01 Jan 2012

Automated Human Screening for Detecting Concealed Knowledge

Nathan W. Twyman
Missouri University of Science and Technology, nathantwyman@mst.edu

Automated Human Screening for Detecting Concealed Knowledge

AUTOMATED HUMAN SCREENING FOR DETECTING CONCEALED KNOWLEDGE

by

Nathan W. Twyman

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2012

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by NATHAN W. TWYMAN entitled AUTOMATED HUMAN SCREENING FOR CONCEALED KNOWLEDGE and recommend that it be accepted as fulfilling the dissertation requirement for the DEGREE OF DOCTOR OF PHILOSOPHY

_______________________________________Date: 4/16/2012
Jay F. Nunamaker, Jr.

_______________________________________Date: 4/16/2012
Judee K. Burgoon

_______________________________________Date: 4/16/2012
Susan A. Brown

_______________________________________Date: 4/16/2012
Mark W. Patton

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: 4/16/2012

Dissertation Director: Jay F. Nunamaker, Jr.

_______________________________________Date: 4/16/2012

Dissertation Director: Judee K. Burgoon

STATEMENT BY AUTHOR

SIGNED: ____________________________________
Nathan W. Twyman

ACKNOWLEDGEMENTS

The Department of Homeland Security's (DHS) National Center for Border Security and Immigration (BORDERS) and the National Center for Credibility Assessment (NCCA) provided significant support for this research. 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 appreciate my committee, Jay F. Nunamaker Jr., Judee K. Burgoon, Susan A. Brown, and Mark W. Patton for their training and insight. I also thank Aaron C. Elkins, Christopher B. R. Diller, Paul B. Lowry, Mary A. Peterson, Douglas C. Derrick, Sean L. Humpherys, Kevin C. Moffitt, and Joseph S. Valacich for their valuable input, feedback, and support.

TABLE OF CONTENTS

TABLE OF CONTENTS -- Continued

TABLE OF CONTENTS -- Continued

TABLE OF CONTENTS -- Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES -- Continued

ABSTRACT

Screening individuals for concealed knowledge has traditionally been the purview of professional interrogators investigating a crime. But the ability to detect when a person is hiding important information would be of high value to many other fields and functions. This dissertation proposes design principles for and reports on an implementation and empirical evaluation of a non-invasive, automated system for human screening. The screening system design (termed an automated screening kiosk or ASK) is patterned after a standard interviewing method called the Concealed Information Test (CIT), which is built on theories explaining psychophysiological and behavioral effects of human orienting and defensive responses. As part of testing the ASK proof of concept, I propose and empirically examine alternative indicators of concealed knowledge in a CIT. Specifically, I propose kinesic rigidity as a viable cue, propose and instantiate an automated method for capturing rigidity, and test its viability using a traditional CIT experiment. I also examine oculomotor behavior using a mock security screening experiment using an ASK system design. Participants in this second experiment packed a fake improvised explosive device (IED) in a bag and were screened by an ASK system. Results indicate that the ASK design, if implemented within a highly controlled framework such as the CIT, has potential to overcome barriers to more widespread application of concealed knowledge testing in government and business settings.

1. INTRODUCTION

The most difficult type of information to obtain is often that which is intentionally kept hidden. Yet hidden information is often the most valuable. The perceived ability to successfully conceal information motivates individuals to hide poor performance, commit fraud, and even engage in acts of terrorism. For decades Bernie Madoff successfully concealed the fact that his financial service was secretly a Ponzi scheme—resulting in a price tag of over $64 billion (Graybow, 2009). No one on board the aircraft knew Umar Farouk Abdulmutallab had smuggled explosives aboard, which act nearly cost the lives of 289 individuals (United States of America vs. Umar Farouk Abdulmutallab, 2010).

The discovery of high-value, purposely concealed information is an important topic in many fields. Financial fraud detection usually involves searching for deliberately concealed data. Criminal forensics and criminal investigations often include searching for evidence that was deliberately hidden. Cyber security involves seeking purposely concealed malware and other concealed intrusions (Morales, 2008).

Hidden information is of interest to more than just criminal detection organizations. Retail establishments seek ways to unveil each customer’s willingness to pay. Employee recruitment and evaluation teams desire to discover hidden malicious intentions or policy non-compliance. Large event planning and management personnel need methods of screening people for potential security threats.

1.1 Security Screening Methods

Though the theory, protocol, and system design in this paper may be applied to many of these contexts, I chose to focus much of this study on human security screening. In this paper I refer to *security screening* as the evaluation and containment of potential human threats, prior to physical entry of a building, territory, or area of interest. I refer to *security screening methods* as tactics employed to generate an evaluation. Metal detector scans, official documentation evaluations, and simple questioning techniques are common examples of security screening methods.

Screening methods have become increasingly costly, time-consuming, and intrusive; yet, performance levels remain considerably lower than desired (Bandyk, 2010). The United States Department of Homeland Security (DHS) processed over 267 million incoming border crossings in 2008 (Bureau of Transportation Statistics, 2008), but estimated a 71.1% failure rate when it came to apprehending “major violations” of laws, rules, and regulations (Department of Homeland Security, 2009). Air passenger violations were reportedly apprehended only 25% of time (Department of Homeland Security, 2009). Anecdotal evidence indicates that the problem is not simply overlooking minor infractions: DHS undercover personnel attempting to smuggle explosive devices through security screening at a Newark airport succeeded in 90% of trials, and failure rates were as high as 70% during similar penetration tests at Los Angeles and Chicago airports (Mosk, Hill, & Fleming, 2010). Given the enormity of the consequences of poor security screening, a promising research opportunity exists to improve these methods.

In the extant literature, several studies have addressed improving security screening methods. Studies have investigated improving operational design to more efficiently allocate resources (Cavusoglu, Koh, & Raghunathan, 2010; Lee & Jacobson, 2011; Menneer, Cave, & Donnelly, 2009; Wang & Zhuang, 2011), enhancing sensors for detecting illicit items (Fainberg, 1992; S.-W. Park, Yuk, Ryu, Kim, & Yi, 2006; Vassiliades, Evans, Kaufman, Chan, & Downes, 2008), and improving decision support and screener decision making tools (Jensen, Lowry, Burgoon, & Nunamaker, 2010; Jensen, Lowry, Burgoon, & Nunamaker Jr., 2011; G. Park & DeShon, 2010; Twyman, Jenkins, Carl, & Nunamaker Jr., 2011).

Though these streams of security screening research have greatly contributed to practice, they stop short of addressing *as-yet unknown* threats. Namely, current screening systems tend to be reactionary in nature: they are designed and redesigned to detect only that which has already been discovered through prior experience (Nakanishi, 2008). Accordingly, they do not include mechanisms for preventing threats that have yet to be discovered. In this area, human security screening lags behind related research areas such as fraud detection (Cecchini, Aytug, Koehler, & Pathak, 2010; Holton, 2009; Humpherys, Moffitt, Burns, Burgoon, & Felix, 2011), behavior-based virus detection (Morales, 2008), intrusion prevention (Green, Raz, & Zviran, 2007; Ryu & Rhee, 2008), and cyberterrorism prevention research (Hansen, Lowry, Meservy, & McDonald, 2007) that have begun developing techniques that could be used to avert illegal or undesirable actions that have not been previously committed.

1.2 Veracity Assessment Tools for Illuminating Unknown Threats

This inability to screen for the unexpected can be addressed by integrating elements of veracity assessment research with security screening protocols. In recent years, some IS research has proposed advanced tools for veracity assessment (e.g., D. C. Derrick, Elkins, Burgoon, Nunamaker, & Zeng, 2010; Fukuda, 2001; Meservy et al., 2005; Twyman, Elkins, & Burgoon, 2011; Twyman, Moffitt, Burgoon, & Marchak, 2010). Though few studies have targeted security screening specifically, research results on these tools suggest the possibility of recognizing threats on a more abstract level. For instance, natural language processing tools have been designed to predict the likelihood of financial fraud, even though the exact nature of the fraud may be undefined (Glancy & Yadav, 2011; Humpherys, et al., 2011). In addition to linguistic analyses, IS veracity assessment research has investigated other tools such as body movement analyses, vocalic feature analyses, pupillometry, thermal measurement, and noncontact cardiorespiratory measures (D. C. Derrick, et al., 2010; Meservy, Jensen, Kruse, Burgoon, Nunamaker, et al., 2005). This research has demonstrated the potential value of certain tools, showing that systems can improve accuracy and can automate processes.

However, these tools cannot be equally applied to every situation. Cues to deception are likely to be heavily influenced by many factors. Psychology, criminal justice, and communication research on deception detection demonstrates that when it comes to screening, the *protocol* used to assess veracity can be just as important as the cue or measurement (Hartwig, Granhag, Stromwall, & Kronkvist, 2006; Levine, Shaw, & Shulman, 2010). The effectiveness of a given tool for measuring veracity is influenced by factors such as interviewer skill (Iacono, 2008) and crime-related knowledge (E. Elaad, 1997), the level of synchrony in the communication (Humpherys, et al., 2011; Zhou, Burgoon, Nunamaker, & Twitchell, 2004), and even the type of questions asked (J. K. Burgoon, Buller, Ebesu, & Rockwell, 1994; Moffitt, 2011). Thus, a natural step forward in this field of research is to design and test screening systems that are rooted in procedures and theories that are well-established and generalizable. To establish validity and reliability in automated human security screening, methods and processes need to be researched in conjunction with effective technologies.

To address these opportunities in security screening research, this study proposes and evaluates an automated security screening system design that is based on a modified version of a successful screening technique called the Concealed Information Test (CIT; Lykken, 1974). Rather than scanning for the threat itself, the system searches for hidden threats at a more abstract level, by asking each individual several questions while recording psychophysiological and behavioral responses, similar to the common lie-detector test (i.e. “polygraph”) (Lykken, 1998;

Raskin & Honts, 2002). This system design is similar in concept to behavior-based computer virus detection, which analyzes the behavior of programs looking for suspicious activity rather than only looking for specific virus signatures (e.g., Kim, Shin, & Pillai, 2011; Morales, 2008; O'Kane, Sezer, & McLaughlin, 2011).

A well-designed screening system for detecting intentionally concealed knowledge could be useful in any situation where the discovery of such information would be valuable. For instance, criminal investigations can benefit from knowing which suspects possess crucial information. Job applications could be accompanied by an objective integrity evaluation. Retraining courses could be personalized for corporate policy topics for which employees prefer not to reveal ignorance or non-compliance. Physical building security could be enhanced. These considerations inspire the research question:

What are design principles for a system and protocol for automated screening of individuals for concealed knowledge?

As the ensuing discussion will show, the investigation into this research question revealed that non-traditional indicators of concealed information are needed. I investigate three potential indicators, two of which are novel to CIT and deception detection research. Thus, in addition to outlining a screening system design, a secondary research question for this work is as follows:

How can measurement of kinesic, oculomotor, and vocalic behaviors serve as automated alternative indicators of concealed knowledge?

The remainder of this dissertation proceeds as follows: existing standard interviewing techniques are reviewed, and justification is provided for using the CIT as a foundation for this study and system design. Following this is a narrative that introduces and explains how oculomotor (eye movement-based) patterns, body movement rigidity, and response time can serve as indicators of concealed information. The experiments used to investigate alternative indicators and the potential of a CIT-based screening system are then reviewed and results reported. Included in the discussion of the final experiment is a description of an implemented automated screening system termed an automated screening kiosk (ASK), which was evaluated in the experiment. The final experiment involved having participants build a mock improvised explosive device (IED) and try to smuggle it through a security screening station that implemented the ASK. Following the reported results of these experiments is a discussion of the implications for research and practice together with future research directions.

1.3 Background on Standardized Interviewing Techniques

As noted, studies investigating veracity assessment and screening tools illuminate the value of a standardized screening process (Hartwig, et al., 2006; Humpherys, et al., 2011; Levine, et al., 2010; Moffitt, 2011). Standardized interviewing techniques such as the CIT have historically been used to assess the veracity of persons under investigation for their involvement in a crime or other illicit activity by asking particular questions and measuring physiological and behavioral responses (Gamer, Verschuere, Crombez, & Vossel, 2008; MacLaren, 2001). Several standardized interpersonal screening techniques are currently used regularly by practitioners, the most common of which include the Control Question Test (CQT), the Behavioral Analysis Interview (BAI), and the CIT (Vrij, 2008). I briefly overview these here; Vrij's (2008) compilation contains a more in-depth review.

