retractable arm. Upon successfully configuring the arm, each participant received instructions on how to proceed[75]. An image of the ASK system is provided in Figure 27. A close-up shot of the monitor and sensor array is provided in Figure 28.

[73] The force platform has decals of footprints on each of the two plates. Participants would step directly onto the footprints, even without explicit direction from the experiment facilitator.

[74] This grouping of devices will be referred to as the 'sensor array' from this point forward.

[75] Participants were informed that they would first complete a calibration task with the eye tracker and then receive further instructions on how to proceed.

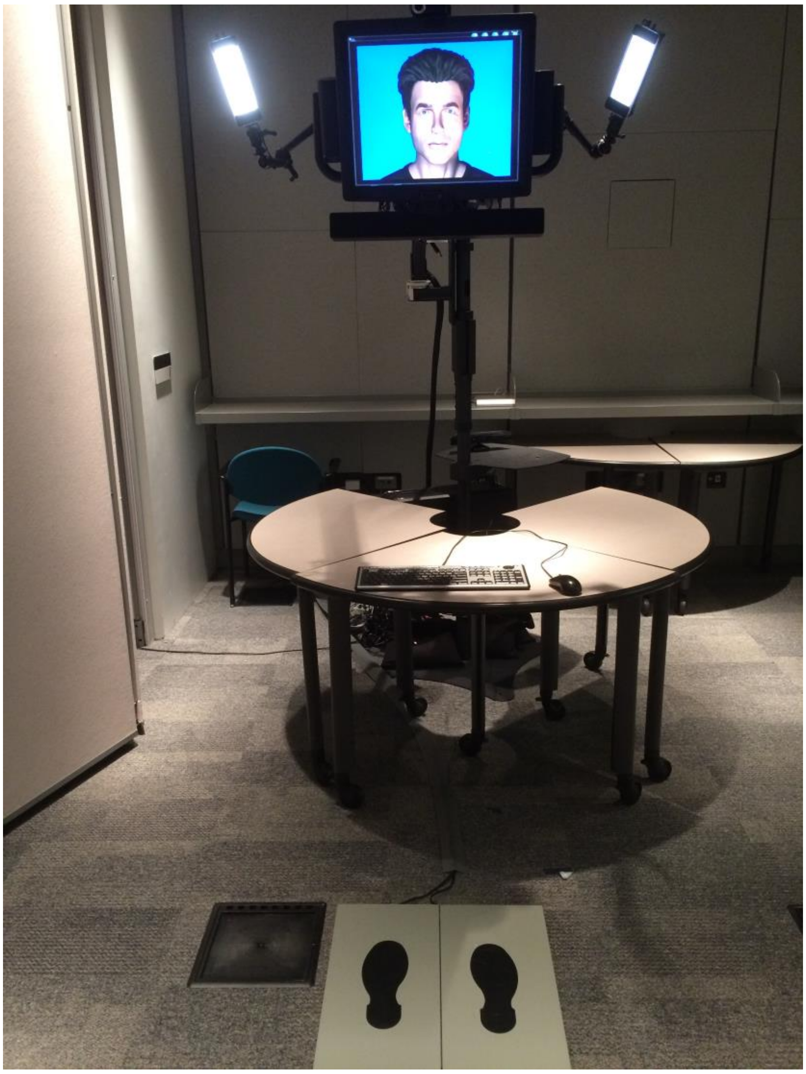


Figure 27 - Automated Screening Kiosk (ASK)

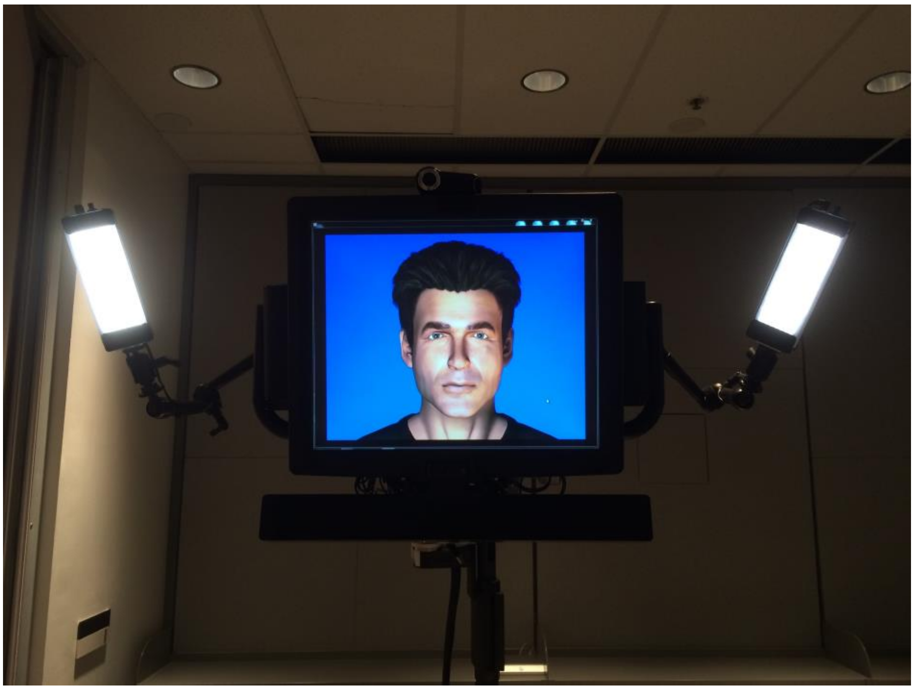


Figure 28 - ASK Sensor Array

Next, participants completed an automated calibration task to calibrate the eye tracker. The system presented nine yellow circles in various locations on the screen and recorded eye movements while participants viewed these circles[76]. When calibration was complete, the system initiated the preliminary interviewing phase.

During the preliminary interview, the ASK provided information about the nature of the interview, defined the purpose of the CIT, listed the sensors that would be used, and completed a practice test to ensure that each participant understood the proper protocol for completing the interview. This preliminary phase was designed to mimic the pretest interview conducted prior to a traditional CIT, during which the examiner talks about the nature of the test, the testing protocol, the sensors that will be used, and so forth (Krapohl et al., 2009). Having instructions presented by the ASK avoided any inadvertent experimenter expectancy effects that could occur with human

[76] Calibration is used to ensure that oculometric differences attributable to variations in ocular features between participants are minimized.

research assistants. The full set of instructions provided to examinees is enumerated in the following list.

1. During this interview you will complete a Concealed Information Test.
2. The Concealed Information Test is designed to identify familiarity and recognition.
3. During this test you will view several images of faces, banned items, and names of criminal organizations. While you view these images, this system will analyze your eye movements and speech.
4. If you are involved with any of these people or organizations, or if you are carrying any items similar to the items displayed, the system will be able to identify this.
5. During the test, an Embodied Conversation Agent (ECA) will ask you a question, then display an image. Please respond by answering “Yes” or “No” out loud. In total, you will view 25 images; the interview will last approximately 4 minutes[77].
6. You will now view a practice set of images to make sure you understand the procedures[78].
7. The screening interview will now begin…Remember to respond “Yes” or “No” out loud. Do not look off of the screen at any time.

---

[77] A small number of participants had difficulty during initial interactions with the ASK. They did not realize that they had to answer out loud, or they would think that the system did not hear them, so they would repeat their answer multiple times. The use of a set of practice images ensured that all participants were familiar with the appropriate protocol prior to beginning the CIT.

[78] Five images of faces were displayed during the practice test. These images were patterned after the face images used during the actual interview, but did not contain any of the face images that were used for the main interview.

An example of one of the images used to display information during the practice test is provided in Figure 29.

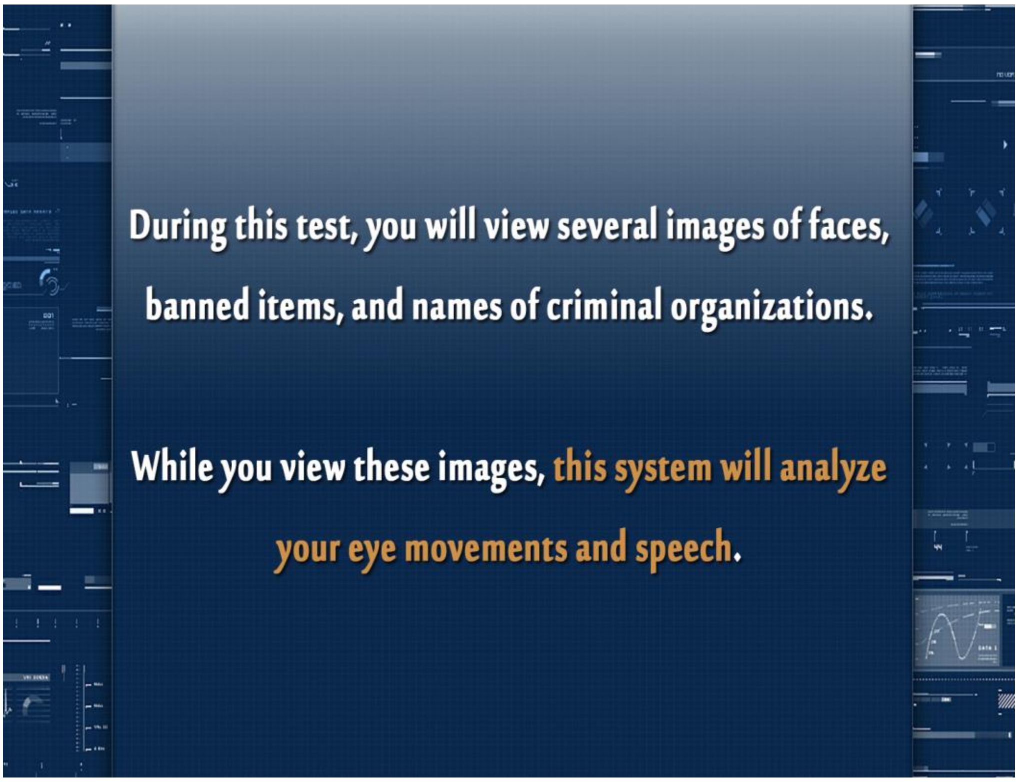


Figure 29 - Image Used during Preliminary Interview Phase

After completing the preliminary interview, the CIT portion of the interview commenced. During the CIT, each participant viewed 25 slides. These images were divided into 5 foils comprised of 5 slides each. Each slide was comprised of four images; each image was centered in one of the four quadrants on the slide[79]. The first three foils contained images of faces, while the fourth and fifth foils contained images of banned items and the names of criminal organizations, respectively[80]. Slides

[79] The locations of the images on each slide were rotated to ensure that no location-based or ordering effects influenced the data.

[80] The corpus of images used for this experiment is the same corpus of images used for the study described in Chapter 4. For additional information concerning pilot testing of images, as well as examples of the slides that were used, please refer to the methodology section in Chapter 4 and Appendix A.

containing target items appeared in different locations within each foil and target items appeared in different locations on each slide; this was implemented to mitigate any location-based effects. Prior to the presentation of each slide, an ECA asked each participant if he or she was familiar with any of the images on the following slide. A fixation cross was displayed for 500ms prior to the presentation of each slide; slides were displayed for 7500ms. Participants were expected to answer by saying “Yes” or “No” out loud. The experiment facilitator sat on the opposite side of the retractable wall visible on the left side of Figure 27. During the interview, the facilitator monitored each participant’s responses to ensure that he or she did not confess. A panoramic view of the room in which interviews were conducted, including a view of the facilitator’s setup relative to the participant’s location, is provided in Figure 30.

Figure 30 - Panoramic View of Automated Interview Location

When the automated screening interview was complete, the facilitator informed members of the manipulation groups that they would not be delivering the IED to the atrium. The facilitator then conducted a manipulation check to ensure that participants tasked with learning target items prior to the screening could still recall them. This was accomplished by displaying images to the participant using an iPad[81]

[81] In the experiment presented in Chapter 4, the manipulation check was conducted as part of the post-measures. However, it was apparent from the responses collected during that experiment that some participants thought they were still tasked with deceiving, and intentionally picked the wrong faces, while others appeared to be lazy or confused and did not pick any faces at all. Manually conducting the manipulation check alleviated each of these issues.

and requesting that they vocally identify target items. When the manipulation check was completed, participants completed post-measures. Post-measures were designed to collect information about each participant's general perceptions of the interview, emotional responses during the interview, perceived performance, and the use of countermeasures. Analyses of data collected by the post-measures are discussed in a following section.

6.4.3 Instrumentation

Oculometric and vocalic data were recorded using the EyeTech™ Digital Systems VT2 infrared eye tracker and a noise-cancelling Andrea™ array microphone, respectively. For more information concerning the features or data-collection configurations of either device, refer to the methodology section in Chapter 4.

6.5 Analysis and Results of Hypothesis Tests[82]

In study one, hypothesis tests warranted the use of multilevel regression models as behaviors were compared both within and between subjects. The need to conduct within-subject comparisons was associated with the need to identify relative variations in behavior due to the presence or lack of target items. The purpose of this study is to identify persisting differences in behaviors between conditions throughout the entire interview; thus, ANOVAs were used to identify mean differences between conditions.

6.5.1 Manipulation Checks

Participants in the manipulation groups were expected to exhibit strategic and subsequent nonstrategic behaviors in an effort to appear truthful and avoid suspicion. Participants in each condition were incentivized to take the task seriously as they were

---

[82] Each hypothesis test was conducted using an ANOVA. To mitigate any limitations of this approach considering the repeated-measures nature of the data, hypothesis tests were also conducted using repeated-measures ANOVAs with maximum likelihood estimation. This alternate approach did not yield any noteworthy findings.

informed that examinees appearing truthful during the interview would receive an additional monetary reward. Participants were also warned that examinees appearing suspicious would be directed to a lengthier screening interview. To ensure that members of the manipulation groups took the task seriously, measures assessing their (1) effort to appear truthful, (2) motivation to appear truthful, and (3) confidence that they did appear truthful were formulated. The questions presented in Table 15 were used to conduct this manipulation check.

Table 15 – Manipulation Check Questions

#	Question
1	During the interview, how important was it to you to succeed in making the interviewer believe you?
2	During the interview, how important was it to you to give convincing answers?
3	How hard did you try to convince the interviewer that you were telling the truth?
4	How hard did you try to avoid suspicion during the interview?
5	How successful do you think you were in convincing the interviewer that you were truthful?

Participants were asked to rate each question using a five-point Likert scale. For questions 1 and 2, response options ranged from 1="Not at all important" to 5="Extremely important". For questions 3 and 4, response options ranged from 1="Not at all" to 5="Very hard". For question 5, response options ranged from 1="Very unsuccessful" to 5="Very successful". Table 16 contains a summary of descriptive statistics by condition for each question.

Table 16 - Manipulation Check Descriptive Statistics

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Importance of Interviewer Belief								
Control	3.39	1.37	4	1	5	4	-0.621	-0.59
No Targets	4.23	0.8	4	1	5	4	-1.74	5.55
Targets	4.29	0.77	4	2	5	3	-0.94	0.67
Importance of Convincing Answers								
Control	3.45	1.28	4	1	5	4	-0.66	-0.47
No Targets	4.2	0.77	4	1	5	4	-1.68	5.87
Targets	4.39	0.59	4	3	5	2	-0.38	-0.64
Attempt to Convince Interviewer								
Control	2.12	1.32	2	1	5	4	1.07	0.05
No Targets	3.16	1.17	3	1	5	4	-0.43	-0.34
Targets	3.74	1.11	4	1	5	4	-0.58	0.04
Attempt to Avoid Suspicion								
Control	3.06	1.09	3	1	5	4	-0.44	-0.9
No Targets	3.69	0.99	4	1	5	4	-0.58	0.07
Targets	4.13	0.96	4	1	5	4	-1.23	1.75
Perceived Success								
Control	4.45	0.62	5	3	5	2	-0.67	-0.43
No Targets	4.14	0.96	4	1	5	4	-1.29	1.75
Targets	3.63	1.1	4	1	5	4	-0.74	0.07

A review of the values in Table 16 indicates that participants in the manipulation groups completed the task with a sufficient level of motivation to warrant hypothesis testing. A more thorough analysis of these measures, including significance tests and explanations for possible variations between groups, is presented in Chapter 7.

An additional manipulation check was conducted to ensure that participants in the targets group could remember the target items that they had become familiar with prior to beginning the automated interview. A majority of participants (75%) could recall all target items; almost all participants (95%) could recall 4 or more target items.

6.5.2 Oculometrics[83]

6.5.2.1 Oculometric Threat Avoidance: Viewing the Center of the Screen

Eye tracking data for 91 participants was used for this analysis; a number of data points were removed due to calibration issues or poor data quality. Oculometric data were processed to yield the percentage of time that a participant was viewing the center of the screen per slide. These data were submitted to an ANOVA to identify significant differences between the mean duration of time that participants in each condition viewed the center of the screen. Members of the control group viewed the center of the screen 11.4% of the time. There was not a significant difference between the means of the control group and the targets group ($t(1693) = -0.39$, $p = .69$). However, there was a significant difference between the control group and the no-targets group ($t(1693) = 2.54$, $p = .011$); members of the no targets group viewed the center of the screen on average 2.9% longer than members of the control group. A Bonferroni pairwise t-test indicated that there was a significant difference between manipulation groups for this measure ($p = .009$). These statistical tests suggest that members of the no-targets group viewed the center of the screen longer than members of the control group; however, the tendency to view the center of the screen for members of the targets group more closely matched members of the control group.

6.5.2.2 Pupil Dilation: Differences between Conditions

Pupil dilation data for 89 participants was used for this analysis; a number of data points were removed due to calibration issues or poor data quality. The eye tracker used for data collection calculates pupil dilation for each pupil as well as an average pupil dilation value. Pupil dilation data constituted of values for both eyes

[83] An initial review of eye tracking data indicated that there may have been issues with calibration and tracking accuracy for the vast majority of participants. Problematic data points were removed in an effort to create a data set that could still be used to yield accurate findings. Pupil dilation data appeared to be accurate.

averaged together were submitted to an ANOVA to identify significant differences in mean pupil dilation between participants in each condition. On average, members of the control group exhibited an average pupil dilation of 4.04 mm. There was not a significant difference between the means of the control group and the no-targets group ($t(2233) = -0.97, p = .33$). However, there was a significant difference between the control group and the targets group ($t(2233) = 2.55, p = .0108$). Members of the targets group exhibited an average increase of pupil dilation of .084 mm or 2% relative to members of the control group. A Bonferroni pairwise t-test indicated that there was a significant difference between manipulation groups for this scale ($p = .0012$).

6.5.3 Vocalics

6.5.3.1 Response Latency

Vocalic data for 110 participants were used for this analysis; a number of data points were removed due to incomplete data sets for several participants. Raw vocalics data were processed and response latency values (in seconds) were generated. Mean response latency values were submitted to an ANOVA to identify significant differences between participants in each condition. Members of the control group exhibited an average response latency of 1.81 seconds. There were no significant differences in response latency between the manipulation groups and the control group ($F(2, 2747) = 0.28, p = .76$). Additionally, there was no significant difference in means between the manipulation groups.

6.5.4 Temporal Effects

It was hypothesized that significant differences between the no-targets group and the control group would diminish over time as examinees would recognize that they were not being asked about their recognition of crime-relevant information, thus, any strategic behaviors that they had been employing to avoid detection would be curtailed after the first several foils. The following analyses were used to investigate any temporal effects on differences between groups.

6.5.4.1 Temporal Effects on Oculometrics

Significance tests on data for viewing the center of the screen revealed that there was a significant difference between the no-targets group and the control group for the duration of the interview; however, it was anticipated that there may have been a larger effect during earlier stages of the interview. An ANOVA on oculometric data for the first foil only revealed that there was not a significant difference in viewing the center of the screen between the no-targets group and the control group ($t(452) = 0.90$, $p = 0.37$) for this subset of the data. Since no significant difference was found for data constituting the first foil, this fails to support the hypothesis that differences between groups would diminish over time. However, due to an overall significant difference between the no-targets group and the control group for this test, these differences must have occurred at other points during the interview.

To identify the portion of the interview during which the significant difference did occur, mean values for viewing the center of the screen were plotted for each condition by question. Refer to Figure 31 for this visualization. To identify variations attributable to each of the four question blocks of the interview (baseline questions, face images, banned-object images, and criminal organizations), a plot of the mean values for viewing the center of the screen for each condition by question block is presented in Figure 32.

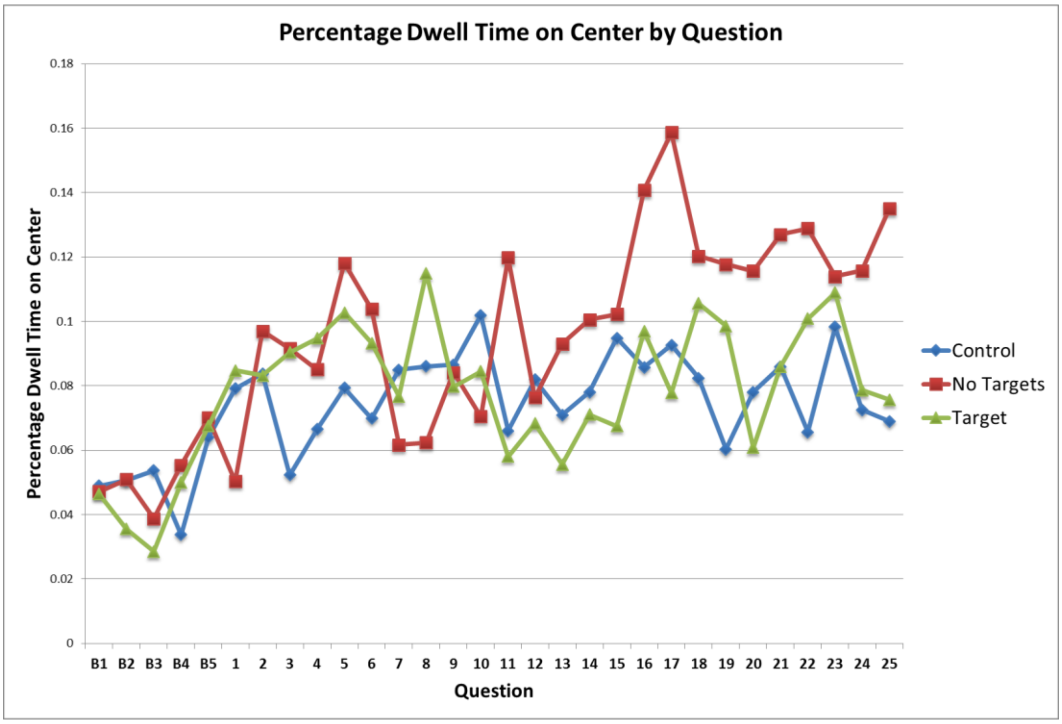


Figure 31 - Percentage Dwell Time on Center by Question

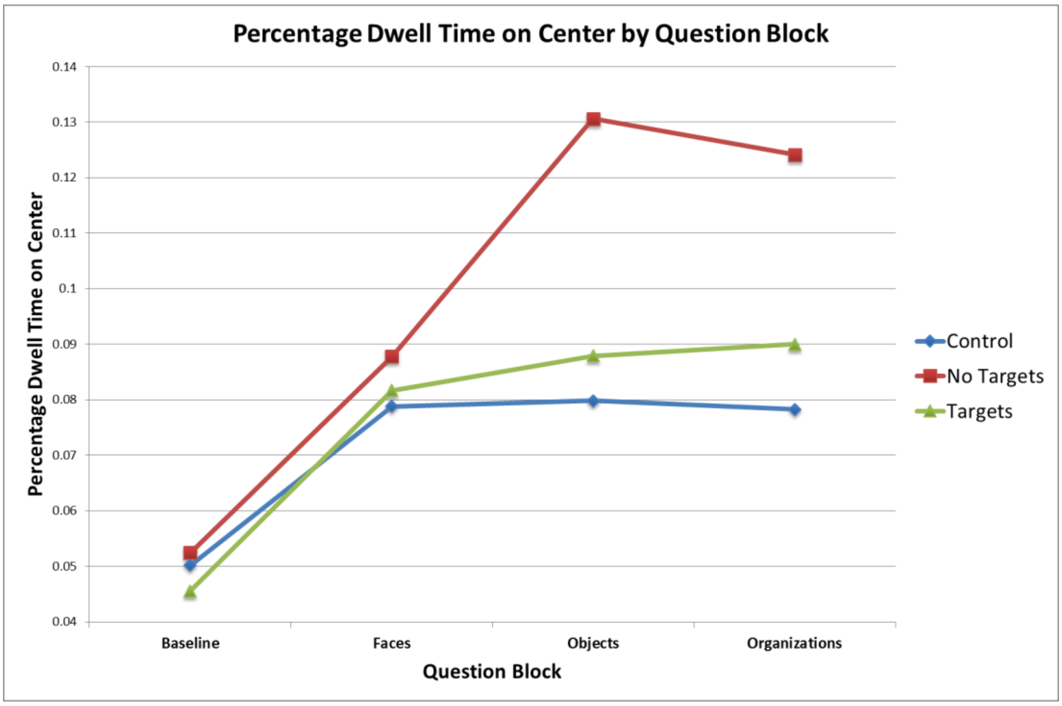


Figure 32 - Percentage Dwell Time on Center by Question Block

It is evident in Figure 31, and clear in Figure 32, that the overall significant difference in viewing the center of the screen is associated with the third and fourth block of interview questions, the banned-objects and criminal organizations blocks, respectively. This difference may be attributable to members of the no-targets group disengaging from the interview as they had viewed three foils of images and had not seen a single image that was relevant to the crime that they were committing.

Significance tests on data for pupil dilation revealed that there was a significant difference between the targets group and the control group for the duration of the interview; however, it was anticipated that there might have been a significant difference between the no-targets group and the control group during earlier stages of the interview. An ANOVA on pupil dilation data for the first foil only revealed that there was not a significant difference between the no-targets group and the control group ($t(442) = -0.14$, $p = 0.89$). Since no significant difference was found for data constituting the first foil, this fails to support the hypothesis that differences between groups would diminish over time. However, this analysis did identify that the significant difference between the targets group and the control group was present during the first foil of interview questions ($t(442) = 2.13$, $p < .034$).

To further investigate differences in pupil dilation between conditions for the duration of the interview, mean values for pupil dilation were plotted for each condition by question. Refer to Figure 33 for this visualization[84]. To identify variations attributable to each of the four question blocks of the interview (baseline questions, face images, banned-object images, and criminal organizations) a plot of the mean values for pupil dilation for each condition by question block is presented in Figure 34.

[84] This figure includes pupil dilation data for the CIT pre-test overview occurring at the very beginning of the interview. X-axis labels ‘PT’ denote this portion of the interview.

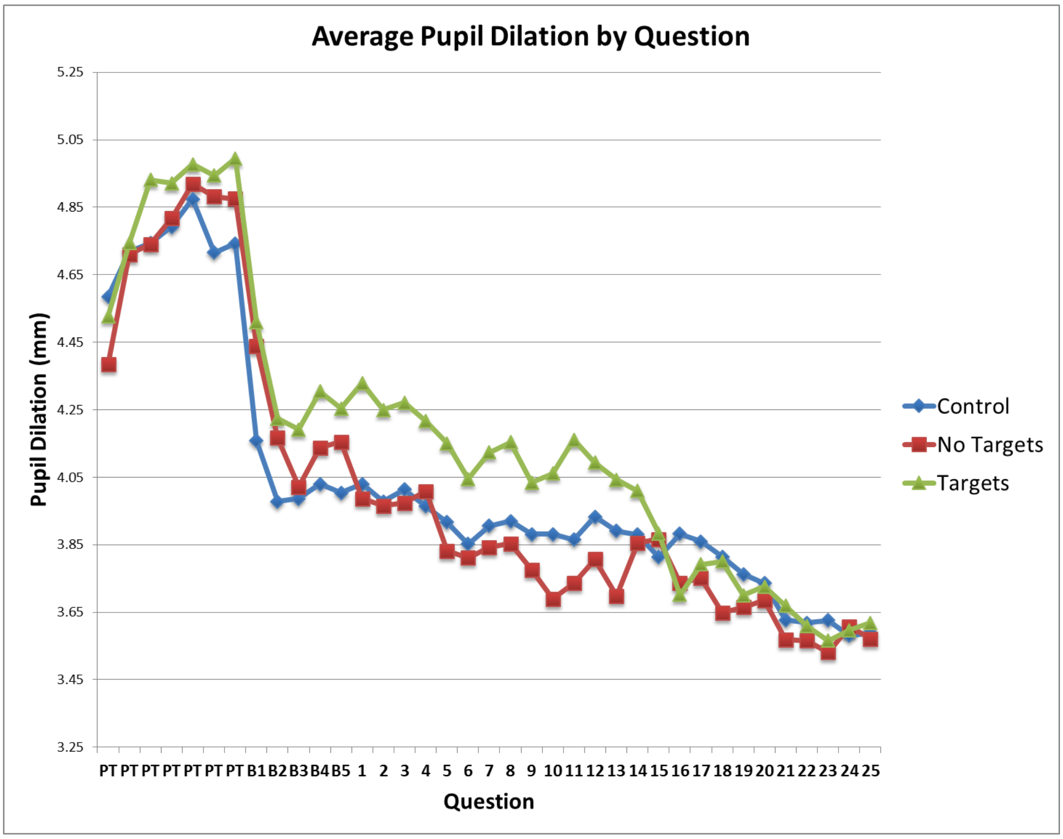


Figure 33 - Average Pupil Dilation by Question

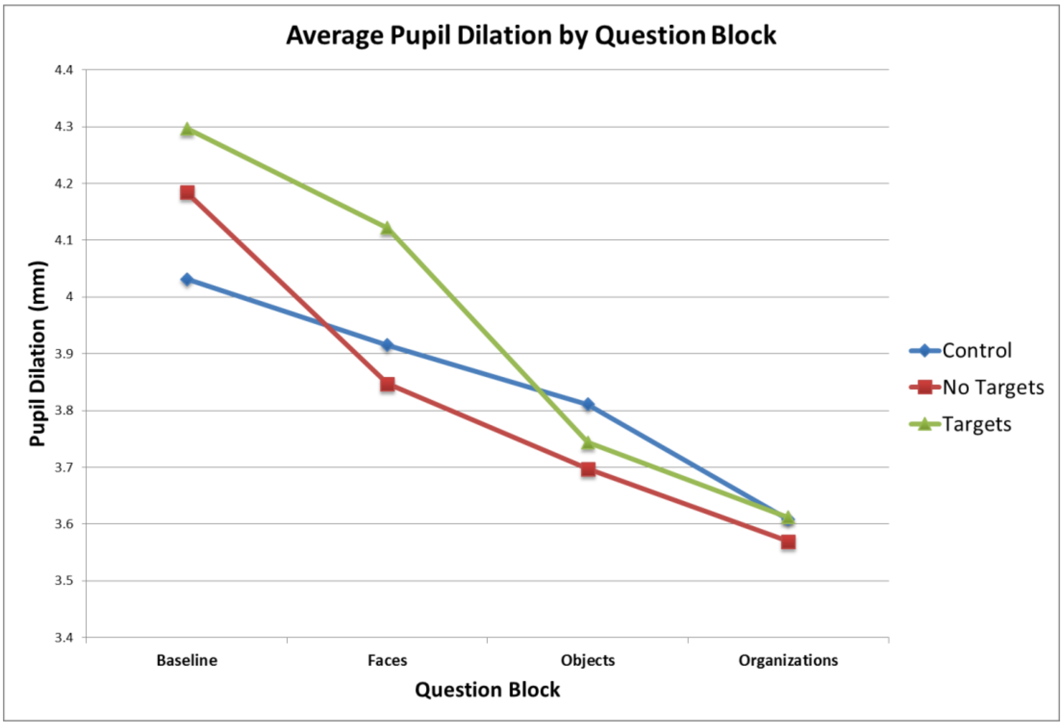


Figure 34 - Average Pupil Dilation by Question Block

It is clear in Figure 33 and Figure 34 that the average pupil dilation of participants in the targets group is larger than the other two conditions for baseline questions and faces questions. However, upon answering questions associated with banned items and criminal organizations, average pupil dilation for members of the targets group falls below members of the control group. All three conditions showed a decrease in pupil dilation over time.

6.5.4.2 Temporal Effects on Vocalics

Significance tests on response latency data revealed that there were no significant differences between the no-targets group and the control group for the duration of the interview; however, it was anticipated that there might have been significant differences during early stages of the interview. An ANOVA was conducted on mean response latency between the control group and the no-targets group using data for the first foil only. This analysis found there to be no significant difference between the no-targets group and the control group ($t(547) = 0.13$, $p = 0.91$).

To identify any differences in response latency between conditions for the duration of the interview, mean values for response latency were plotted for each condition by question. Refer to Figure 35 for this visualization. To identify variations attributable to each of the four question blocks of the interview (baseline questions, face images, banned-object images, and criminal organizations) a plot of the mean values for response latency for each condition by question block is presented in Figure 36.

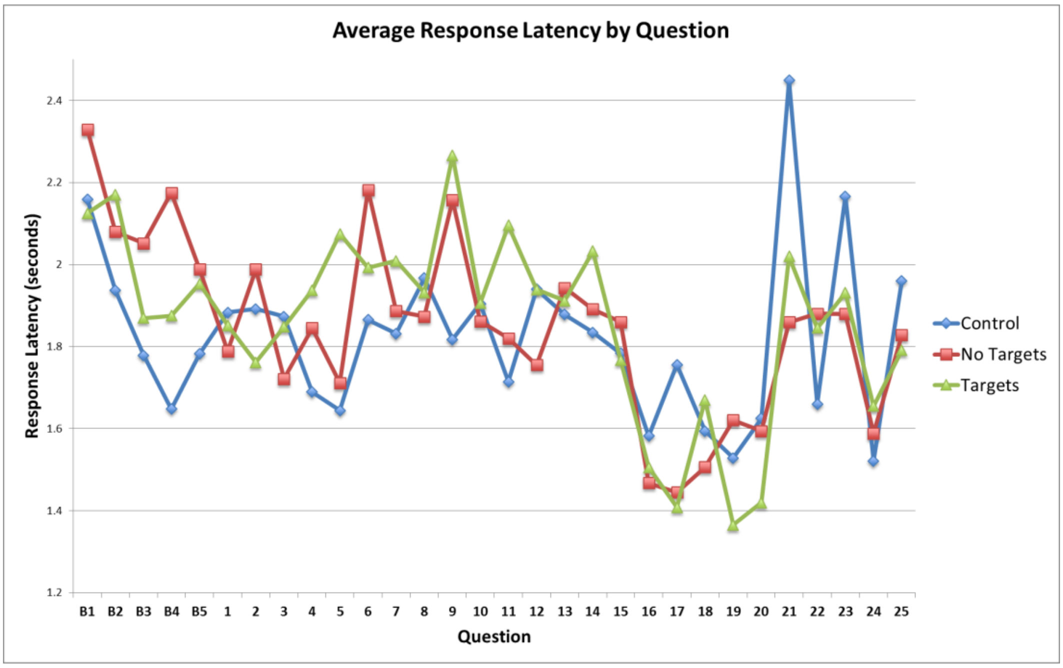


Figure 35 - Average Response Latency by Question

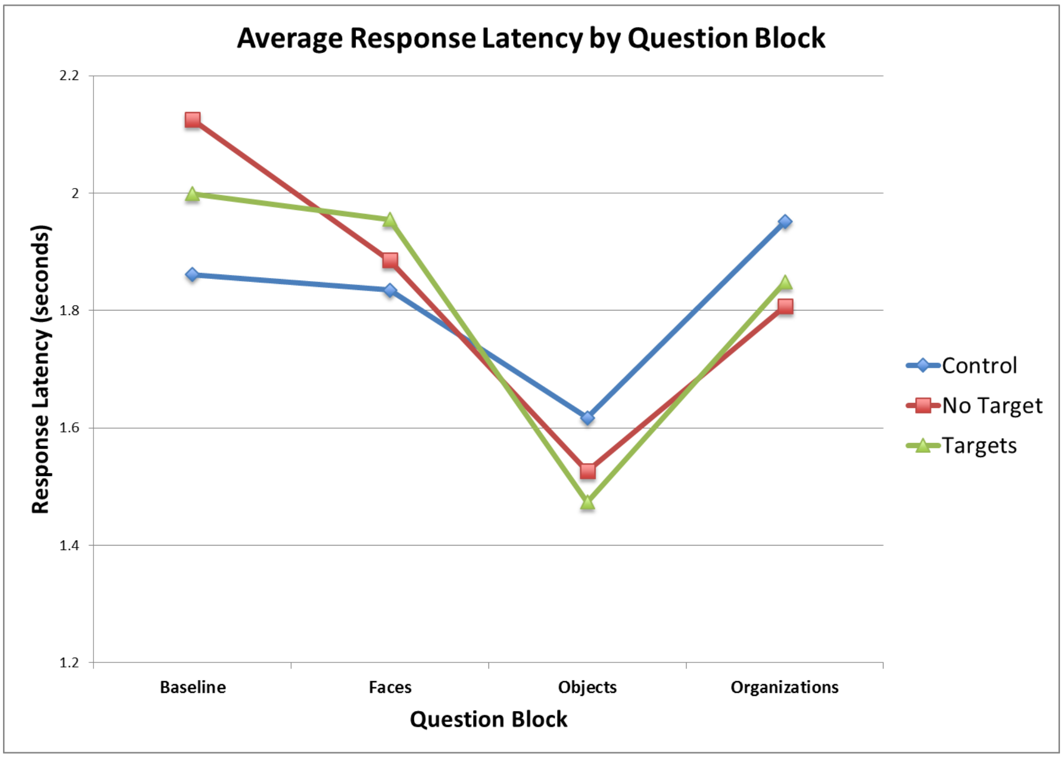


Figure 36 - Average Response Latency by Question Block

The plots contained in Figure 35 and Figure 36 support the lack of significant differences between conditions for response latency. In both figures, mean values for each condition in response latency by question and question block are very similar. The trend lines represented in Figure 36 confirm variations in response latency between question blocks; however, variations by question block for each condition are similar. This is likely attributable to differences in processing images of faces, objects, and text. It can thus be derived that processing face images is similar to processing text, while processing images of objects requires less cognitive functioning, resulting in a shorter response latency.

6.5.5 Exploratory Vocalic Analyses

The vocalic processing tools used to generate response latency values were used to extract data on three additional vocalic features to identify main-effect differences between conditions. These features include mean pitch, mean vocal quality, and mean vocal intensity. ANOVAs were used to identify significant differences between manipulation conditions and the control group for each of these features. Significance tests revealed robust differences between both manipulation groups and the control group for each of these features: Mean Pitch ($F(2, 2726) = 17.51, p < .001$), Mean Vocal Quality ($F(2, 2747) = 8.85, p < .001$), Mean Vocal Intensity ($F(2, 2747) = 4.82, p < .001$). Refer to Table 17 for a listing of mean differences between conditions for each of these vocal measures.

Table 17 - Summary of Means for Exploratory Vocalic Measures

	Pitch (Hz)	Quality (dB)	Intensity (dB)
Control	165.13	11.73	54.83
No-Targets	157.75	11.38	54.33
Targets	149.52	11.16	54.03

These findings suggest that there were significant differences between treatments despite a lack of support for a majority of the specified hypotheses. Specifically, members of the manipulation groups exhibited lower pitch, vocal quality, and vocal intensity than members of the control group.

6.6 Discussion

Only one of the five hypotheses presented in this study was supported. The supported hypothesis was H1, which confirmed the existence of a tendency for members of the no-targets group to fixate on the center of the screen longer (2.9%) than members of the control group. It is worth noting that this main effect difference was also found in the study presented in Chapter 4; however, in that experiment, participants concealing information *viewed* target items. It was thus expected that members of the targets group would have also exhibited this tendency to fixate on the center of the screen; however, this was not supported by the data. H2 was not supported, meaning that members of the no-targets group did not exhibit increased pupil dilation relative to members of the control group. However, members of the targets group did exhibit an increase in pupil dilation (2%) relative to members of the control group. This confirms one of the findings presented in Chapter 4. H3 was not supported, meaning that members of the no-targets group did not exhibit increased response latencies relative to members of the control group. This analysis also revealed that the average response latency of members of the targets group was not significantly different from the control group. This finding is opposite of a finding presented in Chapter 4, as members of the targets group exhibited longer response latencies than members of the control group.

Additionally, it was hypothesized in H4 and H5 that the no-targets group would exhibit the strategic and nonstrategic cues of deception specified in H1-H3 during the first several foils of the interview until they recognized that they were not being presented with crime-relevant information. Hypothesis tests for temporal effects revealed that there were no significant differences between the no-targets group and the control group when restricting the data set analyzed to data from only the first foil.

Despite a lack of support for the hypotheses, exploratory analyses revealed that there were three main effect vocalic differences (mean pitch, mean vocal quality, and mean vocal intensity) between members of the manipulation groups and the control group. Specifically, members of the manipulation groups exhibited lower pitch, vocal quality, and vocal intensity relative to members of the control group. In study one, there was not a significant difference in pitch between conditions, however, the mean pitch for the manipulation group was less than the mean pitch for members of the control group. Improvements in experimental design and stronger effects of the manipulation may have resulted in this increased disparity between conditions. In study one, there was not a significant difference in vocal quality between conditions; however, mean scores for vocal quality followed the pattern of pitch as in this case mean scores for the manipulation groups were less than the control group. Differences in vocal intensity were not analyzed in study one, thus, a comparison between study one and study two for this measure is not possible.

A key distinction between study one and study two was the use of an overview of the CIT during study two. Informing participants of the nature of the CIT and that oculometric and vocalic measures would be used to identify concealed information could have primed participants concealing information to use countermeasures to avoid detection. In study one, participants were unaware of which sensors were used to collect data; thus, it is likely that many of them neglected to use countermeasures. Chapter 8 contains a thorough analysis of the prevalence of countermeasures use in study two as well as a review of the various types of tactics that were used by participants to avoid detection.

6.7 Transition

The analyses and results just described indicate that a number of the hypotheses defined and tested in this study were not supported. An extensive set of post-measures was administered to each participant; these measures were used to collect data concerning various aspects of human-computer interaction during the credibility assessment interview. Key findings derived from these measures can likely be used to further understand the outcomes of this study; thus, additional discussion and the accompanying implications of the findings presented in this chapter will be presented in Chapter 8 after a discussion of the post-measures analyses presented in the following chapter.

7 STUDY 2 PART II: HUMAN-COMPUTER INTERACTION IN CREDIBILITY-ASSESSMENT INTERVIEWS

Using systems and sensor technologies for automated credibility assessments remains a research area in which there are myriad factors meriting further investigation. The emphasis of this dissertation is the evaluation of (1) two novel noncontact sensor technologies and (2) a novel adaptation of the CIT for use in automated credibility assessment interviews. However, a key aspect of automated screening research is the human-computer interaction (HCI) that occurs during the interview. The vast majority of interactions that occur between humans and computers on a daily basis are instigated *by* the user, thus, the user and the computer are often collaborating to achieve a common goal or objective (Nass, Fogg, & Moon, 1996). Automated credibility assessment interviews present a paradigm in which the user is not likely seeking or instigating the interaction and, in the case of an examinee attempting to deceive or conceal information, the objectives of the human and computer are not aligned (Twyman, Schuetzler, et al., 2013).

The post-measures administered in this study contained a variety of question types designed to elicit information from participants concerning their general perceptions, levels of stress and arousal, perceived effort and performance, and countermeasures used during the screening. The findings reported in the following sections are valuable as automated credibility assessment research has not empirically investigated these factors. These data can provide insight on how the act of deceiving or concealing information differentiates deceivers perceptually, emotionally, and behaviorally from truth tellers in this context. The following subsections will report the findings of analyses investigating how condition assignment[85] influenced (1) perceptions of the interview process, (2) self-reported stress and arousal, (3) perceived effort and performance, and (4) the use of countermeasures by participants concealing information to avoid detection.