The CQT is currently the most commonly used interviewing technique for veracity assessment (Vrij, 2008). The CQT takes several hours to complete, and requires a high level of skill on the part of the interviewer to obtain valid results. The interviewer in the main phase of the CQT asks several control questions that every person is likely to lie about (e.g., "Have you ever taken something that does not belong to you?"), and several questions that are directly relevant to the crime or illicit act (e.g., "Did you steal the car?"). Before doing this, the interviewer must lead the examinee to believe that admitting to such a question would necessarily show that he or she is the type of person who would commit the crime or illicit act that is the subject of the CQT. This manipulation is necessary to ensure that an innocent examinee will experience more arousal when presented with these control questions than about crime-relevant questions, whereas a guilty examinee will experience equal or more arousal during the relevant question (Meijer & Verschuere, 2010; Raskin & Honts, 2002). Though commonly used in practice, the CQT is criticized as having a weak theoretical foundation by academic researchers

(Ben-Shakhar, 2002; Iacono, 2000; Lykken, 1998; Meijer & Verschuere, 2010; Vrij, 2008)—most notably the United States National Research Council (2003). Among the major research concerns cited are the inability to measure objectively whether the interviewer has successfully manipulated the interviewee (Fiedler, Schmid, & Stahl, 2002; Iacono, 2008; Lykken, 1998), and uncertainty as to whether psychophysiological indicators can reliably distinguish between a guilty person's fear of being caught and an innocent person's fear of false detection (Meijer & Verschuere, 2010).

The BAI is a method of interviewing sponsored primarily by a single practitioner group (John E. Reid & Associates, 2011). It is probably the second most common interviewing technique used in the United States (Vrij, 2008). The BAI interviewer asks a series of 15-16 standard questions designed to elicit certain verbal and non-verbal responses. Developers of the BAI posit that truth-tellers and deceivers should react differently to these questions as a result of differing attitudes toward the crime or event in question (Horvath, Jayne, & Buckley, 1994; Inbau, Reid, Buckley, & Jayne, 2001). For instance, during what is termed the "motive" question (e.g., "Why do you think someone stole the car?"), guilty examinees are expected to show posture shifts or foot bouncing as a means of reducing a high level of anxiety which purportedly should not be present in innocent examinees. Aside from an initial positive evaluation (Horvath, et al., 1994), little direct scientific investigation validates the BAI. A few direct investigations challenge the validity of the BAI (Blair & Kooi, 2004; Vrij, Mann, & Fisher, 2006), though the ecological validity of these studies has been called into question (Horvath, Blair, & Buckley, 2008). The most recent BAI investigation suggests that portions of the BAI do show promise (Horvath, et al., 2008). The mechanisms underlying the BAI remain underexplored, and much more research is needed before validity of the BAI can be established (Blair & Kooi, 2004; Horvath, et al., 2008).

As noted, the CIT is an interviewing technique used to determine whether an examinee is concealing knowledge (Ben-Shakhar & Elaad, 2003; Lykken, 1959). The CIT actually predates the CQT: the concept was first described in 1908 by Münsterberg (Münsterberg, 1908). But research on the concept did not move forward until much later when Lykken labeled it and proposed it as a more viable alternative to contemporary techniques. In a standard CIT, an interviewer recites several prepared questions or statements regarding the activity (e.g., crime) in question. Prepared with each question are several plausible answers, collectively called a *foil*, which are also recited by the interviewer. For instance, if the activity in question is the theft of a vehicle, one of the CIT questions might read: "If you were involved in the theft of the vehicle, you would know the color of the car that was stolen. Repeat after me these car colors." The interviewer would then verbally recite each item in the associated foil, which would consist of about four to six colors, only one of which would be the correct color. The examinee is usually asked to either repeat the items or reply with a verbal "yes" or "no" after each item is spoken by the interviewer. Once the examinee has spoken, the interviewee and interviewer sit in silence for several seconds while psychophysiological measurements are recorded.

Though not as commonly used by practitioners as the CQT or BAI, researchers widely consider the CIT to be the most scientifically valid approach (Ben-Shakhar & Elaad, 2003; Fiedler, et al., 2002; Iacono & Lykken, 1997; National Research Council, 2003). Unlike the CQT, BAI, and similar techniques, the CIT does not rely heavily on the capabilities of the interviewer. Instead, the CIT interviewer plays only a minor role—requiring little to no skill. Moreover, an innocent person's fear of detection should not affect the outcome in a CIT, because responses to all items should be consistent whether their general arousal level is low or high. For instance, if the relevant (correct) response to the example CIT question above was "blue," an innocent person's fear of being falsely classified should create no different effect than when the response is "green," "white," or any other option, simply because they have no knowledge of which option is correct.

Table 1 summarizes the standard interviewing methods for detecting concealed information.

Table 1. Standard Interviewing Methods used for Detecting Concealed Information

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.

2. HOW THE CONCEALED INFORMATION TEST CAN BE ADAPTED FOR SCREENING SYSTEMS

Whereas aspects of each interviewing technique have potential application to automated screening systems, the CIT has several unique advantages that made it the clear choice for the protocol portion of an automated screening system design. First, it requires the least time and little interviewer skill or intervention, which not only helps to control for interviewer effects but also makes automation easier. The CIT also generates the strongest within-subjects baseline for comparison.

Additional advantages of the CIT stem from its foil structure and length. Each foil is self-contained; a system or evaluator can use a single foil to make a judgment. However, the use of multiple foils reduces the probability of false detects, as long as each foil is associated with a question that is central to the hidden knowledge in question (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003). The low ratio of relevant to non-relevant options within each foil (usually between 1:3 and 1:6) creates a strong baseline that is both person- and question-specific. The CIT process requires much less time compared to other techniques such as the BAI. Evaluators can complete several foils in a matter of minutes, whereas alternative techniques can last hours.

The veracity assessment literature offers three main criticisms of the CIT. These limitations have been reported to be the main reasons the CIT has not enjoyed more widespread adoption in law enforcement practices, as reviews of cases revealed that only a small percentage of criminal cases are reported to meet the necessary criteria (Podlesny, 1993, 2003). To the extent these concerns can be addressed, the CIT could be applied more generally. However, it should be noted that weaknesses exist in the method Podlesny used in the above-cited studies (Meijer & Verschuere, 2010), and in spite of these challenges, Japan has widely adopted this technique for investigations (Hira & Furumitsu, 2002; Nakayama, 2002). Still, these criticisms remain the main arguments against wider adoption.

First, the preparation phase for the CIT is more time-consuming and complicated compared to alternative interviewing techniques. Critical foil items have to be designed from a pool of information that is known and knowable only by the perpetrator of the illicit act. This process can be difficult and sometimes it is impossible to identify enough usable items for testing. When critical foil items are identified, a *lack* of familiarity to innocent parties should be established through pretesting, and non-critical foil items must be pretested as well, all of which can be too costly for a single investigation. However, the nature of human screening is often such that hundreds or sometimes thousands of examinees could undergo the same test in a given location and context, minimizing the relative cost of preparation.

The second criticism is the requirement that guilty knowledge be possessed by only the guilty party, or innocent persons will be improperly accused (Bradley, MacLaren, & Carle, 1996). In security screening, it is plausible that an individual may know of a specific crime in progress, even though he or she is not directly involved. To the extent that this occurs, a rapid screening CIT will face this same difficulty. The National Research Council suggests that this may be addressed by treating the response variables more on a scale rather than as a dichotomy (National Research Council, 2003). Their suggestion rests on the assumption that guilty knowledge will be more poignant with the responsible party than with a witness—an assumption that needs further investigation (Gamer, 2010).

The third criticism of the CIT is similar to the second. For an orienting response to occur, a reasonably strong certainty must exist that the critical items chosen for CIT foils have a high degree of personal significance in the mind of the guilty suspect; otherwise, the difference in response between guilty and innocent suspects will not be as diagnostic (Carmel, et al., 2003; Gati & Ben-Shakhar, 1990). Ensuring this link is especially difficult when little direct information about the crime or event is known, or when a large amount of time has lapsed between the event and the investigation (Honts, 2004). In a security screening scenario, this concern is minimized. The particulars of the event(s) are well understood, because they are pre-specified. For instance, examinees are usually aware of a list of banned items for which they are being screened. Time is unlikely to be a concern, because in screening scenarios the malicious event is in progress (e.g., smuggling) or is intended (e.g., theft) rather than a distant memory; thus, the event is sure to have personal significance to a guilty examinee.

Aside from these criticisms, a practical challenge exists in using the CIT or any of the standard interviewing techniques as the basis for a screening system. The standard measurements typically used in these techniques require sensors that must be strapped on or otherwise physically connected to the examinee. These invasive sensors require human intervention and monitoring, which requires additional time and undermines some of the benefits of automation. To the extent a CIT system design could overcome these challenges, its usefulness in discovering hidden information can be extended not only to security screening, but also to applications ranging from internal auditing to anti-terrorism.

2.1 Design Principles for Automated Screening for Concealed Guilty Knowledge

Several general design principles for an automated screening system and process for discovering purposely concealed knowledge can be derived from the previous review. The design principles I propose for such systems follow:

1. Identify appropriate stimuli that represent the concealed knowledge in question. Ensure there is reasonably strong certainty that the representation has relatively high personal significance for a person who is concealing such knowledge.
2. Identify irrelevant stimuli that arouse the same baseline level of orienting and defensive responses (discussed in the next section).
3. Develop several foils consisting of relevant and irrelevant stimuli in a one-to-many ratio.
4. Automatically present these foils in an environment where potential

distractions are at a minimum, including but not limited to human distractions.

5. During foil presentation, automatically capture human indicators of the orienting and/or defensive responses.
6. Apply categorization algorithms to the collected data to assess concealed knowledge.

While each design principle is an important component of an automated screening system design, this research focuses mostly on advancing principles 4 and 5. However, the system design proposed later in this document will be evaluated against each of the six principles.

3. THEORETICAL FOUNDATION FOR THE AUTOMATED DETECTION OF CONCEALED KNOWLEDGE USING THE CONCEALED INFORMATION TEST (CIT)

This section explains the theoretical foundation of the CIT and how the CIT can be used to detect concealed knowledge about adverse events, including the presence of those that may not have been identified as yet. The theory underlying the CIT has traditionally centered on the orienting response. In extending CIT research to automated screening for hidden knowledge applications, I also incorporate defensive response theory and related theories used in veracity assessment literature.

3.1 The Orienting Reflex

The CIT draws on the *orienting reflex*, which is the autonomic movement of attention toward novel or personally significant stimuli (Pavlov, 1927; Sokolov, 1963). The level of stimulus *novelty* is a function of the degree to which it matches (or does not match) stimuli that precede it in a given context (Gati & Ben-Shakhar, 1990). The level of *personal significance* is a function of the degree to which a stimulus matches one's cognitive representation of a given item of relevant information (Gati & Ben-Shakhar, 1990). When an individual's autonomic system registers a novel or personally significant stimulus, the sympathetic portion of the nervous system activates to mobilize the body to a state of readiness (i.e., arousal) so that the individual is ready to adapt or react to the stimulus (National Research Council, 2003; Pavlov, 1927; Sokolov, 1963; Verschuere, Crombez, Clercq, & Koster, 2004).

The orienting reflex is an autonomic response that creates measurable physiological effects (Ambach, Bursch, Stark, & Vaitl, 2010; Gamer, et al., 2008; Lykken, 1974). This readiness to adapt includes physiological changes such as variance in heart rate, skin sweatiness, pupil dilation, and respiration (Ambach, et al., 2010; Eitan Elaad & Ben-Shakhar, 2009; Lykken, 1959). Stimuli that have stronger relevance or “signal value” (Lykken, 1974, p. 728) such as an out-of-place object or hearing one’s own name (Cherry, 1953) produce a stronger orienting reflex (Bernstein, 1979; Maltzman, 1979). With repeated presentations of stimuli, the magnitude of the reflex decreases as a function of the corresponding decrease in novelty and personal significance (Sokolov, 1963). Figure 1 depicts the process of orienting reflex activation.

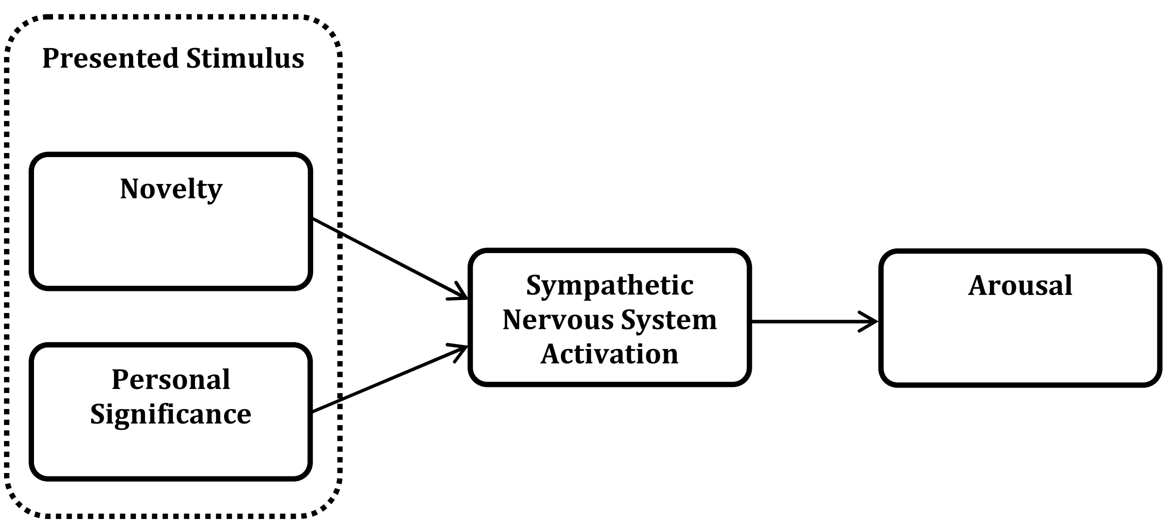

Figure 1. Depiction of the Orienting Reflex

As noted, the CIT uses several multiple choice questions, with only one relevant (i.e., correct) answer per question. Knowledge regarding the correct alternative serves as additional "signal value," which activates the orienting reflex much more strongly than that seen in irrelevant (no "signal value") alternatives (Lykken, 1974). In the CIT, two basic outcomes are compared. The physiological responses that follow presentation of a relevant stimulus are compared to the physiological responses that follow presentation of the several irrelevant stimuli (Gamer, et al., 2008).

The key to successful execution of the CIT is thus to identify stimuli that are relevant only to the guilty party. Traditionally, the knowledge in question is related directly to a crime or similar event, which knowledge usually activates guilt, fear, or similar arousal when accessed. These traditional applications helped inspire the original technique label—the Guilty Knowledge Test (GKT) (Ben-Shakhar, Bar-Hillel, & Kremnitzer, 2002; Lykken, 1959, 1974; Podlesny, 1993). However, the CIT has also been used successfully to uncover hidden knowledge unassociated with highly charged emotions, such as hiding knowledge of a playing card (Fukuda, 2001; Gamer, Bauermann, Stoeter, & Vossel, 2007). The phrase "concealed information" rather than "guilty knowledge" has become the phrasing of choice, even though most practical applications of the CIT still focus on knowledge highly associated with guilt.

3.2 When Stimuli are Threatening

Though it has received somewhat less attention in CIT research, there is another mechanism examinees usually exhibit called the *defensive response*. Whereas the orienting reflex can occur with any stimulus of sufficient novelty or personal significance, the defensive response is a reaction only to stimuli perceived to be aversive or threatening. This reaction includes physiological and behavioral changes.

The defensive response was initially coined the "fight-or-flight" behavior in the early 20th century (Walter B. Cannon, 1929). The defensive response can be broken up into at least two phases—an initial defensive *reflex* followed by defensive *behaviors*. When threatening stimuli are first perceived, an initial sympathetic nervous system activation occurs—driving a defensive physiological reaction thought to help the individual assess the threat and determine the appropriate action to take (Sokolov, 1963; Verschuere, et al., 2004). Many of the physiological changes associated with this initial *defensive reflex* are similar to the orienting reflex (e.g. a sudden increase in skin sweatiness) (Verschuere, et al., 2004), though there are differences in cardiovascular response. In CIT research, this reflex stage of the defensive response is thought to amplify many of the physiological measures of the orienting reflex.

The initial defensive reflex transitions into behaviors designed to escape or combat the threat (W. B. Cannon, 1914; Gray, 1988). For the purpose of this paper I term these *defensive behaviors* to distinguish them from the defensive reflex. Behaviors that stem from responding to a threat are not necessarily autonomic, and may be driven by subconscious or conscious mechanisms. Defensive behaviors are driven by a perceived threat, and therefore can be different than behavioral reactions to stimuli perceived to be non-threatening (Ambach, Stark, Peper, & Vaitl, 2008). Though absent from CIT research, these defensive behaviors have been the focus of much research in veracity assessment literature. Veracity assessment literature has documented various "fight or flight" tactics individuals consciously or subconsciously employ when an important deception is under threat of discovery. For instance, a tendency to freeze or become more rigid has been documented (Vrij, Semin, & Bull, 1996). Other behaviors include avoiding direct answers, attempting to distract the evaluator, and/or controlling message content and nonverbal behavior so as to appear truthful (DePaulo, Kirkendol, Tang, & O'Brien, 1988). Deception detection literature has shown that sometimes these defensive behaviors themselves can indicate deception. Sometimes these defensive behaviors are insufficient or unsuccessful, and cues indicating a deception "leak" out thereby exposing the individual. Figure 2 depicts the defensive response to a perceived threat.

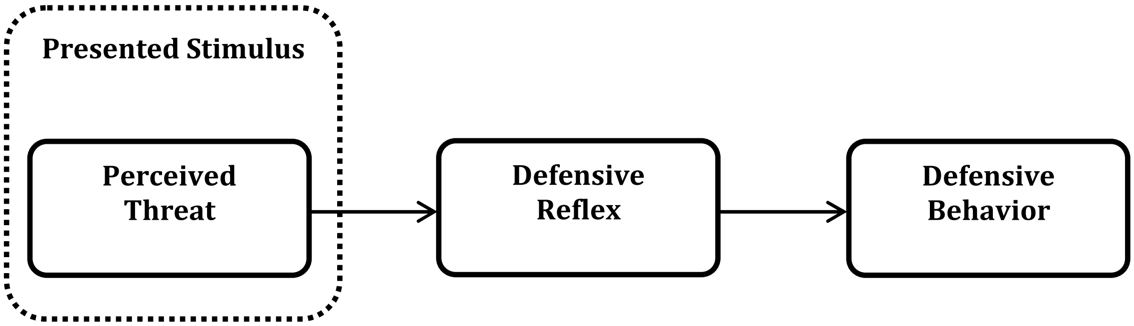

Figure 2. Depiction of defensive responding.

I propose that defensive behaviors can be valuable input for discovering concealed knowledge. In the CIT, aversive or threatening stimuli are those foil items that have potential to expose concealed knowledge about the incident that the individual wishes to keep hidden. When presented with the aversive stimuli, individuals should exhibit defensive behaviors designed to escape or combat the threat. The same stimuli should have no such effect on individuals who do not find the stimuli aversive. Thus, behavior modifications in a CIT can reveal hidden guilty knowledge.

3.3 Measuring Orienting and Defensive Behavior in Screening Applications

Measures of the orienting and defensive reflexes traditionally target skin conductance response (SCR), respiration, and heart rate (Gamer, et al., 2008). Sensors for measuring these physiological reactions require direct contact and manual calibration and supervision—making application to security screening infeasible. These considerations lead us to evaluate alternative measures for detecting concealed knowledge. Though much recent CIT research has begun using functional magnetic resonance imaging (fMRI) or similar brain imaging techniques (e.g., Gamer, et al., 2007; Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011; Hahm et al., 2009; Langleben et al., 2002), the procedures and measurement apparatus for these scenarios are even more invasive than traditional techniques, and likely would require even more specialized supervision. Recent research in deception detection has indicated that eye movement can betray deception (Douglas C. Derrick, Moffitt, & Nunamaker, 2011; Osher, 2007; Steptoe, Steed, Rovira, & Rae, 2010; Twyman, et al., 2010), and advances in eye tracking technology are such that eye movement can be measured non-invasively at a distance. Body movement rigidity and response time have likewise shown potential as cues to deception. For this study I chose eye-movement tracking, movement tracking, and response time as non-contact, automated alternatives for measuring the orienting reflex and the defensive response to aversive stimuli.

3.3.1 Oculomotor Orienting Behavior

When a presented novel or significant stimulus demands visual processing, the eyes reflexively orient toward the stimulus. The rapid movement of the eye from one point of visual focus to another is termed a *saccade*. Saccades are the most common type of eye movement, and can be reflexive (such as when driven by the orienting reflex), or they can be overt (such as when performing a visual search task) (Hollingworth, 2007; Müller & Rabbitt, 1989).

Eye-movement patterns have long been used in cognitive psychology research to explore both the orienting reflex and overt attention shifts (Hollingworth, 2007; Müller & Rabbitt, 1989; Posner, 1980), and some IS research has similarly begun to use eye-movement behavior as a surrogate for visual attention (Cyr, Head, Larios, & Pan, 2009; Djamasbi, Siegel, Skorinko, & Tullis, 2011; Lorigo et al., 2008). The popular spotlight theory of attention (Posner, 1980) posits that stimuli outside the focus of attention are processed by peripheral attention. Visual stimuli are first discovered by peripheral attention; if a stimulus has a sufficient level of significance or novelty the eyes will move toward it. Importantly, saccades can be either reflexive or overt (i.e., consciously controlled) (Duchowski, 2007). To the extent saccades are reflexive they will occur before the stimulus is consciously identified (Posner, 1980).

I propose a system design that exploits this reflexive visual orienting. To do this I propose using visual rather than auditory CIT foils, and presenting foil items simultaneously rather than in a sequence. If visual foils are displayed simultaneously on a screen, those who are hiding knowledge about a particular event should be more likely to orient their initial attention reflexively, and therefore their eyes, toward the visual CIT item that is associated with their guilty knowledge. For instance, if a visual CIT foil consists of the words “bombs,” “knives,” “guns,” and “ammunition,” a person hiding an explosive device should reflexively saccade toward the word “bombs,” as it currently should have the highest level of personal significance relative to the alternative items. In contrast, a person without guilty knowledge would be significantly less likely to orient toward the word, “bombs.” I thus hypothesize that

H1. Guilty knowledge increases the likelihood that an initial saccade will be directed toward the critical item in a collection of simultaneously presented CIT foil items.

As noted, orienting theory posits that over time, the orienting reflex diminishes in a manner corresponding to the associated decrease in novelty and/or personal significance. As an individual gains experience with the format of a rapid screening CIT, he or she could find the novelty of the stimuli diminish. Accordingly, I hypothesize that

H2. Over time, stimuli representing guilty knowledge will be less likely to attract the initial saccade.

3.3.2 Response Time

An alternative measure that has been investigated in CIT literature is response time. The orienting reflex may increase reaction time when an individual is presented with relevant stimuli (Gamer, et al., 2007), because the reorienting of cognitive resources (a result of the orienting reflex) causes a delay. Gronau and colleagues found that response time was not significantly discriminatory measure when answering Stroop-like questions regarding a mock crime (Gronau, Ben-Shakhar, & Cohen, 2005). However, other studies successfully used reaction time in modified CITs (Seymour & Fraynt, 2009; Seymour, Seifert, Shafto, & Mosmann, 2000), and one suggests it may produce discriminating power similar to the polygraph (Verschuere, et al., 2004). Meanwhile, the counterargument for using response time to discriminate concealed knowledge is that it may be easily consciously controlled (Gronau, et al., 2005).

The benefit of response time in this context is that it can be measured non-invasively and automatically. In a rapid screening context, response time may be a useful measure. I replicate the hypothesis of prior CIT work here:

H3: Guilty Knowledge will increase response latency when a critical item is presented.

3.3.3 Oculomotor Defensive Behavior

I propose that upon initial detection of the critical foil item, persons with guilty knowledge will exhibit defensive behavior. Upon detection of the threatening item, attention should orient to the perception of a threat, triggering the defensive response. While there are several possible defensive actions that could be taken, defensive response theory holds that the default defensive response to a threat is usually avoidance or escape (Gray, 1988). In the visual CIT design (detailed in the Methods section), a "safety" point is presented in the center of the screen when each question is recited audibly. Its position is equidistant from all visual foil items, thereby serving as the optimal point of avoidance or point of greatest safety—away from potential threats.

I thus propose that examinees with hidden guilty knowledge will focus more visual attention on the best point of escape—the center point of the screen. Those without guilty knowledge will manifest significantly less propensity to orient to this center point, because they will not share this inherent need for "safety." I thus hypothesize that

H4. Guilty knowledge increases time spent gazing at the safety point.

Unlike the orienting reflex that diminishes with time, defensive behavior should remain constant as long as the examinee has reason to perceive a threat. Namely, a credible threat yesterday does not diminish another credible threat today. This is an important consideration for contexts such as security screening where testing may occur several times for frequent travelers. I thus hypothesize that

H5. Guilty knowledge increases critical stimuli avoidance during repeat exposures.

3.3.4 Kinesic Defensive Behavior

Standard interviewing techniques such as the CIT use multiple sensors to help decrease the potential for error and the possibility of capitalizing on chance. It is likewise prudent in an automated human screening scenario to investigate multiple indicators of concealed knowledge. Though absent from CIT research, much research in deception detection has investigated kinesic (body movement) correlates of veracity. Kinesic rigidity has been documented under conditions of low veracity. Specifically, during high-stakes deception a liar tends to exhibit fewer random movements, such as fewer instances of rubbing hands together or bouncing a leg. Expressive or illustrative movements that do occur tend to be more confined and seem forced, as if they are being resisted (Buller & Aune, 1987; Vrij, 1995; Vrij, et al., 1996).

There are at least two theories that have been used to explain the rigidity phenomenon: cognitive load theory and behavioral control theory. The next subsections review these theories and explore how they relate to the defensive response. Though the rigidity phenomenon has been explored in other research, it has never been explored within the context of the CIT, or an automated screening approach. Thus, this section explores the potential viability of rigidity as a cue to concealed knowledge within a CIT framework, and proposes a method for the automatic detection of rigidity in a CIT context.