7.1 Perceptions of the Interview Process

A user's perception of a system can influence the behaviors and emotions that occur when a user is interacting with a system; these perceptions can then determine whether or not a system will be used in the future. For example, the Technology Acceptance Model (TAM) explores the relationships between perceived usefulness, perceived ease of use, and user acceptance of a technology (Davis, 1989). While TAM can be used to better understand technology acceptance in a variety of personal and organizational computing contexts, systems designed for automated credibility assessment provide a unique context for investigating user perceptions as many users may not *want* to use the system, especially if they have intentions to conceal information or deceive.

In an attempt to glean information about user perceptions of the automated screening system used in this study, participants were asked to list five adjectives describing their interaction with the system once the interaction was complete. Using an open-ended technique to gather user perceptions is valuable as structured surveys are restrictive in nature and do not contain the full domain of possible answers. It was anticipated that the information collected during this exercise could then be used at a high level to identify (1) general impressions of the system and the interaction, (2) potential problems or weaknesses in the system or experimental task, and (3) perceptual differences between members of each of the three conditions.

---

[85] Conditions: the no-targets group, the targets group, and the control group.

7.1.1 Complete Term-List Analysis

The full corpus of terms collected during this task was refined, yielding a list of 199 unique words. The full list was then searched for the number of times that each unique term was used. Refer to Table 18 for a complete term list with accompanying frequencies for each term.

Table 18 - Complete Term List with Frequencies

Term	Count		Term	Count		Term	Count
interesting	42		real	2		intimate	1
long	24		safe	2		investigative	1
repetitive	22		slow	2		irrelevant	1
intense	14		standard	2		irritating	1
different	13		strict	2		judging	1
easy	12		tedious	2		lack-of-fear	1
intimidating	12		tense	2		lie detector	1
boring	11		thought-provoking	2		limited	1
simple	11		understandable	2		location	1
exciting	10		unexpected	2		male	1
fun	10		unnerving	2		meaningful	1
strange	10		abrasive	1		meticulous	1
stressful	10		accuracy	1		monotone	1
weird	10		accurate	1		morally-unsound	1
nerve-racking	9		adrenaline	1		mysterious	1
confusing	8		all-sided	1		neat	1
serious	8		amazing	1		necessary	1
advanced	7		anxiety-inducing	1		new	1
nervous	7		anxious	1		non-expecting	1
scary	7		applicable	1		novice	1
unique	7		avatar	1		pain-free	1
cool	6		awesome	1		peculiar	1
effective	6		bag	1		practical	1
new	6		beneficial	1		precise	1
quick	6		blue eyes	1		predictable	1
straightforward	6		bomb	1		pressing	1
awkward	5		brunette	1		pressure	1
difficult	5		careful	1		pressuring	1
interactive	5		censored	1		profound	1
professional	5		challenging	1		questionable	1
realistic	5		clothes	1		quiet	1

futuristic	5		cold	1		redundant	1
clear	4		comfortable	1		ridiculous	1
annoying	4		comprehensive	1		scientific	1
short	4		concentrated	1		selective	1
technological	4		concerning	1		silly	1
uncomfortable	4		concise	1		similar	1
bright	3		confidence	1		simulated	1
funny	3		confidential	1		sketchy	1
high-tech	3		convincing	1		slow-paced	1
interrogative	3		cool-technology	1		somewhat-difficult	1
monotonous	3		criminal	1		somewhat-scary	1
robotic	3		dangerous	1		sophisticated	1
thorough	3		dark	1		specific	1
unsure	3		daunting	1		stiff	1
useful	3		decent	1		strenuous	1
automated	2		deception	1		structured	1
brief	2		deliberate	1		sudden	1
computerized	2		demanding	1		suspenseful	1
confident	2		detailed	1		technical	1
crazy	2		doubt-inducing	1		technology	1
creative	2		dull	1		time-consuming	1
creepy	2		excitement	1		time-efficient	1
direct	2		exhausting	1		tricky	1
expensive	2		eye-tiring	1		truthful	1
extensive	2		fancy	1		unclear	1
fast-paced	2		focused	1		unconfident	1
hard	2		heart-pounding	1		unconvincing	1
inefficient	2		high-quality	1		unexplainable	1
intriguing	2		important	1		unnecessary	1
invasive	2		impressive	1		unorthodox	1
isolated	2		informational	1		unusual	1
nonintrusive	2		information-seeking	1		virtual	1
novel	2		informative	1		voice	1
odd	2		innovative	1		white	1
random	2		instantaneous	1		worthwhile	1

Refer to Figure 37 for a word cloud of the terms found in Table 18. A word cloud was created because it serves as a more visually compelling representation of the frequency with which each term was used relative to the frequency of every other term reported by participants.

Figure 37 - Word Cloud of Complete Term List

It is clear from both Table 18 and Figure 37 that there is a small subset of terms commonly reported by participants, followed by an extensive set of sparsely used or unique terms with frequencies between 1 and 3. A number of the terms used frequently refer to general elements of the interaction. For example, the use of ‘interesting’, ‘long’, ‘repetitive’, ‘easy’, ‘different’, ‘strange’, ‘interactive’, ‘futuristic’, ‘cool, ‘unique’, and ‘weird’ are terms that participants in any of the three conditions could have used to describe the interaction. Use of the terms ‘long’ and ‘repetitious’ is appropriate considering the CIT is repetitious in nature, and while the length of the interview was only six minutes in duration, it can appear to be long due to the recurrence of answering ‘Yes’ and ‘No’ for each of the 25 stimuli. Additionally, the frequent use of ‘easy’ is promising as a system designed for use without human intervention *must* be easy to use. The condition that a participant was assigned to should not have instigated or inhibited the use of any of these commonly used words.

A review of the complete term list did not identify specific problems or issues with the system; however, the frequent use of 'confusing' (used 8 times) and 'difficult' (used 5 times) indicated that there could have been issues conveying the instructions or experimental-task procedures to participants. In an effort to isolate and identify difficult or confusing aspects of the experiment associated with a specific group or groups of participants, the conditions using these two terms frequently were identified[86].

Finally, common use of 'intimidating', 'stressful', 'scary', 'intense', and 'nerve-racking' were expected from participants in the no-targets and targets groups. However, use of these terms by members of the control group could be problematic. The presence of anxiety and stress in members of the control group could render the discrimination of conditions difficult, as these emotions should have been common only for members of the manipulation groups.

7.1.2 Condition Term-List Analyses

In an attempt to (1) isolate the presence of confusion and anxiety to a specific condition and (2) determine if participants in the control group reported feelings associated with stress and anxiety, term frequencies and percentages[87] were calculated and compared for (1) the control group and the manipulation groups and (2) each of the manipulation groups[88]. A percentage-of-use cutoff point of 2% was specified for each group in an effort to promote parsimony and focus on recurring terms that are

---

[86] This issue will be investigated in a following section.

[87] Conditions were not comprised of equal numbers of participants; to provide a relative frequency for term use in a specific group, percentages were calculated and reported.

[88] Refer to Appendix E for word clouds of terms for the control group, manipulation groups combined (no-targets and targets group), no-targets group, and the targets group.

more generalizable to the "average" participant[89]. Refer to Table 19 for a list of terms frequently used by members of the control group and the members of *both* manipulation groups.

Table 19 - Term Frequencies / Percentages for Control and Manipulation Groups

Control Group	Frequency	%	Manipulation Groups	Frequency	%
repetitive	12	7%	interesting	31	8%
interesting	11	7%	long	16	4%
long	8	5%	intense	13	3%
different	7	4%	fun	10	3%
confusing	6	4%	repetitive	10	3%
advanced	5	3%	stressful	10	3%
intimidating	5	3%	easy	9	2%
strange	5	3%	exciting	9	2%
simple	4	2%	nerve-racking	9	2%
awkward	3	2%	weird	9	2%
boring	3	2%	boring	8	2%
easy	3	2%	intimidating	7	2%
straightforward	3	2%	nervous	7	2%
unique	3	2%	scary	7	2%
useful	3	2%	serious	7	2%
			simple	7	2%
			different	6	2%

Gray shading represents terms that are present in the control group, the combination of the manipulation groups, and the subgroups constituting the manipulation groups[90]. Words falling into this category include 'repetitive', 'interesting', 'easy', and 'long'. Frequent use of these terms is not surprising or problematic. Green shading represents terms found *only* in the control group and the

[89] The terms presented in Table 19 and Table 20 are the most frequently-occurring terms; however, their use relative to the number of terms used by participants in each group remains in the minority. The most frequently-used term in any group is 'interesting' by members of the targets group, which accounted for a 10% share of all terms used by members of that group.

[90] Details of term usage for each manipulation subgroup is presented in Table 20.

combined manipulation group lists. Terms falling into this category include 'different', 'intimidating', 'simple', and 'boring'. It is intriguing that members of the control group would use the term 'intimidating' as this term seems more appropriate for participants tasked with concealing information during the interview. It can be derived from this finding that aspects of the system (e.g., multiple cameras and microphones) or the interview itself can be intimidating regardless of the presence of concealing information or the intent to deceive. It is also noteworthy that members of the manipulation groups used boring[91] to describe the interview.

Other terms frequently used by the control group include 'advanced', 'strange', 'awkward', 'straightforward', 'unique', and 'useful'. These terms are understandable and justifiable in their use as this technology is indeed advanced, strange, unique, and useful. The use of straightforward is promising as a technology designed for automated credibility assessments *without* the direct involvement of human screeners and interviewers *must* be straightforward and easy to use. Terms frequently used by the manipulation groups include 'intense', 'fun', 'stressful', 'exciting', 'nerve-racking', 'weird', 'nervous', 'scary', and 'serious'. A majority of these terms serve as an indirect measure of manipulation effects as the experiment is designed to elicit feelings of intensity, nervousness, and stress for members of the manipulation groups. While the purpose of the experiment was not to be fun or exciting, it appears that smuggling a fake IED through a simulated screening interview while deceiving the system was enjoyable to some participants[92].

---

[91] The use of 'boring' by members of the no-targets and targets groups is discussed in a following section.

[92] A number of participants indicated how fun the experiment was and asked experiment facilitators if any additional experiments of this nature would be occurring in the future.

Perhaps the most important and interesting finding of this analysis is the high ranking of 'confusing' on the term list for the control group[93]. It was previously discussed that the high ranking of confusing in Table 18 could be problematic as a subset of participants may have found the experimental task or procedures difficult to understand. However, since a majority of the instances of 'confusing' are associated with members of the control group, it is likely that the confusion they experienced is attributable to the nature of the task itself. Control group members were not tasked with smuggling a banned item, and had no exposure to faces or criminal organizations during the pre-interview phase. Control group members may have found that completing an automated interview of this nature was confusing from the standpoint that it was strange, unexpected, or irrelevant[94]. In light of this finding, it is assumed that the experimental task and procedures were well-defined and instructions were easy to follow.

To identify term-use differences between members of each of the manipulation groups, a second analysis was conducted. Refer to Table 20 for a list of frequently-used terms by members of *each* manipulation group.

Table 20 - Term Frequencies / Percentages for No-Targets and Targets Groups

Targets	Frequency	%	No Targets	Frequency	%
interesting	19	10%	interesting	12	6%
long	7	4%	long	9	4%
stressful	7	4%	intense	8	4%
exciting	6	3%	boring	6	3%
fun	6	3%	repetitive	6	3%
weird	6	3%	easy	5	2%

[93] Aside from the control group, 'confusing' was used by only two participants in the manipulation groups.

[94] Previously, 'difficult' was linked with 'confusing' as a potentially problematic word; however, it was sparsely used by members of *each* condition; thus, its infrequent use both across and within conditions cannot be linked with any general or condition-specific issues, respectively.

intense	5	3%	serious	5	2%
nerve-racking	5	3%	simple	5	2%
easy	4	2%	exciting	4	2%
intimidating	4	2%	fun	4	2%
nervous	4	2%	interactive	4	2%
realistic	4	2%	nerve-racking	4	2%
repetitive	4	2%	quiet	4	2%
scary	4	2%			
cool	3	2%			
different	3	2%			
short	3	2%			

As referenced previously, terms shaded in gray were reported by members of each of these groups as well as the groups described in Table 20, thus, further discussion of these terms is not necessary in this section. Blue shading was used to highlight terms recurring only in the manipulation subgroups. Terms shaded in blue include ‘exciting’, ‘fun’, ‘intense’, and ‘nerve-racking’. The use of these words was previously identified in Table 20, which contained terms frequently used by *both* manipulation groups, thus, it was expected that participants in *each* of the manipulation subgroups would have used these terms.

A number of terms were used only by participants in the targets group; these terms include ‘stressful’, ‘weird’, ‘intimidating’, ‘nervous’, ‘realistic’, ‘scary’, ‘cool’, ‘different’, and ‘short’. Members of the no-targets group who did not see any target items during the screening interview reported that the interview was ‘boring’, ‘serious’, ‘simple’, ‘interactive’, and ‘quiet’. It is interesting to note the stark contrast between the words used exclusively by each of the two manipulation groups.

7.2 Self-Reported Stress and Arousal

The traditional CIT minimizes stress and arousal as it is used to measure an examinee’s recognition of crime-relevant stimuli (Krapohl et al., 2009). An examinee possessing no knowledge of crime-relevant information should not experience feelings of stress and arousal, resulting in a traditionally low false-positive rate associated with the CIT. However, a CIT without target items cannot be used to measure recognition, thus, other emotional, behavioral, and physiological responses must be measured to identify concealed information. Chapter 5 contains a discussion of deception literature in which deception is described as a complex and dynamic process. IDT posits that deceivers employ strategic and subsequent nonstrategic cues while deceiving, resulting in a leakage of deception-related cues as deceivers have difficulty managing cognitive and behavioral processes simultaneously (Buller & Burgoon, 1996). Additionally, stress and arousal are the fundamental point of interest in a number of credibility-assessment methodologies, thus, identifying the presence of these emotions in examinees completing an automated interview is both relevant and valuable.

Self-reported measures of stress and arousal were collected to (a) inform findings from the hypothesis tests in section 6.5 and (b) determine if the interview process itself instilled feelings of stress and arousal in all participants, or only participants constituting the manipulation groups. The following sections contain (1) an overview of the measures that were used to collect self-report data of emotional states, (2) the results of an exploratory factor analysis (EFA) conducted to identify distinct emotional states felt by examinees, and (3) the results of significance tests conducted to identify the presence of statistically significant differences in emotions felt by members of each condition. The corpus of data (N=114) used for these analyses is comprised of data from 33 members of the control group, 43 members of the no-targets group, and 38 members of the targets group.

7.2.1 Measurement Selection

Two approaches for measuring stress and arousal were considered for this analysis. Thayer (Thayer, 1967) developed an instrument to measure activation based on a list of mood-describing adjectives developed previously (Mackay, Cox, Burrows, & Lazzenini, 1978; Nowlis, 1965). These adjectives were grouped into four factors referred to as the Activation-Deactivation Adjective Checklist (AD-ACL) (Thayer, 1970); each factor represented a different level of arousal. Refer to Table 21 for a table listing each factor with its corresponding mood-describing adjectives.

Table 21 - AD-ACL Factors and Adjectives

Factor Name	Adjectives
General Activation	Lively, active, full-of-pep, energetic, peppy, vigorous, activated
High Activation	Clutched-up, jittery, stirred-up, fearful, intense
General Deactivation	At rest, still, leisurely, quiescent, quiet, calm, placid
Deactivation-Sleep	Sleepy, tired, drowsy

Additional research investigating the use of these factors to measure stress and arousal identified instability issues in the factor loadings (Thayer, 1970, 1971, 1975). Mackay (1978) developed a revised list of terms and found that two, rather than four, factors can be used, resulting in a more parsimonious design. The revised term list developed by Mackay is presented in Table 22.

Table 22 - Revised Term List (Mackay, 1978)

Tense	Relaxed	Vigorous	Stirred-Up
Restful	Active	Apprehensive	Expectant
Worried	Energetic	Drowsy	Insensitive
Bothered	Uneasy	Intense	Dejected

Leisurely	Quiet	Nervous	Placid
Quiescent	Distressed	Fearful	Peaceful
Activated	Tired	Idle	Up-tight
Alert	Lively	Stimulated	Aroused
At rest	Somnolent	Cheerful	Passive
Contented	Jittery	Sluggish	Still
Pleasant	Sleepy	Comfortable	Calm
Excited			

Each participant was tasked with reviewing the list of emotions provided in Table 22 and rating the extent to which they felt each emotion during their interaction with the system. Participants rated how strongly they felt each emotion using a four-point scale adapted from the scale used by Thayer (1967), with 1=Definitely felt and 4=Did not feel.

7.2.2 Exploratory Factor Analysis (EFA)

Of the complete corpus of data points collected using post-measures (N=114), only 106 data points were considered for analysis due to the removal of several records missing data. Due to a lack of evidence supporting the use of either the four-factor or two-factor approach to measuring stress and arousal, an exploratory factor analysis (EFA) was used to develop an optimal loading of factors and variables for this data set. A number of steps were taken to ensure that the data were suitable to be factor analyzed. Raw scores from participant responses were used to create a 45 x 45 correlation matrix. The matrix was reviewed for problematic coefficients (e.g., multicollinearity). No problems were found. A Bartlett test was conducted to ensure that relationships between variables were present, and that the correlation matrix was significantly different from an identity matrix. Bartlett's test was highly significant for these data ($\chi^2(990) = 2917$, $p < .001$), indicating that factor analysis was appropriate.

A Kaiser-Meyer-Olkin (KMO) test was used to ensure that the sample size and data were suitable to be factor analyzed. KMO tests yield coefficients indicating a presence or lack of patterns in the data for the complete set of variables as well as each individual variable. KMO coefficients closer to 1 indicate that the data are suitable to be factor analyzed; coefficients less than 0.5 should not be retained in the data set submitted to an EFA (Hutcheson & Sofroniou, 1999; Kaiser, 1974). A KMO test on the full set yielded a coefficient of 0.7755; three variables had scores near or below 0.5 (Quiscent = 0.4747, Somnolent = 0.5386, and Still = 0.5202). These variables were removed from the data set and a second KMO test was conducted. The refined data set yielded a KMO coefficient of 0.80367; the remaining variables all exceeded 0.5. Refer to Appendix F for complete results from the KMO tests.

Next, the number of factors was selected using an automated script designed to (1) plot and compare eigenvalues, (2) conduct a parallel analysis, (3) identify optimal coordinates, and (4) evaluate acceleration factors to determine the optimal number of factors for extraction. According to Kaiser's rule (Kaiser, 1970), 10 factors should be extracted as 10 factors had eigenvalues greater than 1. The parallel analysis and the optimal coordinates analysis identified 4 factors that should be extracted. Finally, the acceleration factor analysis identified only one factor. Due to the majority finding by two of the four automated factor extraction methods that four factors should be extracted, paired with the four factors approach identified by Thayer (1967), an EFA was conducted using four factors. A graphical representation of the results of the factor-extraction analysis is presented in Figure 38.

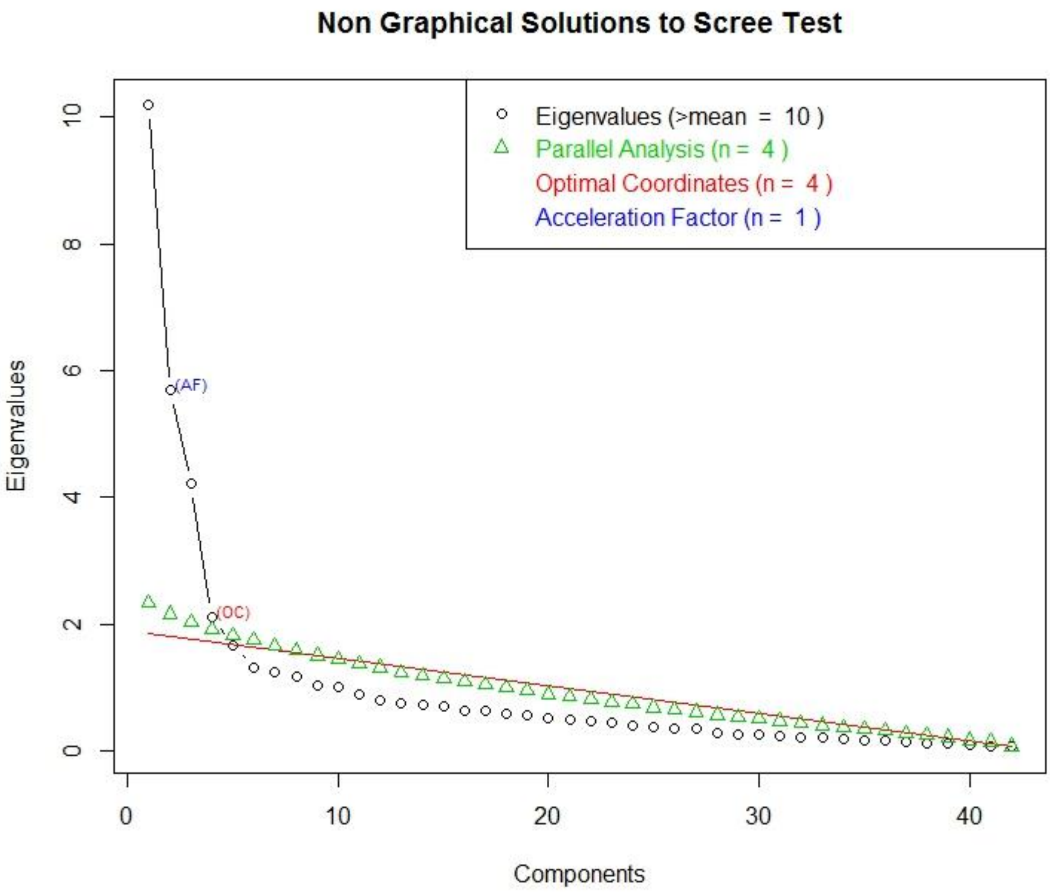


Figure 38 - EFA Factor Extraction Analysis

An EFA was conducted on 42 variables with oblique rotations (promax) as some of the factors should be related. Factor loadings looked promising as variable coefficients supported the presence of convergent and discriminant validity (Campbell & Fiske, 1959); however, a number of variables were cross-loaded or performed poorly. The number of variables was reduced from 42 to 29; the following variables were removed: vigorous, apprehensive, expectant, insensitive, intense, dejected, quiet, idle, alert, stimulated, aroused, cheerful, and contented. A second EFA was conducted to determine the impact of removing these variables on factor loadings. The second EFA contained indications of improvement; however, 8 more variables were identified for removal, including: stirred-up, bothered, passive, pleasant, calm, active, placid and comfortable. A third and final EFA was conducted using a set of 21 variables. The final set of 21 variables yielded factor loadings that were sufficiently high; all cross loadings were resolved. The results of the third EFA are provided in Table 23. Items clustering on the same factor indicate that Factor 1 represents feelings of anxiety, Factor 2 represents feelings of lethargy, Factor 3 represents feelings of tranquility, and Factor 4 represents feelings of enthusiasm.