3.3.4.1 Cognitive Load Theory

Cognitive load theory proposes that lying takes more cognitive effort than telling the truth, and assumes that fabricating events requires more cognitive resources than simply recalling events (Vrij et al., 2008). Because more cognitive resources are allocated to creating a plausible deception, other activities, including movement, are given less attention. Because of the decrease in cognitive resources allocated to body movement, fewer illustrative or communicative gestures are expected as a result.

3.3.4.2 Behavioral Control Theory

People have motivation to hide personal knowledge related to information that if revealed, would lead to adverse consequences. When attention orients toward stimuli that represent information associated with guilty knowledge, the individual initiates purposeful behaviors designed to avoid potential negative outcomes. Consciously controlling actions in order to appear truthful is a phenomenon that has been termed *behavioral control* in the deception detection literature. When behavioral control can be detected, it can be an indicator of deception (DePaulo, et al., 1988).

Postural rigidity during low-veracity communications is a type of behavioral control. Veracity assessment research has shown that the general population holds to a false believe that liars show increased nervousness in their body movements. Interestingly, while the average person believes a person shows increased body movement when lying, the opposite tends to be the case. A deceiver may therefore overtly become more rigid in an attempt to mimic their own false perception of what a truthful communication should look like.

However, one study found that rigidity seems to persist even when the liar is aware that such behavior is suspicious (Vrij, et al., 1996), suggesting that rigidity may not be an exclusively overt behavior, and it may be difficult to consciously counter this effect.

3.3.4.3 Cognitive Load and Behavioral Control as Defensive Responses

Both of these theories describe defensive responding. The underlying driver of the defensive response is the perception of a threat. When a threat is detected, the sympathetic nervous system activates, moving the body to a state of hypervigilance. The initial stage of the defensive response has been called the “stop, look, and listen” reflex (Bracha, Ralston, Matsukawa, Williams, & Bracha, 2004; Gray, 1988). This phenomenon creates cognitive arousal above what is normal.

Thus, from the beginning of the threat perception, cognitive load becomes a factor. The increase cognitive arousal continues at least as long as the threat is present, and those resources are used to combat the threat. One method of combating the threat may be overt behavioral control of movement, or a focus on verbal messaging that decreases resources allocated to non-verbal messaging. In either case, increased cognitive arousal stemming from defensive nervous system activation is the initial driver of the modified behavior.

There is another, more autonomic feature of the defensive response. A “freeze” response is one method of combating a threat, though it is usually characterized as a last-resort defensive tactic. When no other option for fighting or escaping is apparent, individuals instinctively “freeze up,” purportedly in an effort to avoid drawing attention from the threat in hopes it will take little notice.

Whether cognitive load, behavioral control, or freeze response, the hypothesized result is a decrease in overall body movement. This study does not directly compare these theories, but it is the author's opinion that each of them may contribute the rigidity phenomenon to varying degrees under different contexts.

3.3.4.4 Rigidity in the CIT

To date, no known study has investigated rigidity in a CIT. At first glance, analyzing body movement in a CIT seems almost a non-sequitur. The examinee gives only a "yes" or "no" answer to each foil item, or repeats a word spoken by the interviewer, then sits in silence for several seconds. No communicative body movement is required during the interaction. Because the CIT requires no message fabrication, very few cognitive resources will be allocated to creating a believable verbal message. Likewise, illustrative movement will not be present, and therefore cannot be actively manipulated by the examinee in an attempt to appear truthful.

This is not to say cognitive activation will be absent. To the contrary, the presentation of a threatening stimulus such as will occur when the correct foil option is presented will activate the defensive response, which increases cognitive activation autonomically.

In spite of the lack of communicative movement, rigidity should still be present during the presentation of relevant CIT foil items. First, rigidity should occur as a result of the freeze response. Secondly, the simple one-word answer should require few cognitive resources, leaving ample resources to allocate toward attempting to avoid detection by controlling body language. Both of these mechanisms may involve an unconscious or semi-conscious reaction, which may help to explain the findings of Vrij and colleagues (1996), who discovered that individuals have difficulty countering this tendency toward rigidity.

H6: Guilty knowledge will increase body movement rigidity while a critical item is presented.

4. A NOVEL METHOD FOR MEASURING RIGIDITY

Traditionally, rigidity has been measured using human coders, who subjectively rate the appearance of forced versus natural gesturing given the type of gesture and the context in which it was made. Human coding is limited to the gross movement that can be perceived by a given coder and it remains subject to inter-coder error. Minute changes in movement can be imperceptible to human coders. The largest limitation of subjective rigidity coding is the large amount of time and labor cost required.

As an alternative to subjective human rigidity coding, I developed a novel automated method for measuring rigidity that is well suited for a CIT-based human screening system. Automated rigidity measurement via computer vision can introduce potential for more objective, real-time results at a much lower cost than human coding, allowing for more widespread application. Such an advance is critical if rigidity is to be included in an automated human screening system. The method involves using computer vision techniques (e.g., Kanaujia, Huang, & Metaxas, 2006; Meservy, Jensen, Kruse, Burgoon, & Nunamaker, 2005) to identify face and hands in video and use frame-by-frame position changes of these features to calculate overall body movement during CIT segments.

4.1 Skin Blob Tracking

To measure hand/arm movement automatically, I adopted a video analysis approach. For this study, I employ a skin blob tracking technique initially introduced to deception detection research by Meservy and colleagues (Meservy, Jensen, Kruse, Burgoon, & Nunamaker, 2005). To measure head movement, I employ an Active Shape Modeling technique (Kanaujia, et al., 2006). These techniques analyze video frame by frame. For each frame, the face is detected using the Viola-Jones algorithm (Viola & Jones, 2004). Once the face is detected, hand/arm "blobs" are identified by searching for areas of similar skin color. The centroid of each hand/arm blob is among the features identified in each frame.

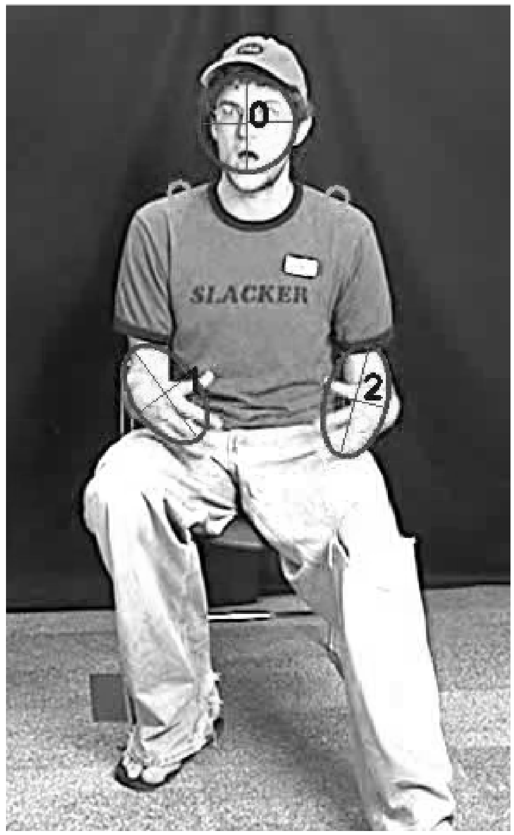

Figure 3. Depiction of skin blob tracking for face and hands. The face is detected and face skin color is used as a reference to detect hands.

4.2 Active Shape Modeling for Tracking Head Movement

Compared to hand/arm movement, minor changes in head movement may be more difficult to detect in standard-definition video. An alternative procedure for measuring head movement is to analyze a close-up video recording of the face. For this study I used a software suite patented by Rutgers University called ASM Face Tracker (Kanaujia, et al., 2006). The software suite tracks the 2-dimensional position of many points on a face in a close-up video feed in standard definition. The computer vision technique is built on an active shape model (ASM), which uses statistical models of face shapes to match identified points on an object in one image to points on an object in a new image. In essence, the ASM algorithm tries to match the statistical model parameters to the image. Thus the model can deform (e.g., stretch), but not beyond what would be naturally seen in a real-world object, given properly defined model parameters (Cootes, Taylor, Cooper, & Graham, 1995). For faces, this means that identified facial points must as a whole represent the image of a face. For instance, a point on the chin cannot be accidentally identified as immediately adjacent to a point on the eye (Judee K. Burgoon et al., 2010).

4.3 Automating Rigidity Measurement

To measure movement in an interview, the centroids of the head and the left and right hand blobs can be determined for each frame between the end of the interviewer utterance and the beginning of the next interviewer utterance. Each time the centroid of the skin blob changes positions, the total Euclidean distance between pre- and post-change is calculated. The sum of these distances results in a total distance moved for a given segment. This calculation is reflected in the following model:

$$M_s = \sum_j^i \sqrt{((y_2-y_1)^2+(x_2-x_1)^2)}$$

The same model can be used to measure movement via the close-up video analysis of the head. A point or set of points can be used in the same manner as the centroids from the skin blob input. Using the output from the ASM Face Tracker software, a point near the center of the head was chosen for this study.

This measure of total distance moved could then be standardized in a within-subject, within-question manner. This is an important step to account for the fact that some people naturally move more than others, and the idea that variations in discussion topic can have a greater or lesser impact on the orienting reflex and defensive responding. In this study, total movement numbers are standardized using within-subject, within-question z-scores.

4.4 Potential Limitations of Automated Rigidity Measurement

This method of automated rigidity measurement is well suited for the CIT. A less controlled interviewing format introduces the possibility of one response naturally requiring different body language than another. For instance, smaller gestures may be used to communicate the concept of "little" in response to one question, while larger gestures may be used to communicate "big" in response to a separate question. Thus, simply comparing aggregated movement in an open-format interview without considering response context would not lead to optimal measurement of rigidity.

Efforts to automatically identify and classify body movements are ongoing, but are currently inadequate for open-format discussions. The CIT's controlled format eliminates this problem because no communicative body movement is required during the interaction. In fact, the lack of communicative movement allows for automated, more precise measurement of rigidity—by aggregating all movement within each foil segment, the amount of movement during the critical foil item can be compared to the amount of movement during the non-critical items, providing a person-specific and question-specific baseline.

5. METHODS

I used two complementary methods to further this research, namely, laboratory experimentation and system building. First, a traditional CIT was conducted during an experiment that involved a mock crime. Next, I created a prototype of an automated security screening system to test the modified CIT process for rapid security screening. Finally, I conducted a laboratory experiment to test the hypotheses and the efficacy of the design. The second experiment involved having participants construct a mock improvised explosive device (IED) and then attempt to bring it through a screening station.

5.1 Experiment 1

Experiment 1 was part of a larger study led by Burgoon and colleagues for the purpose of investigating noninvasive, multimodal measurement in deception detection (J. Burgoon, Nunamaker, & Metaxas, 2010). The experiment involved the commission of a mock crime. Participants were divided into two conditions, termed "Guilty" and "Innocent." Participants in both conditions were given elaborate instructions on activities to accomplish as part of their participation. Activities for participants in the Guilty condition mirrored those in the Innocent condition, except that those in the Guilty condition also stole a ring out of a desk. After the activities were completed, all participants were interviewed by a professional polygraph examiner. Near the end of this interview, examinees underwent a CIT. Video recordings of these CIT portions of interviews were submitted to computer vision algorithms to generate body movement data for this study.

5.1.1 Experiment 1 Participants

Participants (N=164) were recruited from the local community, via advertisements in local and school newspapers, community shoppers, and craigslist. About three quarters of the sample (76%) were Caucasian, 9% were of Hispanic descent, 7% African-American, with the remainder including Asian, Native American, or other ethnic background. Few (8%) were students, though 40% reported some college and 26% reported a four-year degree. 20% reported only a high school education. Of the 164 enrolled participants, 134 (82%) followed instructions and completed the task. The remaining 18% were disqualified because they either did not follow instructions or confessed during the interview portion of the task. The average age of each participant was 39.5 (standard deviation of 14.0).

5.1.2 Experiment 1 Task

Participants in the experiment were instructed to arrive at a room in an upper floor of an old apartment complex. A pre-recorded set of instructions was waiting for them. After listening to the recorded instructions and signing a consent form, participants left the apartment complex and walked to the nearby school.

Per instructions, participants reported to a room on the top floor and asked for a Mr. Carlson. A confederate acting as a new receptionist who did not know Mr.

Carlson asked the participant to wait while he went to locate Mr. Carlson. A hidden webcam in the room verified the participants' activities while they were waiting for the receptionist's return. Participants in the innocent condition simply waited, while those in the guilty condition had been instructed to steal a diamond ring from the desk. Those in the guilty condition took a key from a mug on the top of a desk and used it to open a blue cash box in the desk drawer that was hidden underneath a tissue box. They removed the ring from the cash box and hid it somewhere on their person.