Table 23 - EFA Results

	Promax rotated factor loadings			
Item	**Anxiety**	**Lethargy**	**Tranquility**	**Enthusiasm**
Tense	0.73			
Worried	0.94			
Uneasy	0.72			
Nervous	0.83			
Distressed	0.68			
Fearful	0.79			
Uptight	0.65			
Jittery	0.58			
Drowsy		0.84		
Tired		0.84		
Sluggish		0.52		
Sleepy		1.00		
Relaxed			0.61	
Restful			0.72	
Leisurely			0.84	
Peaceful			0.72	
At rest			0.58	
Energetic				0.66
Activated				0.76
Lively				0.77
Excited				0.81

The factor correlations provided in Table 24 also support the presence of convergent and discriminant validity between factors.

Table 24 - EFA Factor Correlations

	Factor Correlations			
	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.000	0.055	-0.202	-0.309
Factor 2	0.055	1.000	-0.380	-0.360
Factor 3	-0.202	-0.380	1.000	-0.049
Factor 4	-0.309	-0.360	-0.049	1.000

A comparison of the factor loadings between the adjectives used in this analysis and the factors / adjectives proposed by Thayer (1967) indicate many similarities between the two factor structures.

7.2.3 Significance Tests

Factor scores were extracted from the EFA and used for significance testing. Multiple ANOVAs and Bonferroni pairwise t-tests were used to identify significant differences in emotional responses between conditions based on each participant's self-reported emotions during the interview.

Factor scores for Factor 1, labeled the 'Anxiety' scale[95], were submitted to an ANOVA; this analysis indicated that there was not a significant difference between the responses of the control group and the no-targets group ($t(103) = 1.68, p = .095$). However, there was a significant difference between the control group and the targets group ($t(103) = 3.55, p < .001$). A Bonferroni pairwise t-test indicated that there was not a significant difference between manipulation groups for this scale. These findings indicate that participants in the targets group experienced higher levels of anxiety than the other two conditions. It can also be inferred that the no-targets group had anxiety ratings similar to participants in each of the other two conditions, thus, the mean score for this group was not significantly different from the mean scores of the other two

[95] Anxiety scale variables: tense, worried, uneasy, nervous, distressed, fearful, uptight, jittery.

conditions. Refer to Figure 39 for a box plot of factor scores associated with members of each condition. The box plots visually confirm that the distribution of scores for the no-targets group overlaps with the scores associated with the other two conditions.

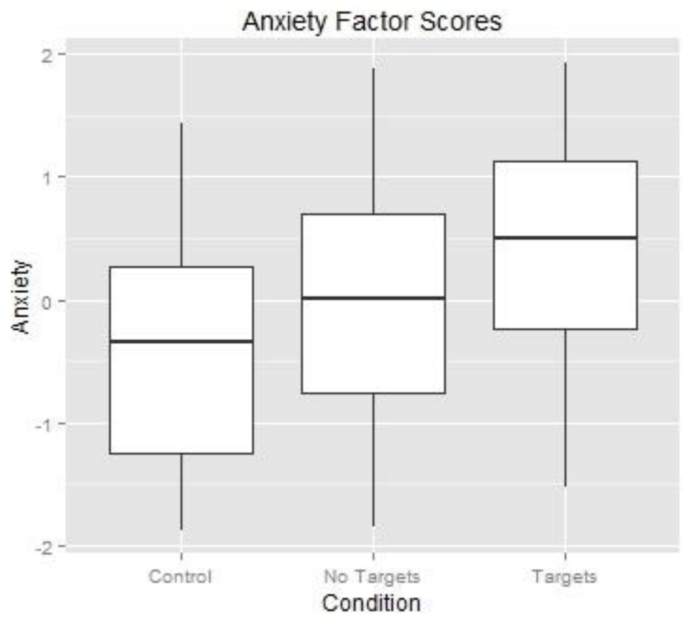


Figure 39 - Box Plots of Anxiety Factor Scores

Factor scores for Factor 2, labeled the 'Lethargy' scale[96], were submitted to an ANOVA; this analysis indicated that there was a marginally significant difference between the responses of the control group and the targets group ($t(103) = -1.87$, $p = .064$). Additionally, there was a significant difference between the control group and the no-targets group ($t(103) = -2.49$, $p = .014$). A Bonferroni pairwise t-test indicated that there was not a significant difference between manipulation groups for this scale. These findings indicate that participants in both manipulation groups experienced lower levels of lethargy relative to members of the control group. Refer to Figure 40

[96] Lethargy scale variables: drowsy, tired, sluggish, sleepy.

for a box plot of factor scores associated with members of each condition. The box plots provide visual support for the significance test results reported in this paragraph.

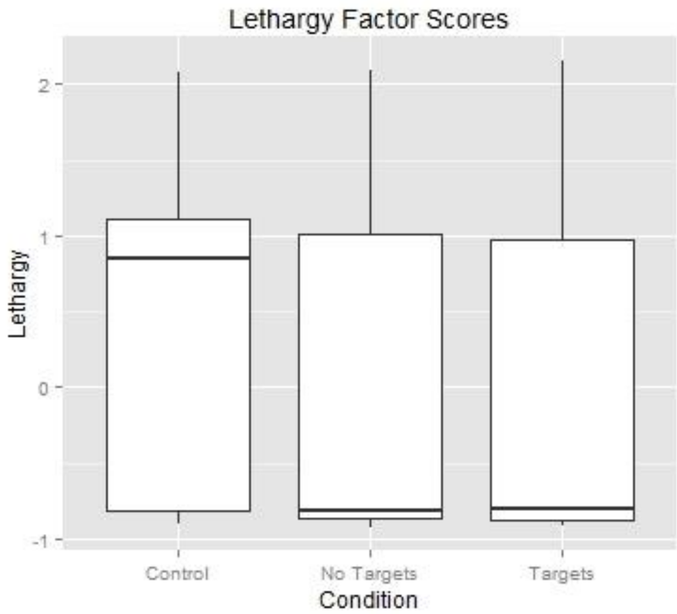


Figure 40 - Box Plots of Lethargy Factor Scores

Factor scores for Factor 3, labeled the 'Tranquility' scale[97], were submitted to an ANOVA; this analysis indicated that there was not a significant difference between the responses of the control group and the no-targets group ($t(103) = -1.46$, $p = .14$). However, there was a marginally significant difference between the control group and the targets group ($t(103) = -1.82$, $p = .07$). A Bonferroni pairwise t-test indicated that there was not a significant difference between manipulation groups for this scale. These findings indicate that participants in the targets group experienced lower levels of tranquility that were marginally different from the participants in the control group. It can also be inferred that the no-targets group had responses similar to participants in each of the other two conditions. Refer to Figure 41 for box plots of factor scores

[97] Lethargy scale variables: relaxed, restful, leisurely, peaceful, at rest.

associated with members of each condition. The box plots visually confirm that the distribution of scores for the no-targets group overlaps with the other two conditions.

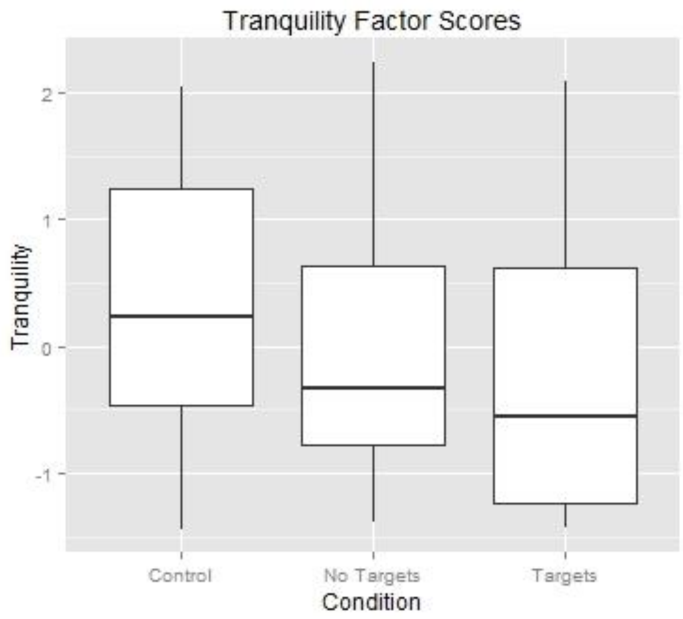


Figure 41 - Box Plots of Tranquility Factor Scores

Factor scores for Factor 4, labeled the 'Enthusiasm' scale[98], were submitted to an ANOVA; this analysis indicated that there was not a significant difference between the responses of the control group and the no-targets group ($t(103) = 1.47$, $p = .14$). However, there was a significant difference between the control group and the targets group ($t(103) = 2.65$, $p = .009$). A Bonferroni pairwise t-test indicated that there was not a significant difference between manipulation groups for this scale. These findings indicate that participants in the targets group experienced higher levels of enthusiasm that were significantly different from the other two conditions. It can also be inferred that the no-targets group had responses similar to participants in each of the other two conditions. Refer to Figure 42 for box plots of factor scores associated with members

[98] Enthusiasm scale variables: energetic, activated, lively, excited.

of each condition. The box plots visually confirm that the distribution of scores for the no-targets group overlaps with the other two conditions.

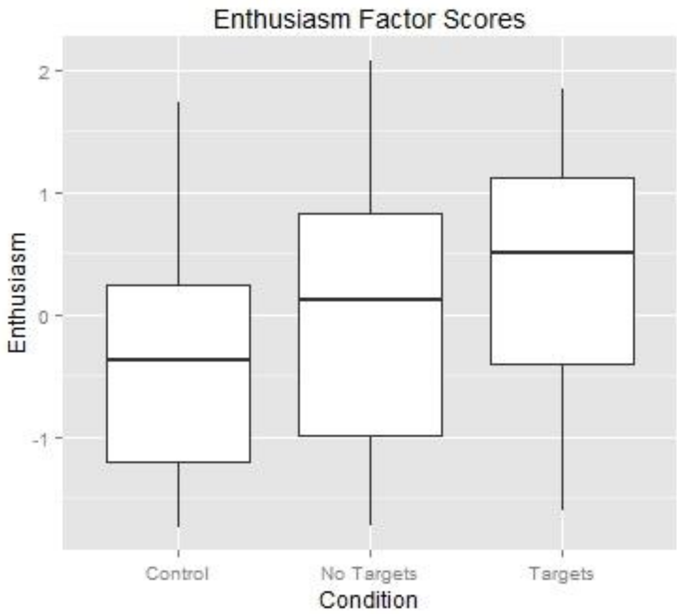


Figure 42 - Box Plots of Enthusiasm Factor Scores

7.3 Perceived Effort and Performance

In an effort to gauge each participant's perceived effort and performance during the interview, as well as differences in perceived effort and performance between conditions, post-measures assessing each participant's (1) effort to appear truthful, (2) motivation to appear truthful, and (3) confidence that they did appear truthful were formulated[99]. The questions presented in Table 25 were used to measure these three aspects of perceived performance.

[99] These analyses are a continuation of the manipulation check presented in Chapter 6. For the purposes of the manipulation check, the descriptive statistics (by condition) for each question were reviewed to ensure that participants in the manipulation groups were sufficiently motivated to avoid detection. The purpose of this analysis is to more closely investigate the presence of significant difference between conditions and understand how condition assignment affects perceptions of the interviewing process.

Table 25 - Questions Measuring Perceived Performance

#	Question
1	During the interview, how important was it to you to succeed in making the interviewer believe you?
2	During the interview, how important was it to you to give convincing answers?
3	How hard did you try to convince the interviewer that you were telling the truth?
4	How hard did you try to avoid suspicion during the interview?
5	How successful do you think you were in convincing the interviewer that you were truthful?

The corpus of data (N=114) used for these analyses is comprised of data from 33 members of the control group, 43 members of the no-targets group, and 38 members of the targets group. To answer each question, participants were asked to rate each question using a five-point Likert scale. For questions 1 and 2, response options ranged from 1="Not at all important" to 5="Extremely important". For questions 3 and 4, response options ranged from 1="Not at all" to 5="Very hard". For question 5, response options ranged from 1="Very unsuccessful" to 5="Very successful". Response data were analyzed using a four-step process. First, data for each question were visually represented and interpreted using boxplots. Second, descriptive statistics for each question and condition combination were calculated. Third, statistically significant differences between each manipulation group and the control group were identified using an analysis of variance (ANOVA). Finally, Bonferroni pairwise t-tests were used to identify statistically significant differences between the two manipulation groups.

7.3.1 Importance of Interviewer Belief

Refer to Figure 43 for box plots of responses to Question 1: "During the interview, how important was it to you to succeed in making the interviewer believe you?"

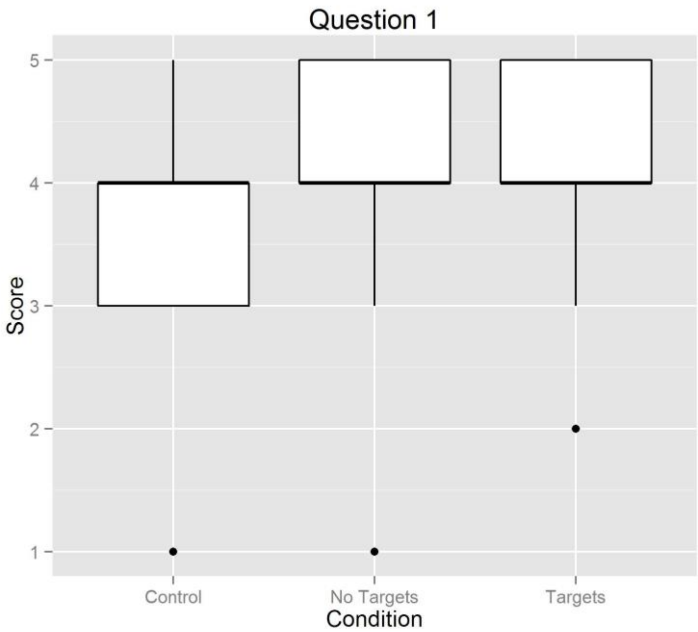


Figure 43 – Importance of Interviewer Belief: Box Plots for Question 1

It is evident from Figure 43 that both manipulation groups found it important to make the interviewer believe them. This is expected as the manipulation groups were tasked with concealing information and were motivated to obtain the monetary reward promised to participants appearing truthful. It is expected that members of the control group would rate this question similarly to the manipulation groups as control group members were also motivated to appear truthful in order to receive the reward; however, as members of the control group did not have to deceive to pass the interview, the importance of convincing the interviewer of their truthfulness was likely subdued.

The box plots indicate that the distribution of scores for the control group was in fact lower than the two manipulation groups, while the distribution of scores for the manipulation groups appear almost identical except for their respective outliers. Refer to Table 26 for descriptive statistics for each condition.

Table 26 - Descriptive Statistics for Question 1

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Control	3.39	1.37	4	1	5	4	-0.62	-0.59
No Targets	4.23	0.80	4	1	5	4	-1.74	5.55
Targets	4.29	0.77	4	2	5	3	-0.94	0.67

An ANOVA and a Bonferroni pairwise t-test were used to determine if inter-condition mean differences were statistically significant. Differences in mean response scores for this question between the control group and each of the manipulation groups were significant ($F(2, 111) = 9.52, p < .001$). A Bonferroni test found no significant difference between the mean response scores for the manipulation groups. The results of the ANOVA and Bonferroni test confirm the interpretation of the box plots presented in the previous paragraph.

7.3.2 Importance of Convincing Answers

Refer to Figure 44 for box plots of responses to Question 2: "During the interview, how important was it to you to give convincing answers?"

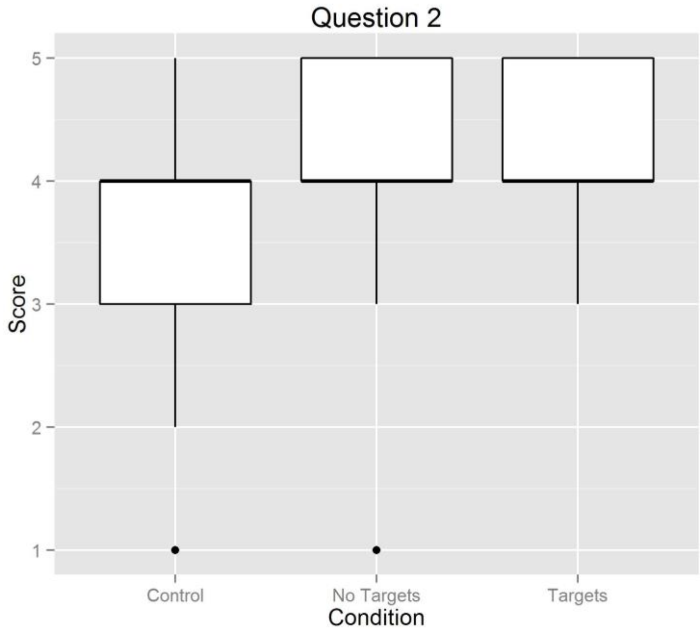


Figure 44 – Importance of Convincing Answers: Box Plots for Question 2

It is apparent in Figure 44 that both manipulation groups found it more important to provide convincing answers to the interviewer relative to the control group. It was expected that responses to Question 2 would follow a pattern similar to the pattern in the box plots for Question 1, as the importance of fostering belief in the interviewer is related to the importance of providing convincing answers to the interviewer.

Visually, Figure 43 and Figure 44 are very similar except for the line below the box plot for the control group. The distribution of scores for the control group was in fact skewed lower than the two manipulation groups, while the distribution of scores for the manipulation groups appear almost identical except for a single outlier in the no-targets group. Refer to Table 27 for descriptive statistics for each condition.

Table 27 - Descriptive Statistics for Question 2

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Control	3.45	1.28	4	1	5	4	-0.66	-0.47
No Targets	4.20	0.77	4	1	5	4	-1.68	5.87
Targets	4.39	0.59	4	3	5	2	-0.38	-0.64

An ANOVA and a Bonferroni pairwise t-test were used to determine if inter-condition mean differences were statistically significant. Differences in mean response scores for each of the manipulation groups relative to the control group were significant ($F(2, 111) = 10.63$, $p < .001$). Additionally, a Bonferroni test found no significant difference between the mean response scores for the two manipulation groups. The results of the ANOVA and Bonferroni test confirm the interpretation of the box plots presented in the previous paragraph.

7.3.3 Attempt to Convince Interviewer

Refer to Figure 45 for box plots of responses to Question 3: "How hard did you try to convince the interviewer that you were telling the truth?"

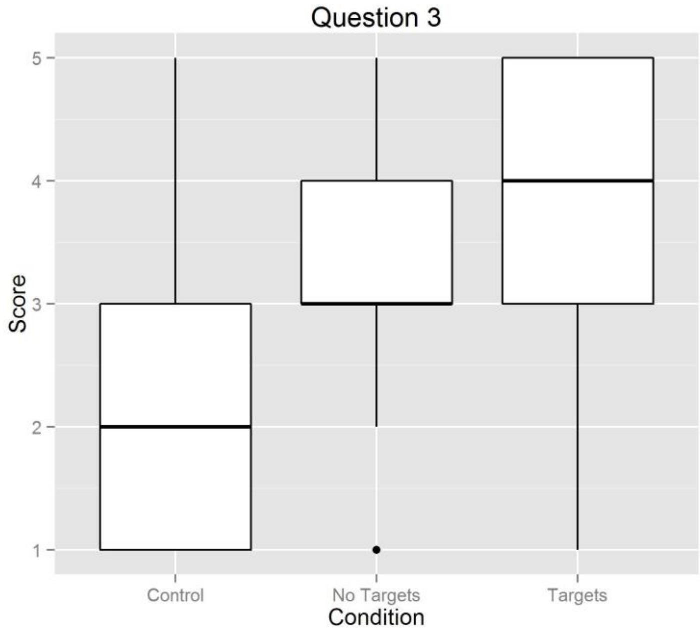


Figure 45 – Attempt to Convince Interviewer: Box Plots for Question 3

While Questions 1 and 2 were used to measure the importance level or motivation of each participant to appear truthful, Questions 3 and 4 were used to determine the *extent* to which participants operationalized this importance. It was anticipated that the response patterns of the manipulation groups would be similar to each other, but different from members of the control group.

It is clear in Figure 45 that there is a tremendous amount of variation between the responses patterns for each condition. A majority of control-group members did not perceive themselves to have tried very hard to appear truthful, while members of the manipulation groups had much higher scores for this question. A key finding based on this box plot is that members of the targets group generally rated themselves higher than members of the no-targets group on this question. This suggests that the presence of target items in the interview caused some members of the targets group to try harder to appear truthful than members of the no-targets group. Refer to Table 28 for descriptive statistics for each condition.

Table 28 - Descriptive Statistics for Question 3

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Control	2.12	1.32	2	1	5	4	1.07	0.05
No Targets	3.16	1.17	3	1	5	4	-0.43	-0.34
Targets	3.74	1.11	4	1	5	4	-0.58	0.04

An ANOVA and a Bonferroni pairwise t-test were used to determine if inter-condition score differences were statistically significant. Differences in mean response scores between the control group and each of the manipulation groups were significant ($F(2, 111) = 16.41$, $p < .001$). However, despite visual differences in the box plots of responses for the manipulation groups, the difference between the mean response scores for these groups was not statistically significant. The results of the ANOVA confirm the interpretation of the box plot presented in the previous paragraph; however, the assumption that there was a statistically significant difference between the mean scores for each manipulation group was not supported.

7.3.4 Attempt to Avoid Suspicion

Refer to Figure 46 for box plots of responses to Question 4: "How hard did you try to avoid suspicion during the interview?"

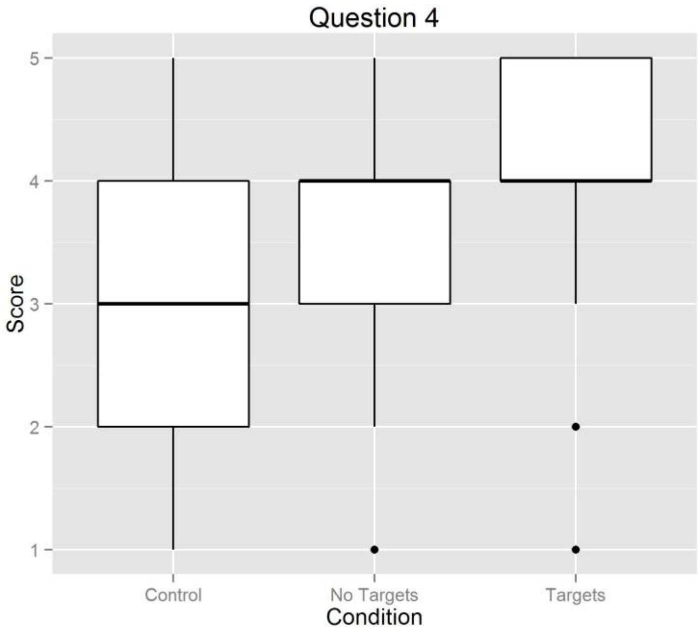


Figure 46 – Attempt to Avoid Suspicion: Box Plots for Question 4

Questions 3 and 4 are both measuring the *efforts* of participants to succeed in the screening interview; however, each question was designed to measure this effort from a different perspective. Question 3 investigates participant efforts to *appear truthful* while Question 4 investigates participant efforts to *avoid suspicion*. It was expected that members of both manipulation groups would exhibit similar behaviors in an attempt to avoid suspicion, while members of the control group would not have employed such behaviors due to a lack of concealed information.

The pattern of responses in Figure 45 is similar to the pattern present in Figure 46; however, it appears as though members of the control group *were* attempting to avoid suspicion, as opposed to an apparent lack of effort to appear truthful (as discussed in the previous section). Responses by members of the no-targets group in Figure 45 and Figure 46 appear to be similar, while members of the targets group reported a higher consensus of attempting to avoid suspicion (depicted in Figure 46) relative to their perceived effort to be convincing about telling the truth (depicted in Figure 45). Refer to Table 29 for descriptive statistics for each condition.

Table 29 - Descriptive Statistics for Question 4

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Control	3.06	1.09	3	1	5	4	-0.44	-0.90
No Targets	3.69	0.99	4	1	5	4	-0.58	0.07
Targets	4.13	0.96	4	1	5	4	-1.23	1.75

An ANOVA and a Bonferroni pairwise t-test were used to determine if inter-condition mean differences were statistically significant. Differences in mean response scores between the control group and each of the manipulation groups were significant as ($F(2, 111) = 9.98$, $p < .001$). However, despite differences in the distribution of scores for the manipulation groups, a Bonferroni test found no significant difference between the mean response scores for these conditions. The results of the ANOVA confirm the interpretation of the box plot presented in the previous paragraph; however, the assumption that there was a statistically significant difference between the mean scores for each manipulation group was not supported.

7.3.5 Perceived Success

Refer to Figure 47 for box plots of responses to Question 5: “How successful do you think you were in convincing the interviewer that you were truthful?”

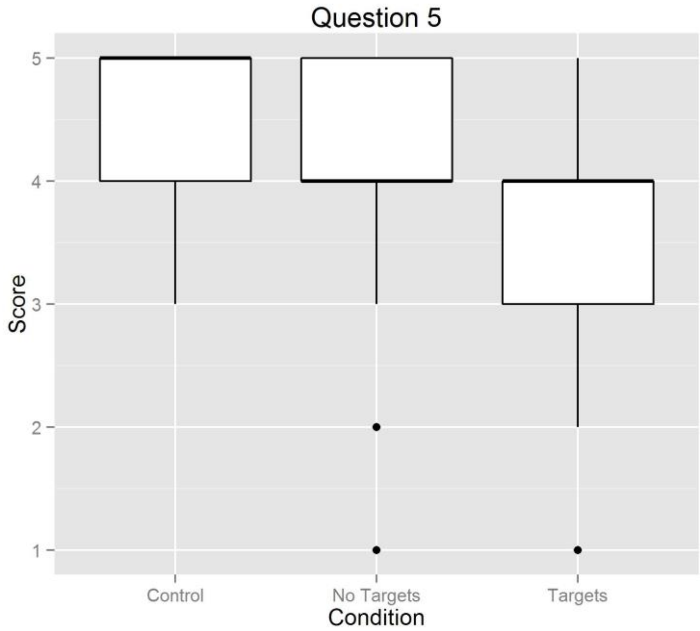


Figure 47 – Perceived Success: Box Plots for Question 5

Questions 1 through 4 were designed to measure (1) how important it was for participants to pass the interview and (2) how hard they tried to do so. Question 5 was used to determine whether or not participants perceived that they had succeeded in appearing innocent during the interview. It was expected that members of the control group would be very confident that they had passed the interview as they were unaware of target items. Figure 47 supports this interpretation. It was also anticipated that the score distribution for members of the no-targets group would resemble the distribution of scores for the control group. This expectation is attributable to the lack of target items for the no-targets group; thus, participants in this group should not have perceived that they had been identified by the system as concealing information. On the other hand, it was expected that participants in the targets group would have less confidence that they had successfully passed the interview.

Figure 47 provides visual support for each of these expectations. The distribution of scores for members of the no-targets group looks similar to the control group. This finding may imply that members of the no-targets group were either not aware of the system's ability to measure persisting physiological and behavioral differences during the interview or that members of the no-targets group had minimal confidence in the system's ability to do so. Additionally, the distribution of scores for members of the targets group is different than the distribution of scores for the no-targets group. This finding may imply that the presence of target items reduces examinee confidence in passing the interview. Indirectly, this means that members of the targets group had confidence in the system's ability to detect concealed information. Refer to Table 30 for descriptive statistics for each condition.

Table 30 - Descriptive Statistics for Question 5

Condition	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis
Control	4.45	0.62	5	3	5	2	-0.67	-0.43
No Targets	4.14	0.96	4	1	5	4	-1.29	1.75
Targets	3.63	1.1	4	1	5	4	-0.74	0.07

An ANOVA and a Bonferroni pairwise t-test were used to determine if inter-condition mean differences were statistically significant. Differences in mean response scores for this question between the control group and the targets group was significant as ($t(111) = -3.715$, $p < .001$). However, there was not a significant difference between the mean response scores for the control group and the no-targets group. Additionally, a Bonferroni test found a significant difference between the mean response scores on this question for the no-targets group and the targets group ($p = .047$, $d = 0.5$). The results of the ANOVA and Bonferroni test confirm the interpretation of the box plot presented in the previous paragraph.

7.4 Countermeasures

It was expected that members of the manipulation groups would exhibit strategic and nonstrategic cues of deception in an attempt to avoid detection. Prior research revealed that examinees concealing information during a visual CIT exhibited a number of main effect differences distinguishing them from members of the control group. These differences included a tendency to fixate on the center of the screen, increased pupil dilation, and increased response latencies. Hypothesis tests provided support for only one of these three behaviors, but due to a pre-test overview of the nature of the CIT and the types of sensors that were being used, it is possible that participants were primed to use countermeasures to avoid detection. A qualitative analysis was conducted to (1) identify the relative use of countermeasures between conditions, (2) identify different types of countermeasures and the extent to which they were used by members of each condition, and (3) identify novel countermeasures reported by participants that may prove useful in identifying attempts to conceal information in future automated credibility assessment research[100]. Refer to Appendix G for a table containing a list of the countermeasures used by each participant who reported using countermeasures, grouped by condition. Results of the analyses reported in the following three sections are based on the information included in Appendix G.

7.4.1 Relative Use of Countermeasures

Roughly half (48%) of all participants reported using countermeasures. The frequency of participants in each condition who reported using countermeasures was

[100] It is worth noting that only one study has previously reported findings on the use of countermeasures against an automated screening system (Twyman, Schuetzler, Proudfoot, & Elkins, 2013). In this study, participants were instructed on the use of mental and physical countermeasures typically used to thwart the effectiveness of sensors used for polygraph interviews. This research provides valuable insight on the types of novel countermeasures that may be employed against automated screening systems measuring oculometric and vocalic behaviors.

compared to the number of total participants in each condition. Surprisingly, one quarter (25%) of members of the control group reported using countermeasures to avoid detection. This is an interesting finding considering that members of this condition did not possess crime-relevant information and thus had no reason to attempt to avoid detection. Roughly half (42%) of members of the no-targets group employed countermeasures to avoid detection, while 66% of members of the targets group reported using countermeasures to avoid detection. It can be inferred from these findings that the presence of concealed information increases the propensity for an examinee to try to avoid detection (a 17% increase from the control group to the no-targets group); however, the interaction effect of concealed information and the presence of target items resulted in the maximum usage of countermeasures (an increase of 22% from the no-targets group to the targets group). Table 31 contains a comparison of the percentages of participants in each condition who reported using countermeasures.

Table 31 - Percentage of Participants Using Countermeasures

	Control	No Targets	Targets
Used Countermeasures	8	18	25
Did Not Use Countermeasures	33	43	38
Percentage	24%	42%	66%

7.4.2 Countermeasures Used

The list of countermeasures used by participants was reviewed and categorized. Some participants reported using only a single countermeasure while others reported using multiple countermeasures. Table 32 contains a summary of the average number of countermeasures used by all participants, as well as the average number of countermeasures used by members of each condition. The targets group, on average, used more countermeasures than any other group.

Table 32 - Average Number of Countermeasures Used

Condition	Average Number of Countermeasures Used
All	1.31
Control	1.13
No Targets	1.33
Targets	1.36

Categories were formulated to group each countermeasure based on the nature of the countermeasures that were used. Countermeasures used by only one participant did not warrant the formulation of a unique category; thus, a category titled ‘Other’ was formed to group countermeasures used infrequently. Table 33 contains a list of countermeasure categories with an accompanying description.

Table 33 - Countermeasure Categories

Countermeasure Category	Description
Blurred Viewing	Blurring eyes to avoid viewing any of the images on the screen.
Center of Screen	Averting eye gaze from images and fixating on the center of the screen.
Consistent Viewing	Using the same pattern of viewing images for each slide (e.g., looking at the image in the top-left quadrant first, then looking at the image in the top-right quadrant, etc.).
Emotion Control	Attempting to control emotional states to avoid detection (e.g., acting calm or bored during the interview).
Equal Viewing	Attempting to view each image on a slide with an equal duration.
Haphazard Viewing	Using a variety of viewing patterns during the interview.
Ignoring Interview	Zoning out during the interview in an effort to avoid viewing stimuli.

Physiological Pain Manipulation	Inflicting pain on oneself to fabricate physiological responses.
Temporal Response Control	Attempting to match the response latency prior to each response.
Tone Control	Attempting to use the same tone of voice for each vocal response.
Other	Tactics comprising this category include: averting eye gaze from targets, attempting to control facial expressions, controlling eye blinks, attempting to forget target items, and matching head movements to verbal responses.

Table 34 contains a list of each countermeasure category with a corresponding value representing the relative use of countermeasures associated with a given category. The relative use of each countermeasure is represented by the percentage of countermeasure use in a given category relative to the frequency of countermeasure use in other categories. Relative category use is displayed for all participants as well as by condition. Percentages are shaded in different colors to improve visual interpretation. Values with darker shades of blue represent categories of countermeasures that were used most frequently, while values shaded in dark red represent categories of countermeasures that were not used frequently. The most prevalently used countermeasure category in any group is the use of equal-viewing countermeasures by members of the control group. The most consistently used countermeasure across conditions is the use of consistent viewing patterns to avoid detection.

Table 34 - Countermeasure Usage

Countermeasure Category	All	Control	No Targets	Targets
Blurred Viewing	3%	0%	0%	6%
Center of Screen	9%	9%	10%	9%
Consistent Viewing	20%	18%	22%	20%

Emotion Control	9%	9%	14%	6%
Equal Viewing	14%	28%	5%	15%
Haphazard Viewing	6%	9%	0%	9%
Ignoring Interview	3%	0%	10%	0%
Physiological Pain Manipulation	5%	0%	5%	6%
Temporal Response Control	11%	0%	19%	9%
Tone Control	12%	9%	5%	17%
Other	8%	18%	10%	3%

7.4.3 Novel Countermeasures

A review of Table 34 reveals that examinees utilized countermeasures that were anticipated due to the findings of prior work in this area. While only one hypothesis test of main effect differences between groups was supported, it is interesting to note that examinees reported using countermeasures related to each of these hypotheses. Examinees in each condition reported using at least one of the following countermeasures: staring at the center of the screen, controlling vocal response latencies, and engaging in physiological response manipulation using pain-inducing countermeasures[101]. The use of these countermeasures may have limited the ability to identify main effect differences between groups.

This analysis reveals that a number of other countermeasures were used. A number of these countermeasures are based on manipulations in eye gaze viewing patterns; namely, consistent viewing, equal viewing, and haphazard viewing. While viewing patterns are highly complex and likely difficult to identify, these countermeasures provide an opportunity for future research to investigate. Other novel countermeasures include the attempt to ignore the interview altogether and control emotions, supported by the use of intentional eye blurring, focusing attention

[101] It is interesting that some participants employed physical countermeasures reportedly used during polygraph examinations. This speaks to the ubiquity of knowledge about the polygraph in the general population as participants in this study were not informed of any countermeasures prior to beginning the automated interview.

elsewhere, and attempts to appear bored or calm. While these countermeasures are new, they are similar to the mental countermeasures used during polygraph examinations (e.g., difficult mental math / counting backwards from a large number). Methods of identifying differences between conditions based on efforts to ignore the interview or control emotion must first be identified before these countermeasures can be measured and identified reliably. An additional countermeasure that was used in each condition is vocal tone control. The work presented in Chapter 4 found that differences in pitch between the manipulation group and the control group were not significantly different; however, this analysis looked at differences in mean pitch, not variations in pitch over time. Future research can investigate the presence of and discriminatory power associated with differences in pitch variations between concealers and members of the control group.

7.5 Discussion

The purpose of this chapter was to provide insight on various factors influencing the human-computer interaction that occurs within an automated credibility-assessment interview. The first analysis discussed in this chapter served as an open-ended means of identifying participants' perceptions of the automated interviewing process. Each participant was asked to list five adjectives, or terms, describing the interview process. Terms were analyzed for overall use and relative use within conditions. This analysis revealed overlap in terms that were used frequently by members of all conditions (e.g., interesting, long, repetitive and easy) as well as terms used only by members of a given condition. For example, members of the control group described the interview as being intimidating and confusing. These reports imply that the system may be intimidating, regardless of the intention to conceal information, and that the task of completing an automated interview seemingly for no purpose was confusing. Members of the manipulation groups described the interview as being scary, stressful, and nerve racking, serving as a direct confirmation that experiment manipulations had succeeded. Additionally, the *presence* of target items during the screening interview appeared to alter examinee perceptions of the interviewing process for members of the targets group as they used words not reported by members of the no-targets group. Based on the findings of this analysis, displaying target items to an examinee concealing information creates a screening environment that is more stressful, intimidating, nervous, and realistic than when target items are not displayed.

The second analysis presented in this chapter investigated differences in emotional states reported by members of each condition. Previous measures of stress and arousal were used to collect data from each examinee. The results of an EFA yielded four factors on which to compare and contrast participants of differing conditions. Based on the variables associated with each factor, factors one through four were relabeled ‘Anxiety’, ‘Lethargy’, ‘Tranquility’, and ‘Enthusiasm’, respectively. An analysis of this data found that (1) members of the targets group reported higher levels of anxiety than members of the control group, (2) members of both manipulation groups reported lower levels of lethargy relative to members of the control group, (3) members of the targets group experienced a marginally lower level of tranquility as compared with members of the control group, (4) members of the targets group experienced higher levels of enthusiasm than members of the control group, and (5) there were no significant differences between members of the two manipulation groups on any of these factors. These findings, coupled with the findings reported in the previous paragraph, suggest that perceptions of the interview itself are moderated by the presence of concealed information and target items.

The third analysis presented in this chapter reported on the perceived effort and performance of examinees. Upon completing the screening interview, examinees provided responses to five different questions used to ascertain the extent to which they tried to avoid detection and how successful they perceived themselves to be at avoiding detection. An analysis of this data found that (1) it was more important for members of the manipulation groups to appear truthful and avoid suspicion relative to members of the control group, (2) both manipulation groups tried harder than the control group to convince the interviewer of their truthfulness and avoid suspicion, and (3) members of the control group and the no-targets group were confident that they had passed the interview whereas members of the targets group were less confident. This analysis directly confirms that the manipulations employed in this interview were effective as members of the manipulation groups were more concerned about appearing truthful and tried harder to avoid detection. Perhaps the most interesting finding resulting from this analysis is the perceived success rates of members of the manipulation groups. Despite concealing information, members of the no-targets group assumed that they had passed the screening. It can be inferred from this finding that members of this condition were either unaware of the system's ability to identify strategic/nonstrategic cues of deception without the use of target items, or they were not confident in the system's ability to do so. Additionally, it is interesting that members of the targets group reported decreased confidence that they had passed the interview simply because target items were presented to them. This can indirectly imply that participants in this condition had some level of confidence in the system's ability to identify concealed information.

Finally, the fourth analysis discussed in this chapter investigated the use of countermeasures by examinees to avoid detection. 48% of all participants reported using countermeasures. 25% of members in the control group reported using countermeasures. This is an intriguing finding as members of this condition were not concealing information and thus had nothing to hide. Members of the manipulation groups employed more frequent use of countermeasures as 42% of members of the no-targets group and 66% of members of the targets group reported using countermeasures, respectively. In addition to measuring the number of participants in each condition that used countermeasures, the number of countermeasures used by each participant was also evaluated. On average, members of the control, no-targets, and targets groups used 1.13, 1.33, and 1.36 distinct countermeasures, respectively. Ten categories of countermeasures were created based on the types of countermeasures that were used; countermeasures used only once were assigned to the 'Other' category. The attempt to view each image an equal number of times and the attempt to use the same pattern of viewing images were the most frequently used countermeasures reported by participants. Finally, without any training or priming, countermeasures associated with the three main-effect differences evaluated in study two were present. Interestingly, a number of participants reported using pain-inducing mechanisms to beat the system (e.g., pressing their nail into their hand). Such countermeasures are used against the polygraph, yet it is interesting that college students, likely without criminal backgrounds or criminal interviewing experience, are aware of such countermeasures. Finally, the prevalence of countermeasure use in this study is an important finding. Participants were not primed to use or trained to use countermeasures, yet a large portion of examinees used a wide range of countermeasures. This suggests a highly varied range of countermeasures that may be used to beat credibility-assessment systems in the future.

7.6 Implications

The findings presented in this chapter are compelling and provide insight on a number of aspects of automated screening systems that can be implemented, or at least considered, when conducting future research in this area.

First, the absence of concealed information, the act of concealing information, as well as the interaction effect of concealing information coupled with the presence of target items, alter how an automated credibility assessment system is perceived. Control group members reported the system as being intimidating. This is something that must be considered when designing future systems, as triggering stress and arousal in innocent persons by asking them to complete an automated interview can be problematic. Second, the way in which examinees perceive and respond to an automated screening system is influenced by whether or not the system confronts them with crime-relevant information. This translates to the need for automated screening systems to be designed to ask questions that are highly relevant to the information that is being concealed or the possible crimes that may be committed. If automated systems fail to present the examinee with relevant questions or crime-relevant information, they will likely fail to elicit behavioral and physiological differences in examinees that can be used to accurately identify deception and concealed information.

Second, the act of concealing information results in different emotional states relative to a lack of concealed information. Examinees concealing information experienced increased arousal and decreased lethargy. Additionally, members of the targets group reported lower levels of tranquility and higher levels of enthusiasm than the control group. While the measurement of emotion is admittedly complex, it may be possible to incorporate the use of interpreting facial expressions to identify the emotional states of examinees. Furthermore, vocalic features can be used to augment the analysis of facial features to create a robust emotional-state classifier. The presence or lack of various emotions could then be used in an attempt to discriminate between innocent examinees and examinees concealing information[102].

Third, a presence or lack of target items appears to influence the perception of the examinee as to whether or not they thought they had passed the screening interview. Aside from serving as a robust manipulation check for study two, this information can also be valuable from a practical standpoint. Examinees concealing information who assume that they have been identified may resort to unexpected behaviors or drastic measures to remove themselves from the screening process. If the targetless CIT approach proves feasible in future research, a valuable aspect of this approach may be the covert nature of identifying persons warranting further screening without their knowledge.

Finally, countermeasures were used by members of each condition in study two. This is troubling considering a probable scarcity of people possessing criminal backgrounds or extensive knowledge of credibility assessment interviews in the sample. A wide spectrum of countermeasures was used by participants without prior priming, training, or time to research methods that may have been effective against this type of system or sensors. The findings presented in this study are important as the countermeasures reportedly used in this study can be further investigated in future experiments in an attempt to identify the robustness of this system to each specific countermeasure. Clearly, the evaluation of countermeasure effectiveness will play a pivotal role in the potential value added of an automated credibility assessment system used in applied screening scenarios.