After his return, the confederate receptionist directed each participant to another room on the bottom floor. Upon arrival, each participant was told that a crime had occurred in the building that day and that they would be interviewed to assess their possible involvement in that crime. Each participant was then interviewed by one of four professional interviewers provided from several intelligence agencies with the assistance of the National Center for Credibility Assessment (NCCA). The interviewers were trained and experienced in criminal interviewing, and were familiar with the purpose and procedure involved in administering a Concealed Information Test.

A CIT test consisting of three CIT foils took place near the end of the interview. Measurements taken during these three questions were included in the analyses. The three questions together with their associated relevant and irrelevant items are included in the table below.

Table 2. CIT Foils used in Experiment 1

Foil Number	Question	Words Repeated by Suspect
1	"If you are the person who stole the ring, you are familiar with details of the cash box it was stored in. Repeat after me these cash box colors:"	1. Green 2. Beige 3. White 4. Blue* 5. Black 6. Red
2	"If you are the person who stole the ring, you moved an object in the desk drawer to locate the cash box containing the ring. Repeat after me these objects:"	1. Notepad 2. Telephone book 3. Woman's sweater 4. Laptop bag 5. Tissue box* 6. Brown purse
3	"If you are the person who stole the ring, you know what type of ring it was. Repeat after me these types of rings:"	1. Emerald ring 2. Turquoise ring 3. Amethyst ring 4. Diamond ring* 5. Ruby ring 6. Gold ring

Note: Relevant items denoted by *.

Many CIT studies have participants review relevant items prior to the interview or testing phase to ensure salience. While this practice may provide some experimental control, it is not a reasonable procedure for real-world applications. Since a goal of this study is to investigate potential for real-world application, no review of relevant items was used. To ensure high validity in field settings, a CIT needs to contain questions such that relevant items have a very high probability of being personally significant to the real criminal (e.g., a murder weapon or an unexpected event that happened during the crime).

At the end of the interview, participants were paid for their time and were given an additional $50 monetary reward if they successfully convinced the interviewer that they were innocent. This large monetary reward together with the realism of the experiment was important to induce motivation to succeed. At the end of the interview, the interviewer made a judgment as to the participant's guilt or innocence. Manipulation checks ensured that the participants conducted their task per their condition. A final questionnaire contained these simple manipulation check questions, together with a question about perceived behavioral control, and measures of arousal and motivation levels.

5.1.3 Experiment 1 Equipment

During the interview, several cameras and other measurement equipment were present in the room, though no equipment was actually attached or touched the participant, as the goal was to assess deception or guilt non-invasively. Two video cameras were placed directly in front of the chair each participant sat in during the interview. One camera captured a full-body frame, while the second camera concentrated on a close-up of the head. The chair had no armrests and a low back. No other furniture or objects were placed within reaching distance. All of this was done to ensure a clear camera view. It also ensured arms and hands could only rest on legs during the CIT portion of the interview. In this way, hand/arm movement also reflects leg movement.

5.2 The Automated Screening Kiosk System

An automated screening kiosk (ASK) system was designed as a means of implementing the design for a rapid screening CIT. I constructed the ASK to test the hypotheses and to discover the technical challenges, limitations, and unexpected findings from building such a system.

The ASK system was designed to conduct a rapid visual CIT automatically while simultaneously gathering oculometric, kinesic (i.e., body movement), and vocalic data. Though the ASK does not currently provide veracity judgments in real-time, future versions of ASK will include real-time judgment capability. The ASK gathers eye movement data using an EyeTech™ TM3 eye tracking device. The ASK gathers kinesic and vocalic data using high-definition video recording and a studio-quality microphone. The kinesic and vocalic data will be analyzed as part of a future study.

The ASK system waits in a readiness state until it recognizes eyes within the field of recognition of the TM3. Once it recognizes eyes, the process begins automatically. A computer-generated voice gives initial instructions to the person being screened. ASK then guides the individual through a 10- to 15-second 9-point calibration process, which allows the device to more accurately track each individual's unique oculomotor activity. Figure 4 depicts the EyeTech TM3 eye-tracking device.

Figure 4. The EyeTech TM3 eye-tracking device. The TM3 is optimally placed directly below a computer monitor.

Following a successful calibration, ASK uses a computer-generated voice to ask CIT-based questions. While a CIT question is asked, the screen remains blank except for a fixation marker in the center of the screen. This fixation marker serves to standardize the starting point for visual attention prior to the presentation of a foil. It also serves as the single point on the screen that is equidistant from all foil items (the "safest" place). Figure 5 depicts this fixation marker.

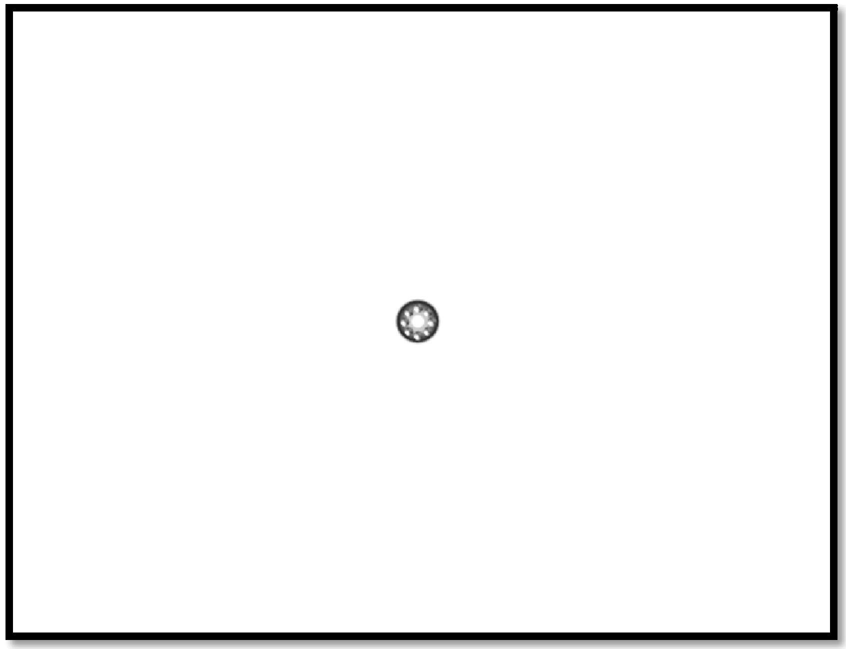

Figure 5. Fixation marker present when each foil question was asked.

Immediately following each question, the fixation marker disappears as ASK simultaneously presents four boxes on the screen, equidistant from one another—one in each quadrant of the screen. These "stimuli screens" are displayed for 7.5 seconds each, allowing time for the participant to examine each of the four stimuli and respond to the question verbally with a "Yes" or "No."

The entire process takes approximately two minutes. After completing the process, the ASK instructs the participant to proceed forward, then returns to a readiness state, awaiting the next participant. These features allow the system to operate automatically, without requirement of human intervention, except in cases where the ASK determines that further screening is desirable. In such circumstances, an ASK could be designed to alert a managing human agent, while simultaneously directing the traveler to a secondary screening station.

5.3 Experiment 2

The experimental design involved two treatments with eight repeated measures presented on two separate days (4 per day). The two treatments were termed Guilty and Innocent. Half of participants were assigned to the Guilty treatment, which involved constructing a mock IED (i.e., bomb) and packing it in a bag. Innocent participants also packed a bag. Both conditions involved bringing the packed bag through a mock building security screening station. The purpose of constructing an IED was to simulate guilty knowledge as closely as possible in a laboratory environment. If there is no arousal associated with the concealed knowledge, defensive responding is unlikely to occur.

Participants interacted with the screening system twice, with the second interaction occurring on the day immediately following the initial interaction. The repeated measures occurred at the screening station, where participants viewed four visual CIT foils, with oculomotor measurements recorded during the presentation of each foil.

5.3.1 Experiment 2 Participants

Participants (N=172) were recruited from an introductory MIS course at a large southwestern university in the United States. This course includes students from across the university. Students participated to fulfill a course requirement to participate in a research experiment. Participants' average age was 23; 61.7% were of Caucasian descent; 13.1% were from Asian or Pacific Islander descent; 17.1% were of Hispanic descent; and 6.3% were of multi-racial descent. Participant gender was evenly split: 50% percent were female and 50% were male.

All human-subjects protocols were followed, and the experimental design was approved by the institution's human subjects review board. Each participant was aware that their participation was strictly voluntary, and each was instructed that he or she could refuse to participate or halt participation at any time, and alternative means of fulfilling the course requirement were available. A review of anonymized, open-ended feedback collected at the end of the experiment revealed that although most participants in the guilty condition found the experience somewhat unnerving, they appreciated the stated goals of the experiment and hoped their participation would help improve security screening technology.

5.3.2 Experiment 2 Task

The task I chose for the experiment focused on security screening for explosives because this is a real-world high-stakes scenario of interest where guilty knowledge is highly salient. Again, the two between-subjects experimental conditions were Innocent and Guilty. After completing a short questionnaire, participants received written instructions telling them to pack a bag and deliver it to an office in a secure portion of the building. To enter the secure portion of the building, the participant needed to pass through a security screening process. Participants packed the bag with items provided (e.g. clothes, pad of paper, book, and other innocuous items). The instructions briefly mentioned that if judged suspicious, a secondary screening process would be necessary, which in the case of innocent participants would take additional time, thereby preventing them from delivering the bag in a timely manner. Guilty participants were reminded that a more in-depth screening would expose their deception and prevent their delivering a bag.

Participants in the Guilty condition had two additional instructions. First, they received instructions on how to construct a mock explosive device (pre-assembled parts were provided) and were told to pack it in their bag along with the other items. Figure 6 is a photograph of an actual device used in the experiment. Second, they were shown a photograph of the face of a person to which they were to deliver the bag. After packing the bag, each participants came to the screening room where they interacted with the ASK.

Figure 6. Improvised explosive device (IED) constructed by half of participants.

For this experiment, the ASK conducted five CIT-based questions. The first four CIT questions read as follows:

> *The following items are not allowed beyond this point. Are you carrying any of these items?*

Of the four boxes containing foil items that were subsequently presented, only one contained a word that was designed to relate to guilty knowledge (i.e., “Bombs,” “Explosives”, “Weapons”), though all boxes contained a word describing a name of a banned class of items. These words were pretested via self-report to ensure they were equally arousing by default. Figure 7 illustrates the screen shown to the participants. The other items were certain to not be carried by participants, since what was packed in their bags was known beforehand[1].

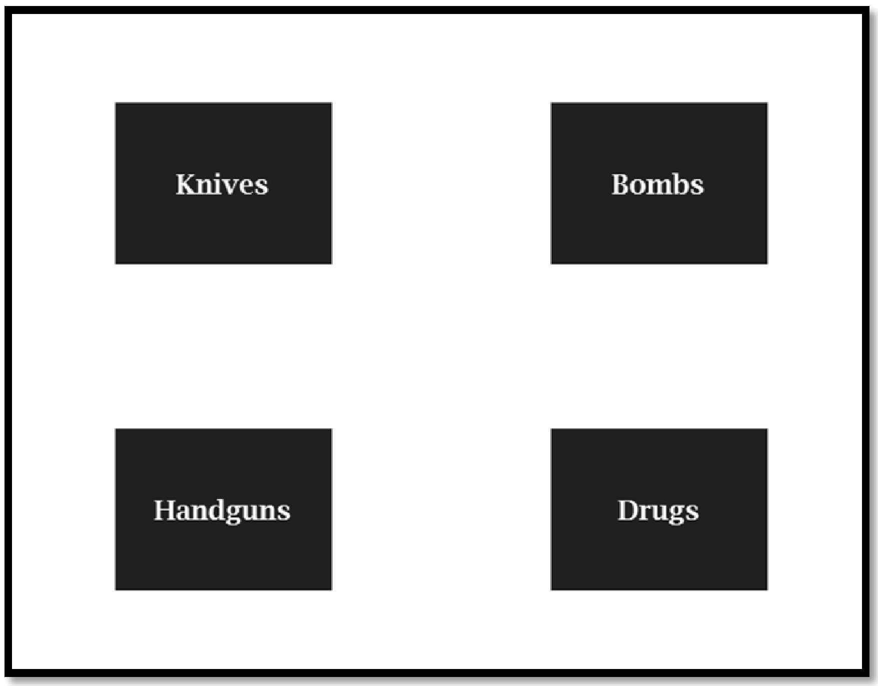

Figure 7. Example of a simultaneous visual foil presented by ASK.

The visual location of the item associated with guilty knowledge was rotated in a balanced fashion. The presentation order of visual foils was similarly rotated.

As an exploratory measure, an additional CIT question presented by the ASK computerized voice read as follows:

[1]In a field environment, control items for the CIT would need to be carefully chosen such that they are certain to not be present. For instance, “rifle” could be an effective control foil item in many security screening field environments: screeners can be certain no rifles are present because they cannot be effectively hidden on one’s person and would be easily detected by a luggage scanner.

The following people are wanted by local authorities. Are you familiar with any of these people?