---

[102] It is important to reiterate that a limited number of emotions were used to collect data in study two. More informative findings could result from using a larger range of emotional states to collect self-reported emotions from examinees.

8 STUDY 2 PART III: EXPLORATION OF COUNTERMEASURE EFFECTIVENESS

A majority of the hypotheses presented in Chapter 6 were not supported. Chapter 7 contained an extensive overview of experiment participants' (1) perceptions of the interview process, (2) self-reported stress and arousal, (3) perceived effort and performance, and (4) efforts to use countermeasures during the interview. The results of analyses 1 through 3 indicated that members of both manipulation groups (1) had similar experiences during the interview, (2) used countermeasures, and (3) that their experiences were very different from members of the control group. However, relative to the no-targets group, members of the targets group perceived the interview to be more stressful, reported marginally higher levels of stress and arousal, tried harder to succeed, and were less confident that they had succeeded.

While some of these differences were not statistically significant, collectively, some of the unexpected variations between the results for the manipulation groups may be attributable to these differences. Exploring each of these variables in detail falls outside of the scope of this dissertation; however, the factor likely causing the unexpected results is the prevalent use of countermeasures. In Study 1, countermeasures were used, but sparsely (25% overall) and only by participants concealing information. Furthermore, the countermeasures that were used were largely related to emotion control (e.g., attempting to appear calm). In this study, members of each condition reportedly used countermeasures, and a majority of the countermeasures that were used were intended to mitigate eye tracking and vocalic measurements specifically.

8.1 Countermeasure Effectiveness by Hypothesis Test

It is probable that the prevalent use of countermeasures resulted in a number of the hypotheses not being supported. To further investigate this issue, hypothesis tests for H1, H2, and H3 were recalculated but with refined data sets constituted of data for participants who did not report using countermeasures. The results of each analysis are presented in the following sections.

8.1.1 Oculometric Threat Avoidance: Viewing the Center of the Screen

A restricted data set excluding participants reportedly using countermeasures was prepared. An ANOVA was specified to identify statistically significant mean differences in viewing the center of the screen between the manipulation groups and the control group. The results of the ANOVA found no statistically significant differences between the manipulation groups and the control group ($F(2, 1297) = 1.45$, $p = .24$). A Bonferroni pairwise t-test indicated that there was not a statistically significant difference between manipulation groups.

8.1.2 Pupil Dilation: Differences Between Conditions

A restricted data set excluding participants reportedly using countermeasures was prepared. An ANOVA was specified to identify statistically significant mean differences in pupil dilation between the manipulation groups and the control group. The results of the ANOVA found a statistically significant difference between the no-targets group and the control group ($t(1372) = -3.82$, $p < .001$). Members of the control group exhibited an average pupil dilation of 4.07 mm; members of the no-targets group exhibited an average *decrease* in pupil dilation of .15 mm or 3.7%. A Bonferroni pairwise t-test indicated that there was not a statistically significant difference between manipulation groups.

8.1.3 Response Latency

A restricted data set excluding participants reportedly using countermeasures was prepared. An ANOVA was specified to identify statistically significant mean differences in response latency between the manipulation groups and the control group. The results of the ANOVA found statistically significant differences between both manipulation groups and the control group ($F(2, 1522) = 13.76$, $p < .001$). Members of the control group exhibited an average response latency of 1.77 seconds; members of the no-targets group exhibited an average *decrease* in response latency of .16 seconds or 9% while members of the targets group exhibited an average *increase* in response latency of .22 seconds or 12.4%. A Bonferroni pairwise t-test indicated that there was a statistically significant difference between manipulation groups ($p < .001$).

8.2 Discussion

An analysis of the full data set for viewing the center of the screen yielded a significant difference between the no-targets group and the control group. Removing data for participants reportedly using countermeasures eliminated this significant difference. This is logical as viewing the center of the screen is a strategic method of avoiding detection that was used by members of the no-targets group more than any other condition. Thus, removing data for participants employing countermeasures should eliminate this effect between the no-targets group and the control group. There are two plausible explanations as to why members of targets group did not view the center of the screen significantly longer than members of the control group. One explanation is the extensive list of countermeasures that were used. Viewing images for consistent durations, using consistent viewing patterns, and using haphazard viewing patterns were some of the most frequently used countermeasures, thus, in this study, simply viewing the center of the screen may not have been as common, especially in the targets group. A second explanation is that participants who did not report using countermeasures may not have been as engaged in the experiment, or may not have taken the experiment seriously. If a participant does not care about the task at hand, or is distracted, they are unlikely to employ countermeasures to avoid detection.

An analysis of pupil dilation differences between groups using the full data set revealed an increase in pupil dilation in members of the targets group relative to the control group. This difference has been eliminated using the restricted data set. The only difference found in the restricted data set is a decrease in pupil dilation for the no-targets group relative to the control group. This finding follows the pattern of the previous analysis as removing participants actively engaged in the interview or in the use of countermeasures results in an apparent decrease in levels of arousal or engagement in the interview. It can be inferred that the presence of target items is a catalyst for countermeasures use, and an accompanying increase in pupil dilation. Subsequently, a lack of target items in the interview and failing to use countermeasures resulted in pupil dilation levels in the no-targets group that were significantly lower than the control group.

Finally, an analysis of response latency differences between groups using the full data set found no significant differences between any of the conditions. An analysis of the restricted data set identified significant differences between all three conditions. In the restricted data set, members of the no-targets group exhibited response latencies that averaged 9% shorter than the control group. This finding follows the pattern identified in the previous two analyses suggesting that participants who were not actively using countermeasures appeared to be disengaged from the interview. These participants likely realized that no crime-relevant information was included in the interview; these participants began responding systematically without carefully inspecting the stimuli[103]. On the other hand, members of the targets group exhibited response latencies 12.4% longer than the control group. Targets-group participants were presented with crime-relevant information and acted accordingly; they likely reviewed stimuli more thoroughly and were more careful in responding, resulting in an average response latency that was longer than the other two conditions.

8.3 Implications

It is clear that the use of countermeasures had a tremendous influence on the lack of support for the hypotheses presented in Study 2. Excluding data points for participants actively employing a variety of countermeasures yielded more significant differences between conditions than the full data set; however, these differences were not the anticipated results. It is worth reiterating that participants were not encouraged or trained to use countermeasures; however, countermeasures use was prevalent. This can be attributed to the pre-test overview of the CIT methodology and the identification of the sensors that would be used during the automated interview.

A key contribution of this work is the realization that within moments of learning about the sensors that would be used for data collection, examinees had identified tactics that they would then use to beat the system. In a field setting, potential examinees will have increased time and resources to prepare countermeasures that can be used to mitigate the effectiveness of specific sensors. Therefore, the main effect differences identified in Study 1 and tested in Study 2 are not feasible for use in a targetless CIT due to the existence of such an extensive set of alternative countermeasures and strategic means of avoiding detection. In other

[103] This is corroborated by a number of participants in the no-targets group describing the interview as being boring.

words, an examinee could be concealing information and strategically attempting to conceal that information, but not be identified using any of the three main effect differences identified in Study 1. It is possible that mechanisms can be put in place to identify various countermeasures as a means of identifying concealed information; however, future research will need to investigate this topic.

An additional implication of this work is the identification of different levels of engagement or involvement exhibited by members of each condition. For example, analysis of the restricted data set found that members of the targets group exhibited longer response latencies while members of the no-targets group exhibited shorter response latencies. It is intriguing that members of the no-targets group would drop below both the average response latency and pupil dilation exhibited by members of the control group. This suggests that members of the control group must be anticipating the presence of a stimulus that they actually recognize, placing their engagement and average behaviors between the targets group and the no-targets group. Members of the no-targets group realize that the stimuli presented to them are not relevant and disengage. It may be possible to measure cues of disengagement to identify persons concealing information.

9 LIMITATIONS, FUTURE RESEARCH, AND CONCLUSION

This dissertation is constituted of research designed to examine the feasibility of using an automated system to conduct CITs with the intent of identifying deception and concealed information. While many of the results are promising and warrant further investigation, there are a number of limitations that should be discussed. Additionally, the results of this research have yielded a number of avenues of further investigation. The following sections contain the limitations of this work, future directions for research, and the conclusion.

9.1 Limitations

The following limitations were identified in this research. Some of these limitations provide opportunities for future work.

9.1.1 Sanctioned Deception

One limitation of this work is the use of sanctioned deception. Participants in both studies were tasked with bringing a banned item through a simulated screening environment; they did not have the option to pass through the screening without the banned item (although they did have the option to discontinue the experiment at any time). Miller and Stiff (1993) identify the practice of sanctioning participants to deceive as a troubling trend in this line of research. They argue that the emotions and feelings of stress and anxiety experienced by college students in a trivial experiment are not generalizable to criminals or other examinees in real-world contexts with much higher stakes. Future research should explore the creation of more realistic experimental tasks to foster a more “high-stakes” environment.

9.1.2 Level of Familiarity

An additional limitation of this research is the weak familiarity association with target items. Participants became familiar with the target faces, object, and criminal organization only minutes before the screening interview. Such a brief exposure could foster only a minimal level of familiarity. In study one, almost one-third of guilty participants were unable to recall the name of the criminal organization for which they were tasked with smuggling the banned item. A manipulation check for the banned item was not implemented, thus no insights concerning the ability of participants to pick out that item are available. Downsizing the number of foils with salient target items from five to three (only the foils containing faces) limits the statistical power of the analyses, and ultimately, the overall findings of this research. Manipulation checks in study two found that the vast majority of participants could recall all of the target items; however, levels of familiarity with these items were weak relative to famous people, colleagues, family members, etc.

9.1.3 Sensor Accuracy

Pupil dilation data appear to be consistent and robust. Measuring vocal utterances does not require any form of calibration or configuration to yield accurate and interpretable data. However, a review of eye tracking data from both studies indicates that data collection for eye gaze fixations is not always accurate and reliable using this type of eye tracker. Eye tracking data for many participants are often riddled with missing values or values that appear to be infeasible. Some researchers have employed the use of an eye tracking apparatus that is mounted on the head of the examinee (Cook et al., 2012; Webb et al., 2009). While these devices may yield more accurate readings, they breach the criteria for automated screening as they require contact with the examinee. Until more accurate methods of collecting eye gaze fixation without contacting the examinee are identified, it may be infeasible to use fixation data to identify concealed information in automated screening.

9.1.4 Countermeasures

Countermeasures are unavoidable within the context of credibility assessment. As long as technologies are developed to identify deception and concealed information, prospective examinees will attempt to create ways in which the accuracy and effectiveness of these systems can be mitigated. A lack of support for a majority of the hypotheses in study two may largely be attributable to the frequent use of countermeasures by participants in both manipulation conditions. The use of countermeasures may have been triggered by a preliminary overview of the nature of the CIT and the types of sensors that would be used to identify concealed information during the screening interview. While the use of countermeasures may have limited the number of significant findings presented in study two, the section in Chapter 7 reporting on the types of countermeasures used can benefit future research in developing more robust methods of identifying concealed information that may be resistant to those countermeasures. Furthermore, the identification of specific countermeasures might also be used as a means of identifying attempts to conceal information.

9.2 Future Directions and Research

Future research is needed to further investigate the feasibility of using automated credibility assessment systems to identify deception and concealed information. This section presents a variety of areas in which future work can extend the research presented in this dissertation.

9.2.1 Level of Familiarity

One of the limitations of both studies was identified as the necessity for participants to memorize novel faces, objects, and organizations only minutes before being tasked with viewing the corpus of all face images, and ultimately, denying their familiarity with them. A study could be designed in which participants submit images of family members, friends, course instructors, or members of a class project group. These target images could then be mixed in with a corpus of unfamiliar faces and tested for the presence of the orienting reflex to the familiar faces. Such a design would increase the generalizability of the findings, as individuals screened at the border working for a local criminal enterprise would have higher levels of familiarity with their criminal counterparts than the level of familiarity generated in the present study.

9.2.2 Sensor Fusion Platform

While some of the hypothesized effects were supported by the data, it was previously identified that the effect sizes are not sufficient to create a classification model capable of discriminating between innocent individuals and those concealing information with an acceptable degree of accuracy. However, the strength of developing a system for automated and noncontact screening is the ability to utilize numerous sensors to measure a variety of cues and manifestations of deception and concealed information that can be used to make an accurate and reliable classification. For this work to have practical implications, future research will investigate methods to fuse data streams collected from disparate sensors to be analyzed in real time, ultimately contributing to a dichotomous classification of truth or deception or a continuous scale risk-based assessment. Finding significant differences in behavior between deceivers and truth tellers is important; however, possessing the ability to measure, interpret, and produce actionable information from this data is paramount.

9.2.3 Refined Deception Experiment

Difficulty understanding the experimental task was identified as one limitation of study one. Future work could focus on the development of a more salient paradigm within which these sensors could be evaluated more realistically. While research does suggest that orienting responses occurring in laboratory CITs are representative of those occurring in legitimate criminal investigations (Verschuere, Meijer, & De Clercq, 2011), laboratory studies designed with higher levels of realism will produce more salient behaviors and more generalizable findings. Extensive pilot testing should also be conducted prior to the core data collection to ensure that the experimental task is clearly defined. Additionally, to address the limitation of using a sanctioned-deception paradigm, an experiment allowing participants to self-select whether or not they will deceive will strengthen the generalizability of the findings.

9.2.4 Thermal Imaging Technology

Future research should also evaluate the ability of other noncontact sensors to conduct a CIT. Thermal imaging technology is relevant to deception research as it has the capacity to monitor the eyes (Pavlidis, Eberhardt, & Levine, 2002; Pavlidis & Levine, 2002) and forehead (Puri, Olson, Pavlidis, Levine, & Starren, 2005; Zhen, Tsiamyrtzis, & Pavlidis, 2007) for variations in temperature associated with differences in blood flow. The fight or flight response and the orienting response (Pavlidis & Levine, 2002) have been linked with temperature variations in the face (Sokolov, 1963a; Vendemia, 2003). A strength of using thermal imaging technology to collect data is that it can be used without contacting the examinee. Extant research reports a relatively high degree of accuracy in using this technology to identify deception (Tsiamyrtzis et al., 2007; Zhen et al., 2007). An additional strength of thermal imaging is the ability to collect thermal data discreetly without the examinee being aware of the camera (Vrij & Granhag, 2008). This could prove extremely valuable in an automated screening setting as examinees will be unable to effectively employ countermeasures if they are unaware of the sensors that are being used. Furthermore, this technology is robust and prior research suggests that it is robust to countermeasures (Tu et al., 2007). Future work can evaluate the feasibility of using this technology for automated screening, both as a single point of data collection as well as in concert with other noncontact screening tools.

9.2.5 Image Types in a Visual CIT

It is apparent in the line charts presented in 6.5.4.2 that there were variations in response latency associated with the type of image presented to the examinee. For example, response latencies associated with face images and text images of criminal organization names were similar, while response latencies associated with banned objects were shorter in duration. This is likely due to the complexity of processing faces and the time required to read text as compared with processing an image of an object. Differences in responses that are attributable to the nature of the stimuli must be identified, thus, future research can utilize various combinations of stimuli types to isolate effects attributable to the nature of the stimuli. Such information would prove critical in the formulation and interpretation of visual CITs.

9.3 Conclusion

Concealing information, one of the many forms of deception, is a pervasive phenomenon as it is present in virtually every facet of interpersonal communication. While innocuous in many cases, some instances of information concealment can have profound implications. Deception research indicates that humans, both trained and untrained, are poor detectors of deception. New technologies are needed that can augment human credibility assessments. There exists a tremendous opportunity to develop a more accurate, automated, and highly-scalable information system for conducting credibility assessments, a platform that harnesses the power of automated interviewing and data collection, fusion of data from disparate noncontact sensors, and real-time objective data analysis.

Researchers in the National Center for Border Security and Immigration (BORDERS), a Department of Homeland Security Center of Excellence, are working to develop such a system. The CIT has been identified as a feasible interviewing technique to be used by this system; however, some challenges persist which prevent the system from being operationally feasible. Research was needed to address the following areas: (a) identification of the most promising sensors that can be used to collect behavioral and physiological data, (b) determination if adaptations of the CIT can be used to reduce the preparation time and effort needed to conduct a CIT (namely, the use of a CIT not requiring the identification and implementation of target items), and (c) investigation of various aspects of human-computer interaction in automated credibility assessment interviews.

The research presented in this dissertation was conducted to investigate the following three topics: (a) compare the relative accuracy of two noncontact behaviors (oculometrics and vocalics) to EDA (the standard sensor used to conduct CITs), (b) determine the practicality of using a targetless CIT to identify concealed information, and (c) investigate various aspects of human-computer interaction in a credibility assessment context. To complete the aforementioned objectives, two experiments were conducted.

In study one, a deception experiment was designed to measure differences in the aforementioned measures based on a presence or lack of concealed information. All participants passed through a simulated screening environment in which they were asked about their familiarity with wanted criminals, banned items, and criminal organizations operating in the area. Prior to the screening, participants assigned to the “guilty” condition became familiar with the target items in each of the three categories. They were then tasked with passing through the screening interview undetected. Logistic regression models, multiple regression models, and ANOVAs were used to analyze the data. These analyses indicated that there exist significant differences in oculometric threat avoidance, response latency, and vocal quality; however, hypothesized differences in initial saccade, pupil dilation, and pitch were not significant. Supplemental analyses yielded interesting findings which merit further investigation. The ability to discriminate between those concealing information and those not concealing information was also attempted, but the classification accuracies based on this data were paltry. Related research using vocalics and oculometrics for automated CITs have reported higher rates of accuracy. Further work will be conducted to investigate new classification approaches that may yield higher accuracy rates.

A second CIT study was conducted to investigate the finding in study one that individuals concealing information attempt to control their behavior to appear innocent, and in doing so, yield main effect differences relative to the control group. In Part I of Study 2, experiment participants were subjected to an automated screening interview (similar to the experimental task utilized in study one) in which a targetless CIT was conducted. The primary purpose of this study was to determine if individuals concealing information would exhibit strategic and nonstrategic cues associated with deception, even without the use of target items. A subsequent point of interest was identifying how long main effect differences would persist despite a lack of target items in the CIT. Multiple ANOVAs were used to compare mean differences between groups for duration of time viewing the center of the screen, pupil dilation, and response latency. The only hypothesis supported was an increase in time that members of the no-targets group viewed the center of the screen relative to members of the control group. Hypotheses associated with pupil dilation, response latency, and temporal effects were not supported. A lack of support for these hypotheses is likely attributable to an insufficient level of arousal and the use of countermeasures to avoid detection. An exploratory review of other vocalic features found significant differences in mean pitch, mean vocal quality, and mean vocal intensity between the members of both manipulation groups and members of the control group. Specifically, members of the manipulation groups exhibited lower pitch, vocal quality, and vocal intensity relative to members of the control group.

Part II of Study 2 entailed an analysis conducted to investigate human-computer interaction in a credibility-assessment context; this analysis included the review of differences in the perceptions, emotions, and use of countermeasures by members of each condition in study two. These analyses indicate that members of the no-targets group were often similar to members of the targets group and members of the control group. For this reason, it is highly likely that there was not a sufficient level of arousal to induce strategic behaviors to avoid detection in some of the members of this condition. However, when strategic behaviors were employed, such a variety of tactics were used that the behaviors specific to H1, H2, and H3 were not used to an extent that resulted in significant differences between conditions.

Part III of Study 2 was conducted on a restricted data set excluding data for participants who reported using countermeasures. The results of this analysis revealed that members of the no-targets group who did not use countermeasures likely disengaged from the interview as they exhibited response latencies shorter than the control group and pupil dilations smaller than the control group. Exploratory vocalic features remained significant even with the refined data set, pointing to their robustness and promise as means of identifying concealed information with or without the use of countermeasures.

In closing, this dissertation makes a number of contributions to research in a number of domains, including: deception detection, concealed information testing, automated-screening technologies, credibility assessment, oculometrics, vocalics, human-computer interaction in automated screening, and countermeasures. Additional work is needed to better understand the theoretical underpinnings and psychophysiological phenomena driving cues of deception and concealed information, as well as the methodologies and technologies that can be feasibly and reliably used to detect and interpret them.

10 APPENDIX A: STIMULI SELECTION PILOT STUDY RESULTS

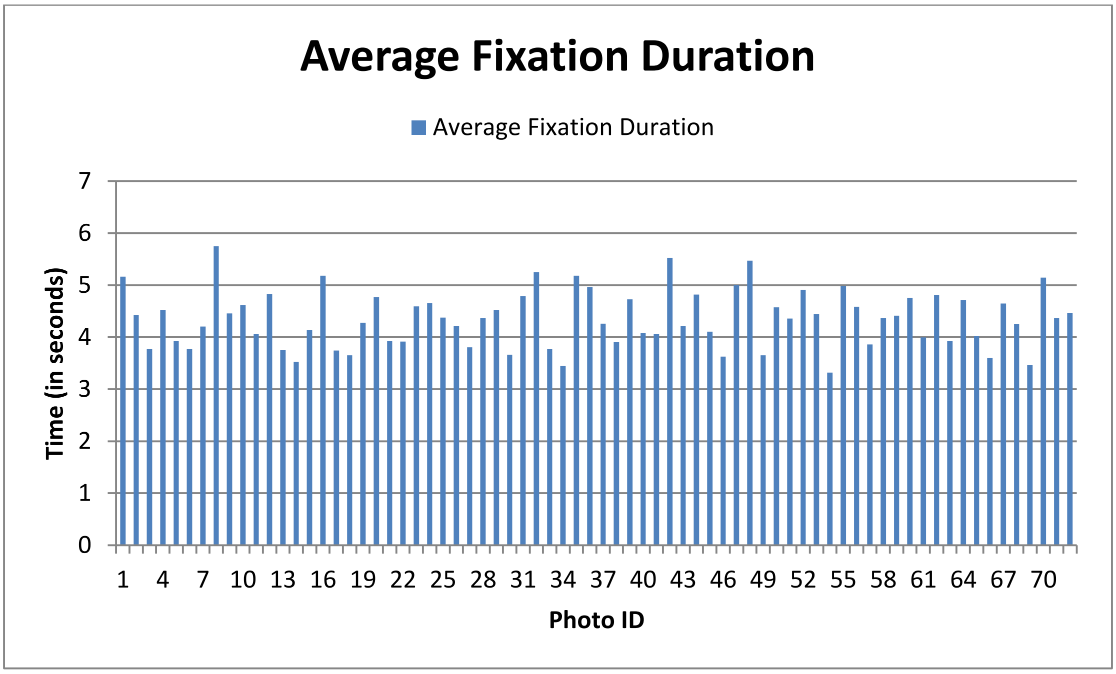


Figure 48 - Average Fixation Duration

The utilization of visual stimuli in a CIT introduces challenges for the examiner as visual stimuli contain an abundance of information and possible distractions (Krapohl et al., 2009). The pilot study was conducted to collect empirical data that could be used for stimuli selection, refinement and configuration. During the pilot, participants viewed a set of 72 face[104] images while the sensor network collected oculometric activity. The average fixation duration per face image was 4.35 seconds; the standard deviation was .547 seconds. Face images falling outside of two standard deviations from the mean were removed. Face images falling closest to the mean were designated as target items. Inter-slide variance was calculated to determine

[104] Slides containing banned items were pilot tested as part of a related experiment (Twyman, Schuetzler, et al., 2013). Slides containing criminal organizations were not pilot tested due to the homogeneity of text.

if a grouping of face images was problematic due to heterogeneity of features. Three slides with the highest inter-slide variance were removed.