Similar in format to the first four CIT questions, the ASK displayed four images on the screen immediately following the question (see Figure 8). For participants in the guilty condition, one of the faces evoked knowledge that the participant would desire to conceal: one image represented the same person to whom they were directed to deliver the IED in the instructions that preceded the screening process. Minear and Park (2004) approved and supplied specially designed face images for this experiment. To encourage a sense of realism, I chose to use images of faces that are most likely to remind participants of a stereotypical individual who might be involved in terrorist activity. I also chose faces that were fairly similar in features, to ensure an inordinate amount of attention would not be drawn to a particular face simply because it possessed features that stood out where others did not. This final visual CIT foil was added as an exploratory item; the analysis of this portion of the screening can be found in Appendix C.

Figure 8. Faces CIT foil. This was presented at the end of each screening, immediately after the question: "Are you familiar with any of these people?"

Participants were asked not to disclose details of the experiment to anyone until a date when data collection would be completed. Each participant was also asked whether he or she had heard any details about the experiment prior to participating, and were promised that full credit would be given regardless of their answer to this question.

5.3.3 Experiment 2 Measures

The EyeTech eye tracking system exported raw data in Cartesian coordinate format. The following measurements were derived for each participant and visual foil: the initial direction of the first saccade after each question (dummy coded as 1 if toward critical foil item), and the percentage of time spent gazing at the safety point during the time provided for a response.

6. ANALYSIS AND RESULTS

The datasets for experiments 1 and 2 were analyzed separately. First reported are the analyses for experiment 1 investigating rigidity. For experiment 2, two analyses were performed. The first used eye movement data and the second involved response time.

6.1 Experiment 1 (Rigidity Detection) Analysis and Results

As part of the post-interview questionnaire, respondents self-reported their level of motivation and arousal on 7-point scales. Respondents also answered two questions regarding non-verbal behavioral control. The items used to measure these are shown in the table below.

Table 3. Reliabilities and Means for Self-Reported Motivation, Arousal, and Non-Verbal Behavioral Control

Measure	Items	Reliability	Mean (S.D.)
Motivation	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?	Cronbach's $\alpha = .90$	6.03 (.14)
Arousal	How did you feel during the interview: …Nervous (1-7) …Flustered (1-7) …Tense (1-7) …Relaxed (1-7) (reverse-coded) …Uneasy (1-7) …Stressed (1-7)	Cronbach's $\alpha = .89$	3.11 (.61)

Non-Verbal Behavioral Control: Effectiveness	How effective were you in controlling your nonverbal behavior during the interview?	n/a	4.53 (1.65)
Non-Verbal Behavioral Control: Effort	How much did you try to control your nonverbal behavior (gestures, posture, etc.) during the interview?	n/a	3.99 (1.98)

The difference in perceived effectiveness of controlling non-verbal behavior was not statistically different between the two conditions. Self-reported arousal and motivation levels likewise were not statistically different between groups. For the question regarding effort allocated to non-verbal control, participants in the guilty condition reported significantly ($F = 7.39_{(1,133)}$, $\eta^2 = .053$; $p = .007$) more effort (M=4.59, s.d. = 1.79) than their counterparts (M = 3.65, SD = 2.01).

The video recordings of each interview were analyzed using the computer vision techniques outlined in an earlier section. Because of technical problems with the video recording and analysis system, only 107 of the initial 134 cases produced usable data for analysis. Of these 107 participants, 64 were female. In this subset, 40 participants “committed” the crime, leaving 67 who did not.

A multilevel regression model was specified for overall movement during each foil item. The summation of standardized movement scores for right hand, left hand, and head movement was used as the dependent variable. Multilevel regression models use adjusted standard errors to reflect the uncertainty that arises from variation within subject. The independent variables included Condition (dummy coded: 1 = Guilty, 0 = Innocent), Participant, Foil Item Type (dummy coded:

1 = Critical Item, 0 = Neutral Item). Question, and Interviewer were initially included as covariates but were not significant predictors and were subsequently dropped from the model. The effect of greatest interest is the Condition and Foil Item Type interaction. The results of the multilevel regression model are shown below.

Table 4. Overall Movement: Multilevel Regression Model Results

Fixed Effects	β	β Standard Error
Intercept	0.044 (n.s.)	0.069
Hidden Guilty Knowledge	0.102 (n.s.)	0.109
Foil Item Type	-0.197 (n.s.)	0.171
Hidden Guilty Knowledge : Foil Item Type	-0.624*	0.267

Notes: N = 1887. Model fit using maximum likelihood.
* p < .05, (n.s.) not significant.

To test if the Condition and Item Type interaction provides a significant improvement to the fit of the data, the model was compared to an unconditional model, which omits any fixed effects, using deviance-based hypothesis tests. The fit of the current model was significantly better than the unconditional model, $\chi^2(1, N = 1887) = 17.15, p < .001$.

Depictions of the overall movement patterns are shown in the next three figures. Foil item 4 was the relevant (i.e. correct) item for CIT foils 1 and 3, and foil item 5 was the relevant item for foil item 2.

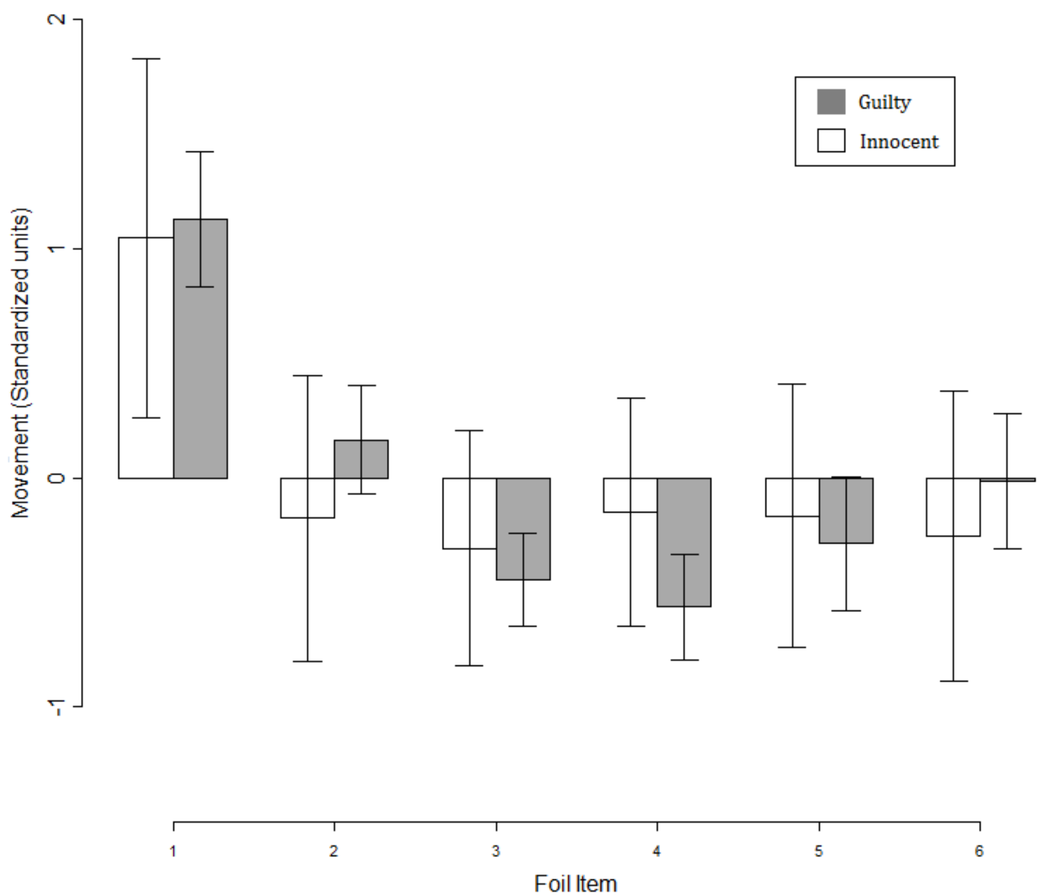

Figure 9. Movement During CIT Foil 1

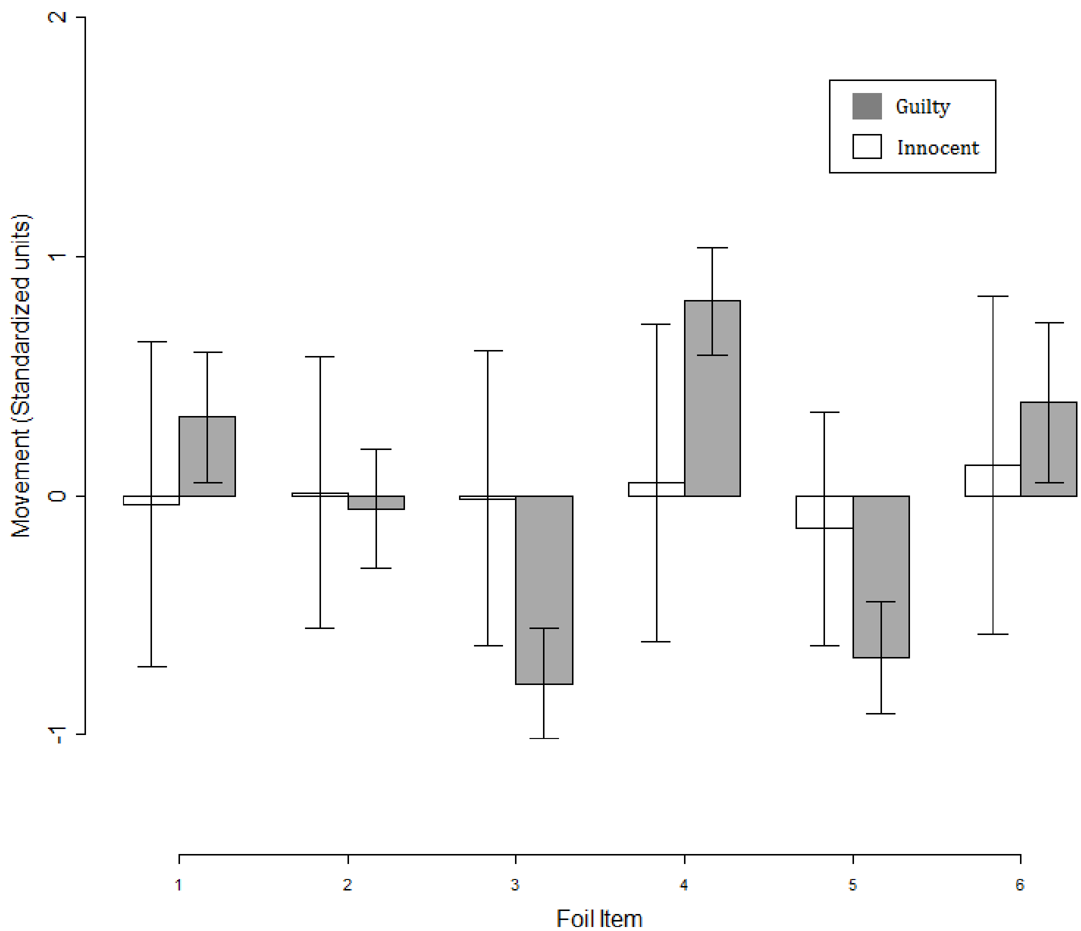

Figure 10. Movement During CIT Foil 2

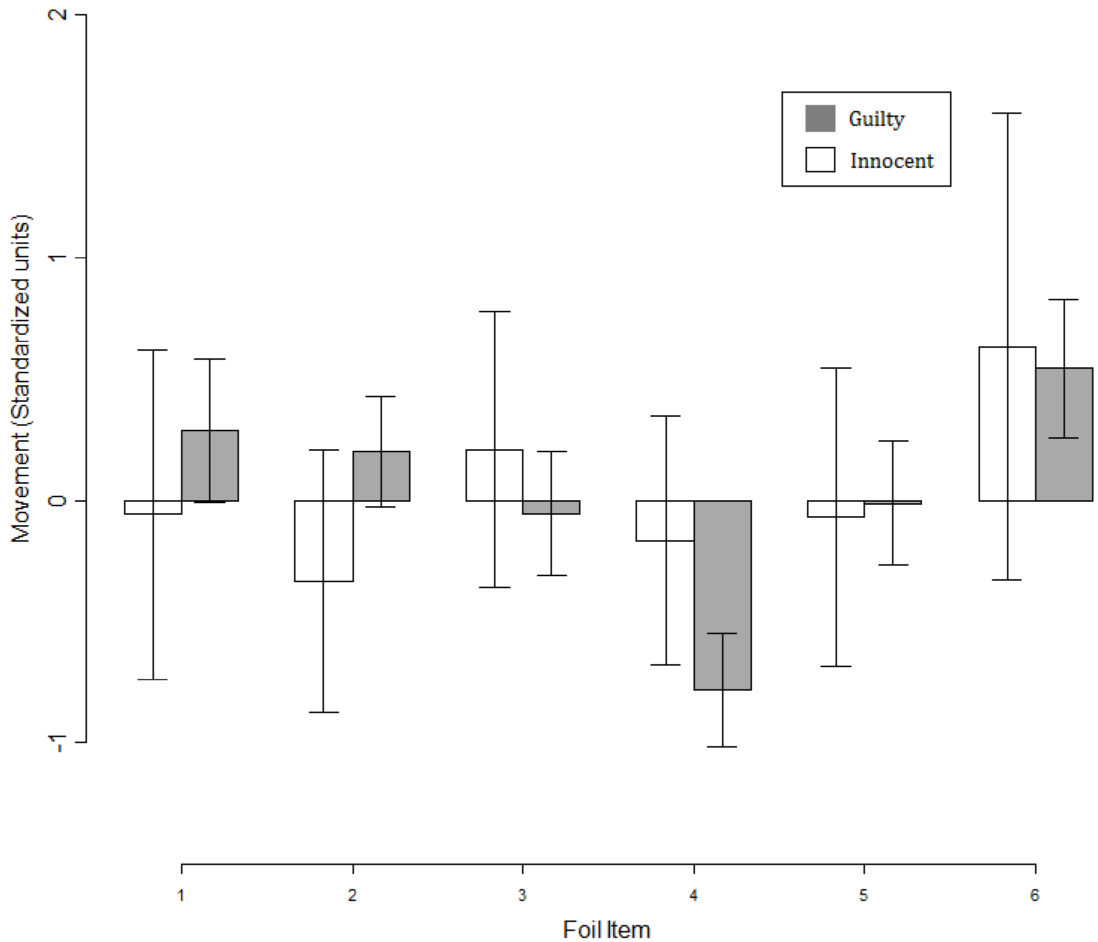

Figure 11. Movement During CIT Foil 3

An exploratory analysis was undertaken to identify whether rigidity was evenly dispersed across the body, or whether it was concentrated in one or more areas. Separate multilevel regression models were specified for the head, right hand, and left hand movement. The results of these models are detailed below.

Table 5. Overall Movement: Results of Separate Multilevel Regression Models

Fixed Effects	Right Hand Movement	Left Hand Movement	Head Movement
	β (S. E.)	β (S. E.)	β (S. E.)
Intercept	.004 (.030)	.001 (.030)	.038 (.030)
Foil Item Type	.010 (.073)	-.006 (.074)	**-.201**** (.073)
Hidden Guilty Knowledge	.044 (.047)	.022 (.047)	.040 (.047)
Hidden Guilty Knowledge : Foil Item Type	**-.293*** (.116)	-.125 (.116)	**-.228*** (.115)

Notes: N = 1887. Model fit using maximum likelihood.
* p < .05, ** p < .01.

Each model was compared to an unconditional model. Compared to the unconditional model, the fit for the right hand and head movement models was significantly better: Head Movement $\chi 2$ (1, N = 1887) = 30.55, p = .000; Right Hand Movement $\chi 2$ (1, N = 1887) = 9.87, p = .020; Left Hand Movement $\chi 2$ (1, N = 1887) = 2.16, p = .540.

The figures below illustrate the relationship between guilt and right hand movement for CIT relevant vs. irrelevant items for two sample CIT foils. On average, a guilty participant's right hand moved less during relevant foil items than irrelevant foil items. Innocent participants showed no difference in their right hand movement. Head movement followed a trend similar to the right hand.

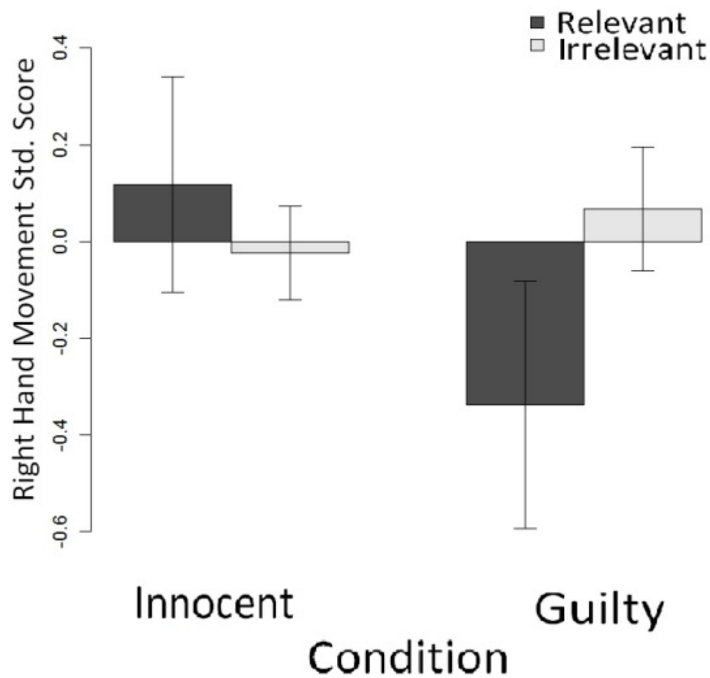

Figure 12. Right hand movement during CIT Foil 2

Figure 13. Right hand movement during CIT Foil 3

6.1.1 Comparison of Head Movement Measurement Techniques

Both the ASM and skin blob tracking computer vision techniques provide usable data for measuring head movement. The ASM data was produced using video of a close-up view of the head, while the skin blob tracking data was produced using a full body frame. The ASM data was chosen for the above analyses because the close-up view provided the ability to measure head movement at the more fine-grained level. In other words, more head movement was visible in the close-up view than in the full-body view. Both camera angles were recorded in standard definition.

Does the close-up view actually improve measurement? To test this assumption, I compare the model of head movement based on the close-up video to a model produced using the full body video. Both models are based on the same participants and experiment, yet the model that employs facial close-up video produced statistically significant results while the other did not. A comparison of the two models is shown in the table below.

Table 6. Comparison of head movement models

Fixed Effects	Head Movement (using full-body video)	Head Movement (using close-up of face)
	β (S. E.)	β (S. E.)
Intercept	.010 (.030)	.038 (.030)
Foil Item Type	-.065 (.073)	**-.201**** (.073)
Hidden Guilty Knowledge	.011 (.047)	.040 (.047)
Hidden Guilty Knowledge : Foil Item Type	-.019 (.116)	**-.228*** (.115)

Notes: N = 1887. Model fit using maximum likelihood.
* $p < .05$, ** $p < .01$.

6.1.2 Classification Accuracy of Rigidity in the CIT

To estimate how effective this pattern might be in a predictive application, I performed a receiver operating characteristic (ROC) analysis to assess the ability of overall rigidity for predicting guilt. The ROC curve reflects the tradeoffs between a true positive rate (sensitivity) and false positive rate (1 minus specificity) when selecting a classification model. If, for example, all interviewees were classified as guilty, guilt detection accuracy would be 100%, but all innocent parties would be incorrectly classified as guilty as well, producing a 100% false positive rate. The standardized movement scores for head and hands for critical items were included as predictor variables. This produced an area under the curve (AUC) of .77, which can be interpreted as 77% prediction accuracy. The underlying logistic prediction model produces a Nagelkerke R^2 of .28 and a Cox & Snell R^2 of .21. The ROC curve is represented graphically in the figure below.

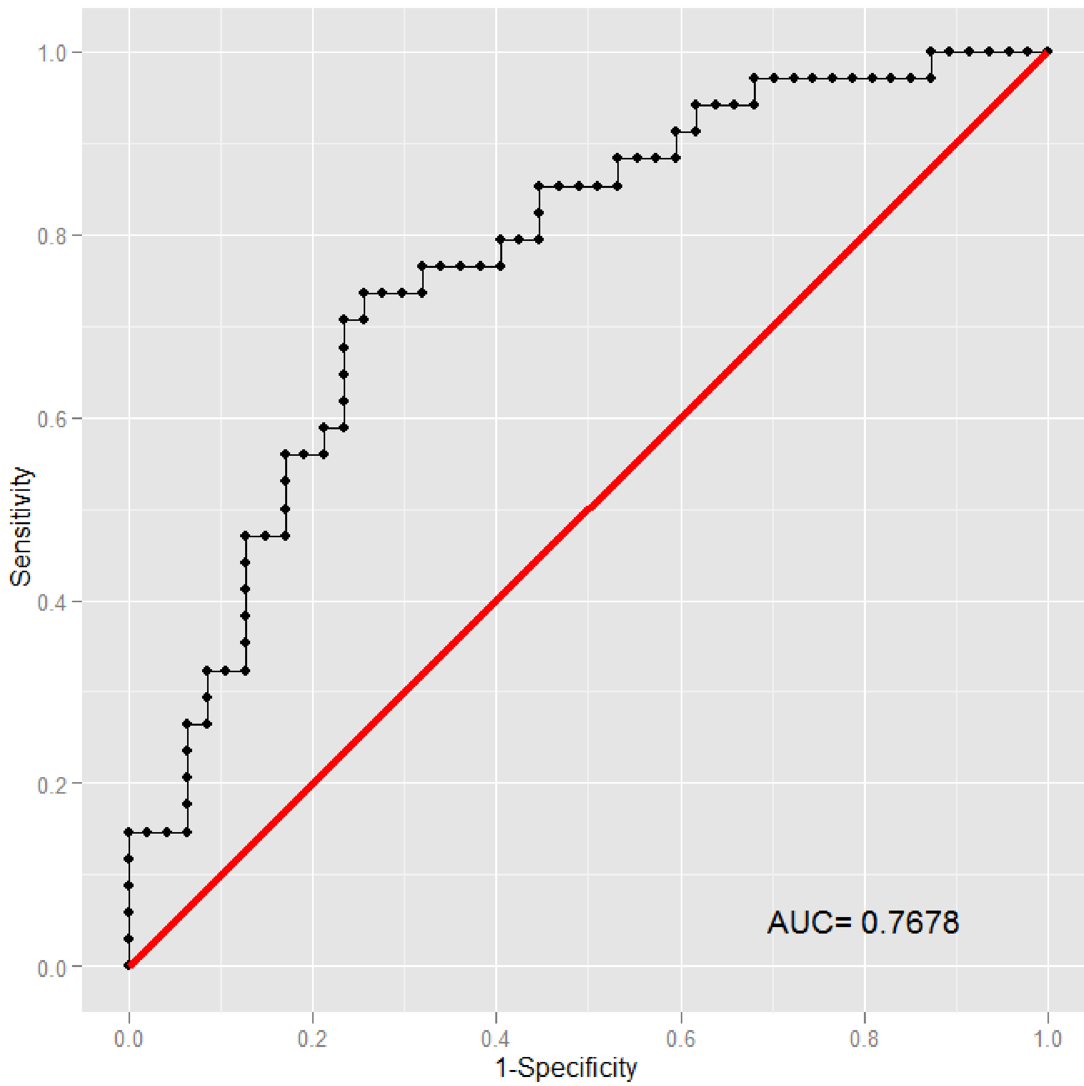

Figure 14. ROC Curve Estimating the Predictive Power of Rigidity in the CIT in Experiment 1.

The 77% prediction accuracy is comparable to the 77% accuracy produced by the professional polygraph examiners on the same subset of participants. These interviewers possessed extensive training and made judgments based on both the CIT and an interrogation conducted prior to the CIT, whereas the automated kinesic classification reported here used only CIT data.

6.2 Experiment 2 (ASK Oculomotor) Analysis

Of the 172 participants, 36 were disqualified either because they did not follow instructions (23 cases), or because of system failures[2] (13 cases). Data from the remaining 136 participants were used in the analysis.

To assess whether the manipulation worked properly, manipulation check questions were added to the post-survey to determine: (1) whether guilty participants possessed concealed knowledge, and (2) whether the concealed knowledge was perceived as aversive by participants. Experiment conductors had visual confirmation that those in the guilty condition constructed and packed the mock IED. However it was important to ensure that participants understood that these actions were illicit, or there would be little reason for a defensive response. The vast majority of participants (57 of 61 or 93%) in the guilty condition correctly reported that they had carried an illicit item through security screening. On the second day, 54 of 57 (95%) in the guilty condition correctly reported their attempt

[2] Physical eye damage and a small percentage of glasses proved problematic. Additionally, the narrowness of the range of allowed head movement caused some of the failures. Alternative eye-tracking systems and modified calibration algorithms will be used to minimize these types of system failures in future iterations of the ASK.

to carry an illicit item through screening. Those participants who either confessed or All participants in the innocent condition correctly reported on both days that they did not carry an illicit item. For each stimulus word, all participants reported how much attention that stimulus captured on a 7-point scale, to gauge their perceived signal value. On day 1, participants in the guilty condition rated the signal value of critical items higher than non-critical items ($F = 107.9_{(1, 120)}$, $\eta^2 = .47$, $p <$ 0.001), while innocent participants reported no significant difference ($F = 2.8_{(1,144)}$, $\eta^2 = .02$, $p = .095$). Day 2 showed similar results, with guilty participants reporting higher signal value on critical items ($F = 37.7_{(1,112)}$, $\eta^2 = .25$, $p < .001$), and innocent participants showing no such effect ($F = 2.1_{(1,132)}$, $\eta^2 = .02$, $p = .15$). Table 2 displays self-reported signal value statistics.