11 APPENDIX B: CHAPTER 4 SUPPLEMENTARY ANALYSIS: EDA

The supplementary analysis presented in this section utilized a restricted data set comprised of data collected during the presentation of only the three foils containing images of faces. Data collected during the presentation of banned items and criminal organizations were not considered for these analyses. To view the results based on the use of the complete data set, refer to section 4.5.2.

11.1 Electrodermal Activity

Due to the results of the manipulation check analysis, accuracy, sensitivity, and specificity calculations were made using the restricted set of data. A table summarizing the accuracy, sensitivity, and specificity results for both data sets is presented in Table 35.

Table 35 - Accuracy, Sensitivity, and Specificity Results

	Complete Data Set	Restricted Data Set	Change
Accuracy	77%	78%	1%
Sensitivity	60%	74%	14%
Specificity	92%	82%	10%

With the restricted data set, accuracy improved from 77% to 78.3%. Sensitivity also improved from 60% to 74.3%; specificity decreased from 92.3% to 82%. The improved accuracy rate of 78.3% is ten percentage points below the overall estimated accuracy rate of the CIT, estimated to be roughly 88% (Krapohl, 2010).

12 APPENDIX C: CHAPTER 4 SUPPLEMENTARY ANALYSES: OCULOMETRIC HYPOTHESIS TESTS

All three of the supplementary analyses presented in this section utilized a restricted data set comprised of data collected during the presentation of only the three foils containing images of faces. Data collected during the presentation of banned items and criminal organizations were not considered for these analyses. To view the results based on the use of the complete data set, refer to section 4.5.3.

12.1 Orienting Response to Target Items

A logistic regression model was specified to measure the presence of an orienting response to target items by concealers. The results again do not support the presence of a significant orienting reflex toward the target image. Results from the model are summarized in Table 36.

Table 36 - Results of Logistic Regression Model for Orienting Response

Fixed Effects	β	β Standard Error
Intercept	-1.217***	0.232
Concealed Information	-0.463 (n.s.)	0.362

Notes: model fit by maximum likelihood. *** $p < .001$; ** $p < .01$; * $p < .05$; (n.s.) not significant.

12.2 Oculometric Threat Avoidance: Eye-Gaze Aversion from Target Items

A model (n = 4254) was specified to determine if concealers exhibited a tendency to avert eye gaze from the target item upon seeing it. This model indicates that the fixed effect measuring the interaction of concealed information and the presence of a target image resulted in a shorter duration of eye gaze (6%) on the target item ($t(4250) = -3.03$, $p < .01$). Restricting the data set resulted in a 2.1% decrease in fixation time from the model utilizing the complete data set. Results for the remaining fixed effects are similar to the previously specified model. Table 37 summarizes the results of this model.

Table 37 - Oculometric Threat Avoidance (Eye-Gaze Aversion)

Fixed Effects	β	β Standard Error	t Value
Intercept	0.265***	0.010	27.165
Concealed Information	-0.006 (n.s.)	0.013	-0.444
Target Image	0.005 (n.s.)	0.014	0.372
Position of Target Item within the Foil	0.001 (n.s.)	0.002	0.742
Concealed Information and Target Image	-0.06**	0.020	-3.030

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

12.3 Pupil Dilation: Differences between Slides

A multilevel regression model was specified to identify increases in pupil dilation exhibited by concealers when a slide containing a target item was presented, but using data from only the first three foils containing images of faces (n = 1110). The fixed effects in this second model were not significant, including the position of the target item within the foil, which *was* significant in the previous model. It can be derived from this finding that there were no significant differences in pupil dilation across target or nontarget items, or between conditions. Table 38 summarizes the results of this model.

Table 38 - Pupil Dilation (Differences between Slides)

Fixed Effects	β	β Standard Error	t Value
Intercept	2.537***	0.086	29.327
Concealed Information	-0.053 (n.s.)	0.124	-0.430
Target Image	0.034 (n.s.)	0.026	1.304
Position of Target Item within the Foil	-0.007 (n.s.)	0.005	-1.320
Concealed Information and Target Image	0.004 (n.s.)	0.037	0.117

Notes: model fit by maximum likelihood. *** $p < .001$; ** $p < .01$; * $p < .05$; (n.s.) not significant.

12.4 Conclusion

Supplementary analysis using a restricted data set confirmed the findings found in section 4.5.3. Results that were not significant using the complete data set were not rendered significant using the restricted data set. However, significant findings using the complete data set remained significant and were strengthened using the restricted data set.

13 APPENDIX D: CHAPTER 4 SUPPLEMENTARY ANALYSES: EXPLORATORY OCULOMETRIC FINDINGS

Both of the supplementary analyses presented in this section utilized a restricted data set comprised of data collected during the presentation of only the three foils containing images of faces. Data collected during the presentation of banned items and criminal organizations were not considered for these analyses. To view the results based on the use of the complete data set, refer to section 4.5.6.

13.1 Oculometric Threat Avoidance: Fixating on the Center of the Screen

A multilevel regression model (n=1110) was specified to determine if concealers had a tendency to fixate on the center of the screen as a countermeasure to reduce suspicion. A minor increase (from 6.1% in the model using the full data set to 7.2% in this model using a restricted data set) in the duration of eye-gaze fixation on the center of the screen is present ($t(1106) = 2.62$, $p < .001$). Similar to the model including data from all five foils, the remaining fixed effects were not significant. Table 39 summarizes the results of this model.

Table 39 - Oculometric Threat Avoidance (Fixation on Center of Screen)

Fixed Effects	β	β Standard Error	t Value
Intercept	0.046*	0.019	2.317
Concealed Information	0.072***	0.028	2.624
Target Image	0.004 (n.s.)	0.009	0.464
Position of Target Item within the Foil	0.002 (n.s.)	0.002	1.030
Concealed Information and Target Image	0.009 (n.s.)	0.013	0.658

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

13.2 Pupil Dilation: Differences within Slides

A multilevel regression model (n = 4254) was specified to determine if concealers exhibited significant differences in pupil dilation when viewing a single slide containing a target item. A minor increase (from .092 mm. or 3.64%, in the previous model to 0.101 mm, or 3.78%, in this model) in pupil dilation is present ($t(4250) = 2.38$, $p < .05$). Similar to the model including data from all five foils, the remaining fixed effects were not significant. Table 40 summarizes the results of this model.

Table 40 - Pupil Dilation (Differences within Slides)

Fixed Effects	β	β Standard Error	t Value
Intercept	2.643***	0.082	32.140
Concealed Information	-0.114 (n.s.)	0.119	-0.960
Target Image	0.007 (n.s.)	0.029	0.250
Concealed Information and Target Image	0.101*	0.043	2.380

Notes: model fit by maximum likelihood. *** p < .001; ** p < .01; * p < .05; (n.s.) not significant.

13.3 Conclusion

Supplementary analysis using a restricted data set only confirmed the findings found in section 4.5.6. Significant findings using the complete data set remained significant and were strengthened using the restricted data set.

14 APPENDIX E: PERCEPTIONS OF THE INTERVIEW PROCESS WORD CLOUDS

14.1 Word Cloud for Control Group

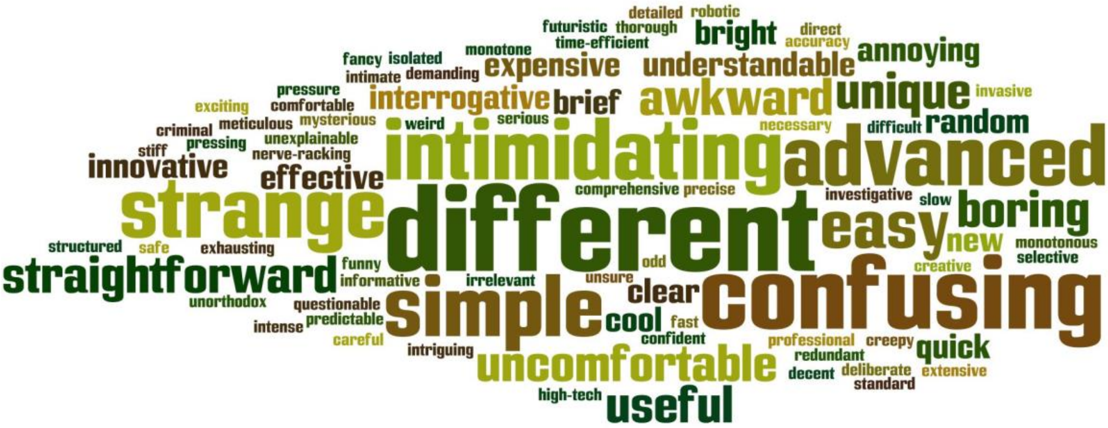


Figure 49 - Word Cloud for Control Group

14.2 Word Cloud for Manipulation Groups

Figure 50 - Word Cloud for Manipulation Groups

14.3 Word Cloud for No-Targets Group

Figure 51 - Word Cloud for No-Targets Group

14.4 Word Cloud for Targets Group

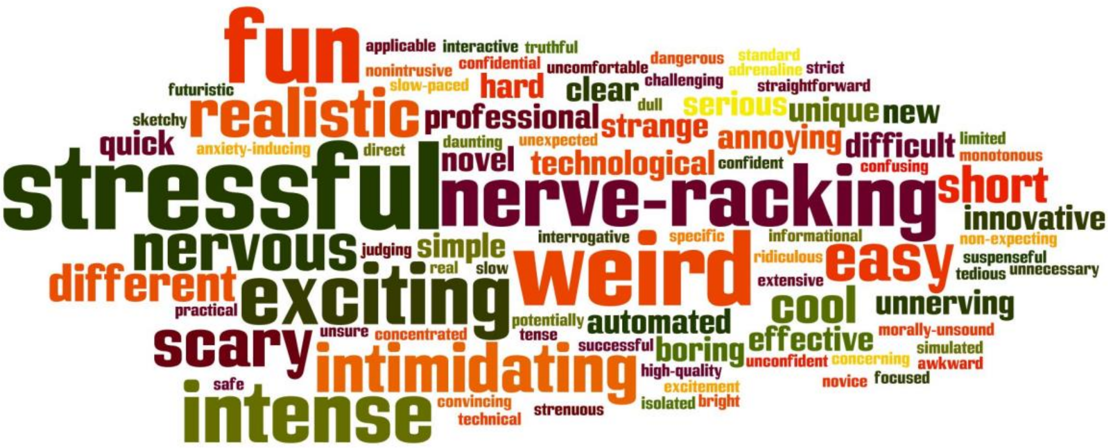


Figure 52 - Word Cloud for Targets Group

15 APPENDIX F: STRESS AND AROUSAL EFA: KMO TEST RESULTS

15.1 Results of KMO1

KMO1 Overall: 0.775542

Individual

Up.tight	0.881	Jittery	0.813	Restful	0.731
Stirred.Up	0.878	Bothered	0.811	At.rest	0.731
Calm	0.875	Comfortable	0.809	Tired	0.721
Distressed	0.866	Aroused	0.797	Placid	0.719
Fearful	0.849	Active	0.793	Drowsy	0.705
Uneasy	0.848	Stimulated	0.787	Leisurely	0.685
Nervous	0.843	Energetic	0.786	Dejected	0.672
Lively	0.839	Passive	0.776	Insensitive	0.654
Vigorous	0.828	Apprehensive	0.774	Sleepy	0.649
Intense	0.828	Activated	0.765	Idle	0.649
Worried	0.828	Sluggish	0.752	Cheerful	0.591
Relaxed	0.828	Excited	0.751	Quiet	0.572
Alert	0.824	Contented	0.750	Somnolent	0.539
Tense	0.818	Pleasant	0.745	Still	0.520
Peaceful	0.818	Expectant	0.739	Quiescent	0.475

15.2 Results of KMO2

KMO2 Overall:
0.803672

individual

Stirred.Up	0.891	Tense	0.835	Restful	0.767
Fearful	0.889	Vigorous	0.831	Expectant	0.761
Up.tight	0.881	Active	0.830	Tired	0.744
Calm	0.874	Apprehensive	0.811	Excited	0.741
Intense	0.873	Sluggish	0.810	Leisurely	0.725
Relaxed	0.868	Stimulated	0.803	At.rest	0.724
Jittery	0.864	Bothered	0.803	Drowsy	0.699
Distressed	0.857	Energetic	0.802	Dejected	0.687
Alert	0.855	Peaceful	0.794	Placid	0.685
Worried	0.855	Contented	0.792	Sleepy	0.667
Uneasy	0.852	Activated	0.787	Insensitive	0.649
Comfortable	0.846	Aroused	0.787	Cheerful	0.635

Nervous	0.844	Pleasant	0.784	Idle	0.632
Lively	0.838	Passive	0.769	Quiet	0.628

16 APPENDIX G: LIST OF COUNTERMEASURES USED

Condition	Countermeasure(s)
Control	After looking at the 4 subjects and saying "no," I tried not to look back at any of them for too long, or I tried to stare at the middle of the screen.
Control	I changed the order I looked in each of the four directions.
Control	I never focused too long on a single picture or object and tried to look at each picture the same amount.
Control	I tried to look around at every face the same amount of time. I also tried to stop shaking my head because I felt the need to shake my head and say no but I wanted to see if I could say no without shaking my head.
Control	I kept the same eye movements throughout reading and looking at the faces in the same pattern. I also tried not to look at any one image too long, and answered questions firmly.
Control	Shook my head, made my facial features agree with my words.
Control	Tried to answer with the same tone.
Control	Tried to look only at the monitor.
No Target	Acted bored.
No Target	Before interviewing, forgot everything I memorized before.
No Target	Looking at all the pictures an equal number of times and for roughly the same amount of time. Appearing to get annoyed and answering faster. Looking at pictures I knew were wrong a little shorter or longer.
No Target	I rarely moved my eyes for any of the images I was shown. I just looked at the center of the screen.
No Target	Tried not to pay close attention so I actually wouldn't recognize any of the images.
No Target	Make a point to at least look at all four images on each slide, to look like I'm actively participating and not just blindly stating "no".
No Target	Had fictional purpose of trip and bag contents pre-planned in case of questioning and role-played that story internally. After two screens of questions about involvement in cartels, realized I wasn't involved in any cartels at all and answered "No" without really looking at the names since the answer is no by default. Acted like I always do at TSA screening procedures where I have an academic interest in methodologies; tried to find the fun in being shown mug shots of interesting-looking people.
No Target	Tried to keep the same facial demeanor and not have a consistent tone of voice. Acted bored with it so that I wouldn't make any sudden moves or be jittery.
No Target	I tried not to look at one image for too long, and once I answered "yes" or "no", I would look at the middle of the screen.
No Target	Squeezed my thumb so I could feel a small pain and not look like I was lying.

No Target	I kept the same eye pattern as I looked at the different pictures on the screen every time, moving from the top left in a clockwise circle each time.
No Target	Used the same eye motions for every set of pictures, not replying until I had looked at all 4, replying at about the same interval, blinking at the same interval, using the same tone of voice, swallowing at random intervals.
No Target	Took the same amount of time to answer on every question during the screening.
No Target	Appeared calm.
No Target	Look at all the pictures in a general order and not focus on any one too long.
No Target	I just tried to stay calm and to act normal for every picture shown to me. Trying not to show any variation in my reactions or time to answer the questions.
No Target	I looked at all of the pictures at the same rate.
No Target	Control speed of answers.
Target	Tried to dig a nail into my hand to distract my mind from the interview.
Target	Looked inquiringly at each image with a skeptical look, tried to be patient, "calm" and respond with obvious focus and consideration.
Target	Rapid eye movement and inconsistency in my period between time lapses in my answers.
Target	Relaxed.
Target	Quickly looked at the pictures and then stared at the screen without focus.
Target	I blurred my vision so that the pictures would be hazy as I looked over them.
Target	I tried to keep myself aware of how my eyes would move from image to image and then always track in the same direction and speed.
Target	I would try to view the pictures in the same order. And continue the same eye motions for every image.
Target	Changed the pitch of my voice to tried and sound the same. Tried to stick to patterns or random movements with my eyes.
Target	I looked at each photo for the same amount of time. I made sure not to show that I knew any of the people.
Target	I tried to use the same tone in every answer.
Target	I tried to respond similarly to every question.
Target	I used the same eye movement patterns throughout.
Target	Conformed all my answers and eye movement in the same manner.
Target	Answered the question at the same pace.
Target	I tried to say "no" after the same amount of time had passed during each round of pictures. I also tried to look at the people again even if I did recognize them in order to not look like I was avoiding them.
Target	Looked straight the entire time.

Target	I tried to keep my voice even and the same during lies as during truths. I tried to avoid looking at the item that I was carrying (or person I recognized) when they appeared on the screen. Instead I focused on the other things because saying no to those would be true, and may make my lie more believable.
Target	Press my nails into my palm during every question
Target	I tried to use the same eye movements every time when looking at each screen. I think I got better as the interview went along, but I don't think I did that well overall.
Target	I tried to look at all possible answers then stare at the middle of screen. I didn't look at an image twice and also just said yes or no.
Target	Tried to keep my eyes from wandering.
Target	Responded the same way to every question, and I did not always look at all the pictures.
Target	I tried to use the same eye motion as I looked at each picture.
Target	I tried to use the same process of viewing the pictures every time. I started at the top and went clockwise around the pictures. After every step I starred at the middle of the screen.

REFERENCES

Abercrombie, H., Chambers, A., Greischar, L., & Monticelli, R. (2008). Orienting, emotion, and memory: Phasic and tonic variations in heart rate predicts memory for emotional pictures in men. *Neurobiology of Learning and Memory*, *90*(4), 644–650.

Abrams, S. (1989). *The Complete Polygraph Handbook*. Lexington, MA: Lexington Books.

Adachi, K., & Suzuki, A. (1991). A comparison of detectability of deception among physiological measures. *Japanese Journal of Applied Psychology*, *16*, 33–43.

Aftergood, S. (2000). Polygraph testing and the DOE national laboratories. *Science*, *290*(5493), 939–940.

Allen, J. J. B. (2011). Clinical applications of the Concealed Information Test. In B. Verschuere, G. Ben-Shakhar, & E. H. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 220–240). Cambridge: Cambridge University Press.

Allen, J. J. B., & Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, *34*, 234–240.

Allen, J. J. B., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology*, *29*, 504–522.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010.

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258–267.

Ambach, W., Stark, R., Peper, M., & Vaitl, D. (2008). An interfering Go/No-go task does not affect accuracy in a Concealed Information Test. *International Journal of Psychophysiology*, *68*(1), 6–16.

Ansley, N. (1992). The history and accuracy of guilty knowledge and peak of tension tests. *Polygraph*, *21*(3), 174–247.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*(5), 715–727.

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219–224.

Baracaldo, N., & Joshi, J. (2012). A trust-and-risk aware RBAC framework: Tackling insider threat (pp. 167–176). Presented at the Proceedings of the 17th ACM symposium on Access Control Models and Technologies, ACM.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*, 371–372.

Ben-Shakhar, G. (1994). The roles of stimulus novelty and significance in determining the electrodermal orienting response: Interactive versus additive approaches. *Psychophysiology*, *31*(4), 402–411.

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). Cambridge: Cambridge University Press.

Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior*, *26*(5), 527–541.

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

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.

Ben-Shakhar, G., Lieblich, I., & Kugelmass, S. (1975). Detection of information and GSR habituation: An attempt to derive detection efficiency from two habituation curves. *Psychophysiology*, *12*(3), 283–288.

Ben-Shakhar, G., & Meijer, E. (2012). Skin conductance, respiration, heart rate, and P300 in the concealed information test: a meta analysis. *International Journal of Psychophysiology*, *85*, 324.

Berger, C. R., Karol, S. H., & Jordan, J. M. (1989). When a lot of knowledge is a dangerous thing: The debilitating effects of plan complexity on verbal fluency. *Human Communication Research*, *16*, 91–119.

Blair, J. P., & Kooi, B. (2004). The gap between training and research in the detection of deception. *Journal of Police Science & Management*, *6*(2), 77–83.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound (Vol. 17, pp. 97–110). Presented at the Proceedings of the Institute of Phonetic Sciences, Amsterdam.

Bond, J. C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). New York, NY: Springer.

Boyle, R., Clements, J. A., & Proudfoot, J. G. (2012). Predicting deceptive behavior: Exploring antecedents to deceptive behaviors. Presented at the National Communication Association Convention, Communication and Social Cognition Division, Orlando, FL.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Bradley, M. T., & Ainsworth, D. (1984). Alcohol and the psychophysiological detection of deception. *Psychophysiology*, *21*(1), 63–71.

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 176–188). Cambridge: Cambridge University Press.

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology*, *18*(3), 307–315.

Brenner, M., Branscomb, H. H., & Schwartz, G. E. (1979). Psychological stress evaluator—two tests of a vocal measure. *Psychophysiology*, *16*(4), 351–357.

Brown, T. E., Senter, S. M., & Ryan Jr, A. H. (2003). Ability of the Vericator to detect smugglers at a mock security checkpoint. *International Journal of Psychophysiology*, *45*, 28.

Buller, D. B., & Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. In J. A. Daly & J. M. Wiemann (Eds.), *Strategic Interpersonal Communication* (pp. 191–223). Hillsdale, N.J.: Erlbaum.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Deception Theory. *Communication Theory*, *6*(3), 203–242.

Burgoon, J. K. (1996). Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research*, *23*(6), 724–48.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*(4), 572–599.

Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. A. (1994). Interpersonal Deception: V. Accuracy in deception detection. *Communication Monographs*, *61*(4), 303–25.

Burgoon, J. K., Derrick, D. C., Elkins, A. C., LaMarc Humphreys, S., Jensen, M. L., Diller, C. B. R., & Nunamaker Jr., J. F. (2008). Potential noncontact tools for rapid credibility assessment from physiological and behavioral cues (pp. 150–157). Presented at the 42nd Annual IEEE International Carnahan Conference on Security Technology, IEEE.

Burgoon, J. K., Fang, C., & Twitchell, D. P. (2010). Deception and its detection under synchronous and asynchronous computer-mediated communication. *Group Decision & Negotiation*, *19*(4), 345–366.

Burgoon, J. K., & Floyd, K. (2000). Testing for the motivation impairment effect during deceptive and truthful interaction. *Western Journal of Communication*, *64*(3), 243–267.

Burgoon, J. K., & Nunamaker Jr., J. F. (2004). Toward computer-aided support for the detection of deception. *Group Decision and Negotiation*, *13*(1), 1–4.

Burgoon, J. K., & Qin, T. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, *25*(1), 76–96.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validition by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Cannon, W. B. (1929). *Bodily changes in pain, hunger, fear and rage: An account of recent research into the function of emotional excitement*. New York, U.S.A.: Appleton-Century-Crofts.

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments. *Journal of Experimental Psychology-Applied*, *9*(4), 261–269.

Cestaro, V. L., & Dollins, A. B. (1994). *An analysis of voice responses for the detection of deception*. Ft. McClellan, Alabama: Department of Defense Polygraph Institute.

Christie, M. J. (1981). Electrodermal activity in the 1980s: a review. *Journal of the Royal Science of Medicine*, *74*, 616–622.

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S., Woltz, D. J., & Kircher, J. C. (2012). Lyin' eyes: Ocular-motor measures of reading reveal deception. *Journal of Experimental Psychology: Applied*. doi:10.1037/a0028307

Crane, H. W. (1915). A study in association reaction with an attempted application of results in determining the presence of guilty knowledge. *Psychological Monographs*, *18*(4), 1–61.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Davidson, P. O. (1968). Validity of the guilty-knowledge technique: The effects of motivation. *Journal of Applied Psychology*, *52*, 62–65.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203–243.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, *12*(3), 177–202.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74.

Derrick, D. C. (2011). *Special-Purpose, Embodied Conversational Intelligence with Environmental Sensors (SPECIES) Agents: Implemented in an Automated Interviewing Kiosk* (Dissertation). University of Arizona, Tucson, AZ.

Derrick, D. C., Elkins, A. C., Burgoon, J. K., Nunamaker Jr., J. F., & Zeng, D. (2010). Border security credibility assessments via heterogeneous sensor fusion. *Intelligent Systems, IEEE*, *25*(3), 41–49.

Derrick, D. C., Moffit, K., & Nunamaker Jr., J. F. (2011). Eye gaze behavior as a Guilty Knowledge Test: Initial exploration for use in automated, kiosk-based screening. Presented at the Hawaii International Conference on System Sciences.

Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York: Norton.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*, 88–106.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *29*(3), 288–298.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, *10*(3), 263–266.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, *75*(5), 521–529.

Elaad, E. (1994). The accuracy of human decisions and objective measurements in psychophysiological detection of knowledge. *The Journal of Psychology*, *128*(3), 267–280.

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology*, *26*(4), 442–451.

Elaad, E., & Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the Guilty Knowledge Test. *International Journal of Psychophysiology*, *11*(2), 99–108.

Elaad, E., & Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *International Journal of Psychophysiology*, *61*, 226–234.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*(5), 757–767.

Elkins, A. C. (2011). *Vocalic Markers of Deception and Cognitive Dissonance for Automated Emotion Detection Systems* (Dissertation). University of Arizona, Tucson.

Elkins, A. C., Derrick, D. C., & Gariup, M. (2012). The voice and eye gaze behavior of an imposter: Automated interviewing and detection for rapid screening at the border. Presented at the Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France.

Elkins, A. C., & Stone, J. (2011). The effect of cognitive dissonance on argument language and vocalics. Presented at the Forty-Fourth Annual Hawaii International Conference on System Sciences, Kauai, HI.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features. *Perception*, *8*, 431–439.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (lie detection) with event-related brain potentials. *Psychophysiology*, *28*(5), 531–547.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning*, *31*, 1–38.

Fiedler, K., Schmid, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology*, *24*(4), 313–324.

Gamer, M. (2011). Detecting concealed information using autonomic measures. *Memory Detection: Theory and Application of the Concealed Information Test*, 27–45.

Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Human Brain Mapping*, *28*(12), 1287–301.

Gamer, M., Rill, H. G., Vossel, G., & Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology*, *60*(1), 76–87.

Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior*, *95*(3), 333–340.

Gardner, R. M., Beltramo, J. S., & Krinsky, R. (1975). Pupillary changes during encoding, storing, and retrieval of information. *Perceptual and Motor Skills*, *41*(3), 951–955.

Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*, *77*(5), 340–355.

Graham, F. K. (1979). Differentiating between orienting, defense, and startle reflex. In H. D. Kimmel, E. H. van Olst, & J. H. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 137–167). Hillsdale, N.J.

Grey, S. J., & Smith, B. L. (1984). A comparison between commercially available electrode gels and purpose-made gel, in the measurement of electrodermal activity. *Psychophysiology*, *21*(5), 551–557.

Gronau, N., Ben-Shakhar, G., & Cohen, A. (2005). Behavioral and physiological measures in the detection of concealed information. *Journal of Applied Psychology*, *90*(1), 147–158.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, *143*, 29–36.

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Ryan, J. D., & Cohen, N. J. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*.

Harnsberger, J. D., Hollien, H., Martin, C. A., & Hollien, K. A. (2009). Stress and deception in speech: Evaluating layered voice analysis. *Journal of Forensic Sciences*, *54*(3), 642–650.
Hartwig, M., & Bond Jr., C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659.
Hartwig, M., Granhag, P., Strömwall, L., & Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, *7*(4), 429–456.
Heaver, B. (2011). *Psychophysiological Indices of Recognition Memory*. University of Sussex, Brighton, England.
Hevner, A. R., & Chatterjee, S. (2011). *Design research in information systems: Theory and practice*. New York: Springer.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.
Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Application of the guilty knowledge test in forensic investigations. *International Journal of Police Science & Management*, *4*, 16–27.
Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communication Monographs*, *47*, 120–131.
Hollien, H., Harnsberger, J. D., Martin, C. A., & Hollien, K. A. (2008). Evaluation of the NITV CVSA. *Journal of Forensic Sciences*, *53*(1), 183–193.
Honts, C. R., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, *79*(2), 252–259.
Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*, *1*(3), 241–247.
Honts, C. R., Raskin, D. C., & Kircher, J. C. (2002). The scientific status of research on polygraph techniques: The case for polygraph tests. In *Modern Scientific Evidence: The Law and Science of Expert Testimony, Volume 2* (pp. 446–483). St. Paul, Minnesota: West.
Horvath, F. (1978). An experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception. *Journal of Applied Psychology*, *63*(3), 338–344.
Horvath, F. (1979). Effect of different motivational instructions on detection of deception with the psychological stress evaluator and the galvanic skin response. *Journal of Applied Psychology*, *64*(3), 323–330.
Horvath, F., Blair, J. P., & Buckley, J. P. (2008). The Behavioral Analysis Interview: Clarifying the practice, theory, and understanding of its use and effectiveness. *International Journal of Police Science and Management*, *10*(1), 101–118.
Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal of Forensic Sciences*, *39*, 793.
Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage.
Iacono, W. G. (2000). The detection of deception. In *Handbook of Psychophysiology* (2nd ed., pp. 772–793). Cambridge, UK: Cambridge University Press.
Iacono, W. G., & Lykken, D. T. (2002). The scientific status of research on polygraph techniques: The case against polygraph tests. In *Modern scientific evidence:*

The Law and Science of Expert Testimony (Vol. 2, pp. 582–618). St. Paul, MN: West.

Janisse, M. P. (1977). *Pupillometry: The Psychology of the Pupillary Response*. Washington D.C.: Hemisphere Publishing Corporation.

Janisse, M. P., & Bradley, M. T. (1980). Deception, information and the pupillary response. *Perceptual and Motor Skills*, *50*(3), 748–750.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401–415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Kline, P. (1999). *Handbook of Psychological Testing (2nd ed.)*. London: Routledge.

Kozel, F. A., Padgett, T. M., & George, M. S. (2004). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, *118*(4), 852–856.

Krapohl, D. J. (2010). Concealed Information Test. Stoelting.

Krapohl, D. J., McCloughlan, J. B., & Senter, S. M. (2009). How to use the Concealed Information Test. *Polygraph*, *35*(3), 34–49.

Krapohl, D. J., & Velez, Y. (2001). A polygraph technique for evidentiary applications. *Polygraph*, *30*(1), 43–46.

Kurohara, A., Kensuke, T., Hiromi, T., & Akio, U. (2001). Respiratory changes during detection of deception: Mechanisms underlying inhibitory breathing in response to critical questions. *Japanese Journal of Physiological Psychology and Psychophysiology*, *19*, 75–86.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*(4), 262–272.

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., & Childress, A. R. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *NeuroImage*, *15*, 727–732.

Lippold, O. (1970). Oscillation in the stretch reflex arc and the origin of the rhythmical 8-12 C/S component of physiological tremor. *The Journal of Physiology*, *137*, 473–487.

Lippold, O. (1971). Physiological tremor. *Scientific American*, *224*(3), 65–73.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual Guilty Knowledge Test: New technique for the detection of deception. *Journal of Experimental Psychology-Applied*, *2*(2), 164–177.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388.

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*(4), 258–262.

Lykken, D. T. (1974). Psychology and the lie detector industry. *The American Psychologist*, *29*(10), 725–39.

Lykken, D. T. (1991). The lie detector controversy: An alternative solution. *Advances in Psychophysiology: A Research Annual*, *4*, 209–214.

Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York, NY: Plenum Trade.

Mackay, C., Cox, T., Burrows, G., & Lazzenini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social and Clinical Psychology*, *17*(3), 283–284.

MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology*, *86*(4), 674–683.

Matsuda, I., Nittono, H., & Allen, J. J. B. (2012). The current and future status of the Concealed Information Test for field use. *Frontiers in Psychology*, *3*.

Matte, J. A. (1996). *Forensic Psychophysiology Using the Polygraph*. Williamsville, NY: J.A.M. Publications.
Maw, N. N., & Pomplun, M. (2004). Studying human face recognition with the gaze-contingent window technique (pp. 927–932). Presented at the Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society, Cognitive Science Society.
Meijer, E. H., & Verschuere, B. (2010). The polygraph and the detection of deception. *Journal of Forensic Psychology Practice*, *10*(4), 325–338.
Meijer, E. H., Verschuere, B., & Ben-Shakhar, G. (2011). Practical guidelines for developing a CIT. In B. Verschuere, G. Ben-Shakhar, & E. H. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 293–302). Cambridge: Cambridge University Press.
Miller, G. R., & Stiff, J. B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage Publications, Inc.
Mossman, D., & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *3*, 330–333.
Münsterberg, H. (1908). *On the Witness Stand*. New York: The McClure Company.
Nachshon, I., & Feldman, B. (1980). Vocal indices of psychological stress: A validation study of the psychological stress evaluator. *Journal of Police Science Administration*, *8*, 40–52.
Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*, 733–744.
Nakayama, M. (2002). Practical use of the Concealed Information Test for criminal investigation in Japan. In *Handbook of Polygraph Testing*. San Diego, CA: Academic Press.
Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678.
National Research Council. (2003). *The polygraph and lie detection*. Washington, D.C.: National Academies Press.
Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175.
Nose, I., Murai, J., & Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *Neuroimage*, *44*, 1380–1386.
Nowlis, V. (1965). Research with the mood adjective checklist. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, Cognition, and Personality*. New York: Springer.
Nunamaker Jr., J. F., Burgoon, J. K., Twyman, N. W., Proudfoot, J. G., Schuetzler, R., & Giboney, J. S. (2012). Establishing a foundation for automated human credibility screening. Presented at the IEEE International Conference on Intelligence and Security Informatics (ISI), Washington, D.C.
Nunamaker Jr., J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, *28*(1), 17–48.
Nunamaker Jr., J. F., Elkins, A. C., Twyman, N. W., & Derrick, D. C. (2012). AVATAR system for automated credibility assessment and interviewing. Presented at the IEEE International Conference on Intelligence and Security Informatics, Washington, D.C.

Nunamaker Jr., J. F., Twyman, N. W., & Giboney, J. S. (2013). Breaking out of the design science box: High-value impact through multidisciplinary design science programs of research. Presented at the Americas Conference on Information Systems, Chicago, IL.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Nunnally, J. C. ., Knott, P. D., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation. *Attention, Perception, & Psychophysics*, *2*(4), 149–155.

O'Gorman, J. G. (1979). The orienting reflex: Novelty or significance? *Psychophysiology*, *16*, 253–262.

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* (pp. 253–275). Cambridge: Cambridge University Press.

Patrick, C. J. (2011). Science on the rise: birth and development of the Concealed Information Test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press.

Pavlidis, I., Eberhardt, N. L., & Levine, J. A. (2002). Seeing through the face of deception. *Nature*, *415*(6867), 35.

Pavlidis, I., & Levine, J. (2002). Thermal facial screening for deception detection. Presented at the 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society.

Pavlov, I. V. (1947). *Complete Collection of Works, IV, 351*. Moscow-Leningrad, USSR.

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.

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, *15*(4), 344–359.

Podlesny, J. A., Raskin, D. C., & Barland, G. H. (1976). *Effectiveness of techniques and physiological measures in the detection of deception* (No. 76-5). University of Utah: National Institute of Law Enforcement and Justice, Law Enforcement Assistance Administration, U.S. Department of Justice.

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Proudfoot, J. G., Twyman, N. W., & Burgoon, J. K. (2012). A comparison of invasive and noninvasive sensors in the Concealed Information Test. Presented at the European Intelligence and Security Informatics Conference, Odense, Denmark.

Proudfoot, J. G., Twyman, N. W., & Burgoon, J. K. (2013). Eye tracking and the CIT: Utilizing oculometric cues to identify familiarity with wanted persons. Presented at the 46th Annual Hawaii International Conference on System Sciences, Maui, HI.

Puri, C., Olson, L., Pavlidis, I., Levine, J., & Starren, J. (2005). StressCam: Noncontact measurement of users' emotional states through thermal imaging (pp. 1725–1728). Presented at the Supplement to Conference on Human Factors in Computing Systems, New York, NY: ACM.

Randazzo, M. R., Keeney, M., Kowalski, E., Cappelli, D., & Moore, A. (2004). *Insider threat study: Illicit cyber activity in the banking and finance sector*. Joint SEI and U.S. Secret Service Report.

Reid, J. E., & Inbau, F. E. (1977). *Truth and Deception: The Polygraph ("lie detector") Technique (2nd ed.)*. Baltimore, MD: Williams & Wilkins.

Rockwell, P., Buller, D. B., & Burgoon, J. K. (1997). Measurement of deceptive voices: Comparing acoustic and perceptual data. *Applied Psycholinguistics*, *18*, 471–484.

Rosenthal, R. (1982). Handbook of Methods in Nonverbal Behavior. In P. Ekman & K. Scherer (Eds.), . New York: Cambridge.

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15*(5), 508–525.

Ryan Jr., A. H., Pavlidis, I., Rohrbaugh, J. W., Marchak, F., & Kozel, F. A. (2003). Credibility assessments: Operational issues and technology impact for law enforcement applications. In E. M. Carapezza (Ed.), (Vol. 5071, pp. 168–182). SPIE. Retrieved from http://link.aip.org/link/?PSI/5071/168/1

Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. *Psychology, Public Policy, and Law*, *5*(1), 203–223.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist*, *40*(3), 355.

Schwedes, C., & Wentura, D. (2011). The revealing glance: Eye gaze behavior to concealed information. *Memory & Cognition*, 1–10.

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

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge." *Journal of Applied Psychology*, *85*(1), 30–37.

Siddle, D. A. T., Kyriacou, C., Heron, P. A., & Mathews, W. A. (1979). Effects of change in verbal stimuli on the skin conductance response component of the orienting response. *Psychophysiology*, *16*, 34–40.

Sokolov, E. N. (1963a). Higher nervous functions; the orienting reflex. *Annual Review of Physiology*, *25*, 545–80.

Sokolov, E. N. (1963b). *Perception and the Conditioned Reflex*. New York: Macmillan.

Sokolov, E. N. (1966). Orienting reflex as information regulator. In A. Leontyev, A. Luria, & Smirnov (Eds.), *Psychological Research in U.S.S.R.* (pp. 334–360).

Sommers, M. S., Brown, T. E., Senter, S. M., & Ryan, A. H. (2002). Evaluating the reliability and validity of Vericator as a voice-based measure of deception. *International Journal of Psychophysiology*, *45*, 28.

Somoza, E., & Mossman, D. (1991). Neuropsychiatric decision making: Designing nonbinary diagnostic tests. *Journal of Neuropsychiatry*, *3*, 197–200.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology*, *96*(4), 407–422.

Suzuki, R., Nakayama, M., & Furedy, J. J. (2004). Specific and reactive sensitivities of skin resistance response and respiratory apnea in a Japanese Concealed Information Test (CIT) of criminal guilt. *Canadian Journal of Behavioural Science*, *36*(3), 202.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, *20*, 663–678.
Thayer, R. E. (1970). Activation states as assessed by verbal report and four psychophysiological variables. *Psychophysiology*, *7*, 86–94.
Thayer, R. E. (1971). *Studies of controlled self-reports of activation* (No. MH-14248-01). National Institute of Mental Health Public Health Service.
Thayer, R. E. (1975). Major factors of the Activation-Deactivation Adjective Checklist.
Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I., Frank, M. G., & Ekman, P. (2007). Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, *71*(2), 197–214.
Tu, P. H., Doretto, G., Krahnstoever, N. O., Perera, A. G. A., Wheeler, F. W., Liu, X., & Harding, K. G. (2007). An intelligent video framework for homeland protection. *The International Society for Optical Engineering*, 6562.
Twyman, N. W. (2012). *Automated Human Screening for Detecting Concealed Knowledge* (Dissertation). University of Arizona, Tucson, AZ.
Twyman, N. W., Burgoon, J. K., Elkins, A. C., & Proudfoot, J. G. (2013). Alternative cues in concealed information testing. Presented at the 46th Annual Hawaii International Conference on System Sciences, Maui, HI.
Twyman, N. W., Schuetzler, R., Proudfoot, J. G., & Elkins, A. C. (2013). A systems approach to countermeasures in credibility assessment interviews. Presented at the International Conference on Information Systems, Milan, Italy.
Van Damme, G. (2001). Forensic criminology and psychophysiology: Truth verification tools, with a special study of TrusterPro. *Crime Research in South Africa*, *2*(2).
Vandenbosch, K., Verschuere, B., Crombez, G., & Clerq, A. D. (2009). The validity of finger pulse line length for the detection of concealed information. *International Journal of Psychophysiology*, *71*, 118–123.
Vendemia, J. M. C. (2003). Detection of deception. *Polygraph*, *32*(2), 97–106.
Verschuere, B. (2011). Psychopathy and the detection of concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 204–219). Cambridge: Cambridge University Press.
Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: Theory and application of the Concealed Information Test*. New York, NY: Cambridge University Press.
Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. H. W. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology*, *41*(3), 461–466.
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.
Verschuere, B., Crombez, G., Koster, E. H. W., & De Clercq, A. (2007). Antisociality, underarousal and the validity of the Concealed Information Polygraph Test. *Biological Psychology*, *74*(3), 309–318.
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). Cambridge: Cambridge University Press.

Verschuere, B., Meijer, E., & De Clercq, A. (2011). Concealed information under stress: A test of the orienting theory in real-life police interrogations. *Legal and Criminological Psychology*, *16*(2), 348–356.

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2nd ed.). West Sussex, England: John Wiley & Sons.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*(4), 141–142.

Vrij, A., & Granhag, P. A. (2008). Interviewing to Detect Deception. In *Offenders' Memories of Violent Crimes* (pp. 279–304). John Wiley & Sons.

Waid, W. M., Orne, E. C., Cook, M. R., & Orne, M. T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of information. *Journal of Applied Psychology*, *63*(6), 728–733.

Waid, W. M., & Orne, M. T. (1980). Individual differences in electrodermal lability and the detection of information and deception. *Journal of Applied Psychology*, *65*(1), 1–8.

Webb, A. K., Hacker, D. J., Osher, D., Cook, A. E., Woltz, D. J., Kristjansson, S., & Kircher, J. C. (2009). Eye movements and pupil size reveal deception in computer administered questionnaires. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (pp. 553–562). Springer Berlin Heidelberg.

Yokoi, Y., Okazaki, Y., Kiriu, M., Kuramochi, T., & Ohama, T. (2001). The validity of the guilty knowledge test used in field cases. *Japanese Journal of Criminal Psychology*, *39*(1), 15–27.

Zhen, Z., Tsiamyrtzis, P., & Pavlidis, I. (2007). Forehead thermal signature extraction in lie detection. Presented at the Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 1–59). New York, NY: Academic Press.