Table 7. Means and standard deviations of self-reported signal value of foil items

	Day 1 Guilty	Day 2 Guilty	Day 1 Innocent	Day 2 Innocent
Critical Foil Items	M=5.46, SD=1.42	M=4.68, SD=1.87	M=4.26, SD=2.32	M=3.96, SD=2.26
Non-critical Items	M=2.69, SD=1.52	M=2.71, SD=1.55	M=3.70, SD=1.65	M=3.45, SD=1.70

A multilevel regression model was specified (n = 1020) using mean time gazing at the safety point (center of the screen) as the response variable. Participant (n = 136) was treated as a random factor, while the experiment condition, foil type (baseline or charged), and participation day were treated as fixed effects. Foil question order and critical item position on the screen were included as covariates.

When a visual foil was charged (i.e., contained a critical item), only participants in the guilty condition spent significantly more time (4.5%) gazing at the safety point ($t_{(1013)} = 3.06$, $p < .01$). The strength of this effect significantly increased (another 3.5%) on day 2, again only among participants in the guilty condition ($t_{(1013)} = 2.70$, $p < .01$). Location of the critical item on the screen (which quadrant) was initially included in the model but was not significant and was subsequently removed. Time likewise showed no significant effect and was removed from the model. Foil presentation order was significant. When the foil containing the word “bombs” was the first charged visual foil presented, gaze effects were more pronounced than when the critical item was “explosives” or “weapons.” Table 3 summarizes the multilevel regression results.

Table 8. Oculomotor Threat Avoidance (Gazing at the Center of the Screen) as Response Variable for the Word CIT Foils

Fixed Effects	β	β Standard Error
Intercept	0.110^{***}	0.014
Hidden Guilty Knowledge	$0.005^{(n.s.)}$	0.019
Participation Day	$0.000^{(n.s.)}$	0.009
Threatening Foil	$0.016^{(n.s.)}$	0.010
Hidden Guilty Knowledge : Participation Day	0.035^{**}	0.013
Hidden Guilty Knowledge : Threatening Foil	0.045^{**}	0.015
Foil Presentation Order 2	-0.021^{*}	0.010
Foil Presentation Order 3	-0.022^{*}	0.011

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

To test model fit, the model was compared to an unconditional model which omitted any fixed effects, using deviance-based hypothesis tests. The fit of the current model was significantly better than the unconditional model, $\chi^2(1, N = 1020) = 64.69$, $p < .001$.

An overall logistic multilevel regression revealed no main effect of condition on the direction of the initial saccade. However, a near-significant interaction effect of condition and participation day was noted, $z_{(765)} = 1.79$, $p = .07$. Separate analyses for each day revealed that for participants with guilty knowledge, the initial saccade was biased toward the critical item during the second day of screening, $z_{(360)} = 2.34$, Nagelkerke $R^2 = .14$, $p = .02$, but not the first day, $z_{(404)} = -0.04$, Nagelkerke $R^2 = .12$, $p = .88$. The model fit for the second day was significant compared to an unconditional model, $\chi^2(1, N = 360) = 8.41$, $p < .05$.

The exploratory CIT question involving faces was analyzed separate from words. Multilevel regression models for the faces question were specified similar to those used for questions involving word stimuli. Hidden guilty knowledge was associated with a 6% increase in the amount of time gazing at the center of the screen, $t_{(249)} = 3.00$, $p < .01$. There were no main or interaction effects of participation day. There were no significant effects of condition on initial saccade for the faces CIT foil.

The results indicate that guilty knowledge significantly affected both the tendency to look toward the critical item in a foil and the tendency to avoid looking at any foil stimuli after initial detection. The hypothesis that guilty knowledge causes visual attention to orient toward a critical item in a CIT foil (H1) is partially supported, with significant results occurring in day 2 but not in day 1. The hypothesis that the orienting reflex would diminish over time (H2) was not supported. The hypothesis that defensive behavior would encourage visual attentiveness toward the safety point (H4) was supported. Finally, the hypothesis that the defensive behavior effect would remain over time was supported (H5).

6.2.1 Oculomotor Defensive Behavior Classification Accuracy

An accuracy analysis can help establish that the initial system has some potential value. The problem of trying to detect concealed knowledge in security screening can be conceptualized as a signal detection problem (Basuchoudhary & Razzolini, 2006; Ben-Shakhar, Lieblich, & Kugelmas.S, 1970).

Since oculomotor defensive behavior showed the strongest results, a receiver operating characteristic (ROC) analysis was performed on those data for each day. Condition was positioned as the response variable, with gaze patterns during word visual foils positioned as the predictor variables. For day 1, the oculomotor defense patterns produced an area under the curve (AUC) of .67; an AUC of .68 was produced from day 2 data. Graphs representing the ROC curves are depicted below.

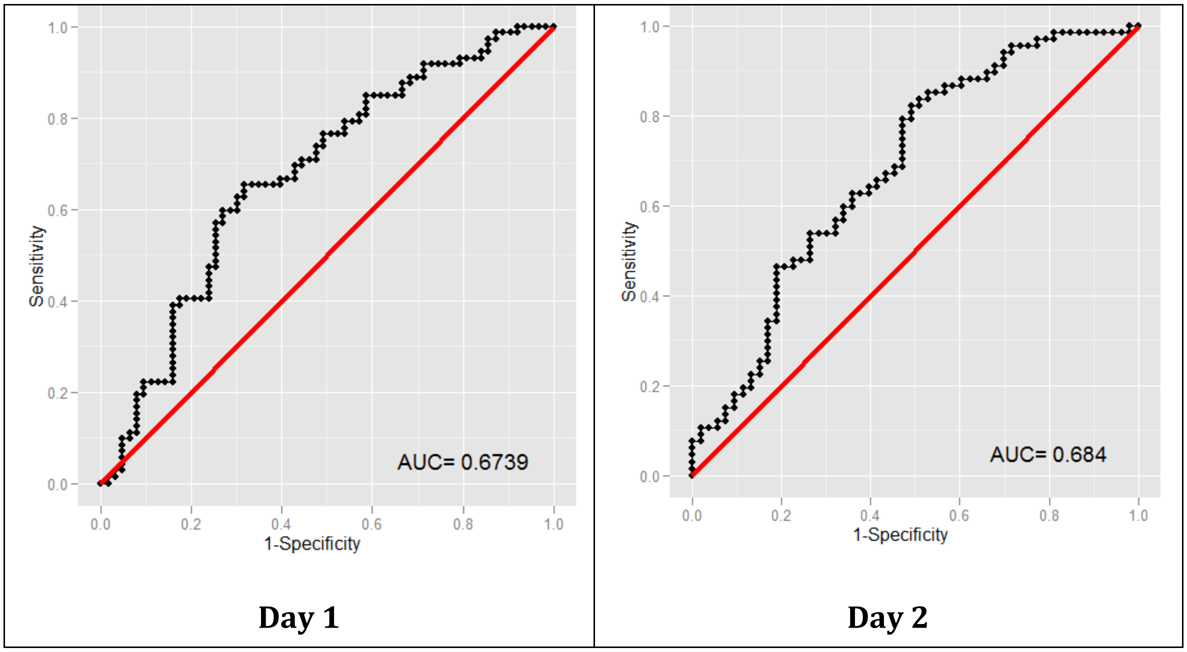

Figure 15. ROC Curves for Day 1 and 2 Oculomotor Defensive Behavior

6.3 Experiment 2 (ASK Response Time) Analysis

Response times were calculated by manually coding and calculating the time lapsed between the end of each question and the beginning of each response. Because these response time analyses were performed as part of a separate research project, some of the results summarized here are expected to also be reported by Hsu (2012). However, here I include additional analyses and discussion.

A multilevel regression model was specified (n = 1020) using response time as the dependent variable. Participant (n = 136) was treated as a random factor, while the experiment condition, foil type (baseline or charged), and participation day were treated as fixed effects. Foil question order and critical item position on the screen were included as covariates.

As reported in Table 9, there was an overall main effect of participation day and foil charge. There was no main effect of condition, but there was a three-way interaction effect among condition, foil charge, and participation day. There was also a significant two-way interaction between foil charge and participation day. The order of presentation likewise had a significant effect: response time decreased slightly but significantly when the foil with the word “Bombs” was presented second rather than first or last.

When a visual foil was charged (i.e., contained a critical item), all participants regardless of condition exhibited significantly shorter response times. Response times also decreased the second day of participation; however, this effect was moderated by the presence of a threatening foil item and completely disappeared for those in the guilty condition. In other words, on the second day of participation, guilty participants took longer to respond when foils that contained threatening items were presented (compared to those in the innocent condition).

Table 9. Response Time Multilevel Regression Results

Fixed Effects	β	β Standard Error
Intercept	3.230***	0.092
Hidden Guilty Knowledge	0.105 (n.s.)	0.127
Participation Day	-0.699***	0.089
Threatening Foil	-0.534***	0.071
Hidden Guilty Knowledge : Participation Day	-0.129 (n.s.)	0.132
Hidden Guilty Knowledge : Threatening Foil	-0.140 (n.s.)	0.103
Participation Day : Threatening Foil	0.378***	0.102
Hidden Guilty Knowledge : Threatening Foil : Participation Day	0.398**	0.151
Foil Presentation Order 2	-0.124*	0.052
Foil Presentation Order 3	-0.041(n.s.)	0.058

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

The response time model fit was strong compared to an unconditional model, $\chi^2(1, N = 1020) = 228.62$, $p < .001$.

6.3.1 Response Time and Combined Classification Accuracy

A ROC analysis was performed that again specified condition as the response variable and used response times from the first 4 slides as predictor variables. For day 1, the ROC analysis produced an AUC of .57; an AUC of .66 was produced from day 2 data. Graphs representing the ROC curves are depicted below.

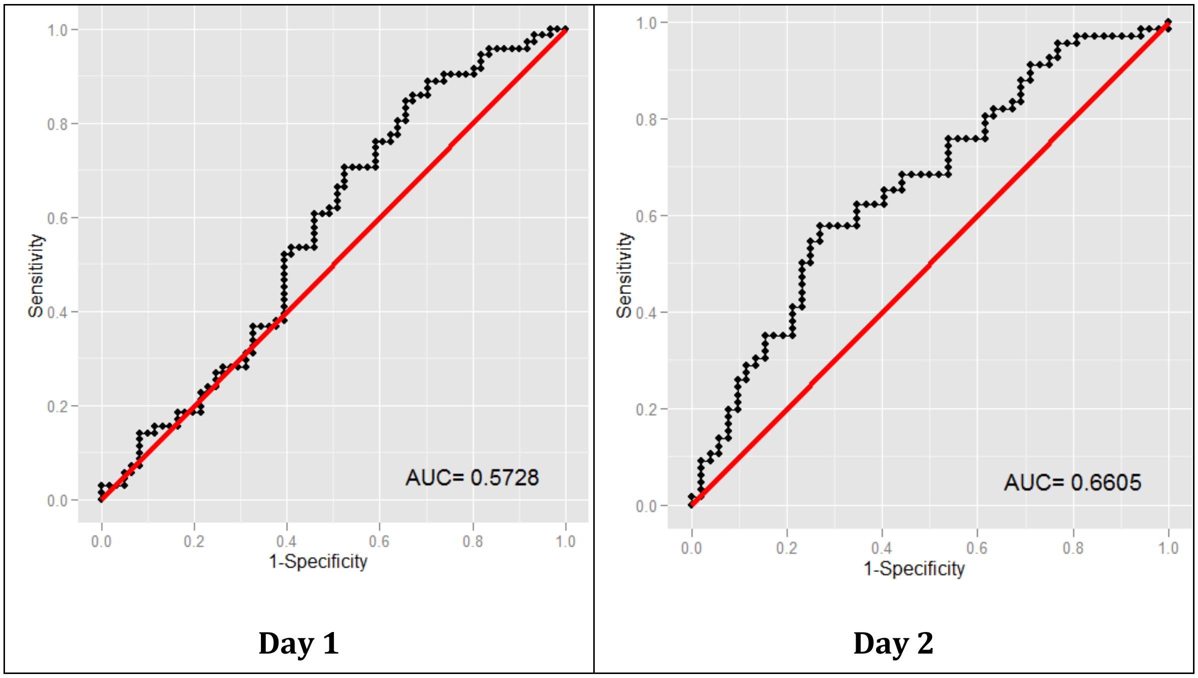

Figure 16. ROC Curves for Day 1 and 2 Response Time

A final ROC analysis was performed that combined both response time and oculomotor defensive behavior. Time spent gazing at the center of the screen for the first four questions were included with response times for the same questions as predictor variables. For day 1, the ROC analysis produced an AUC of .70; an AUC of .72 was produced from day 2 data. Graphs representing the ROC curves are depicted below.

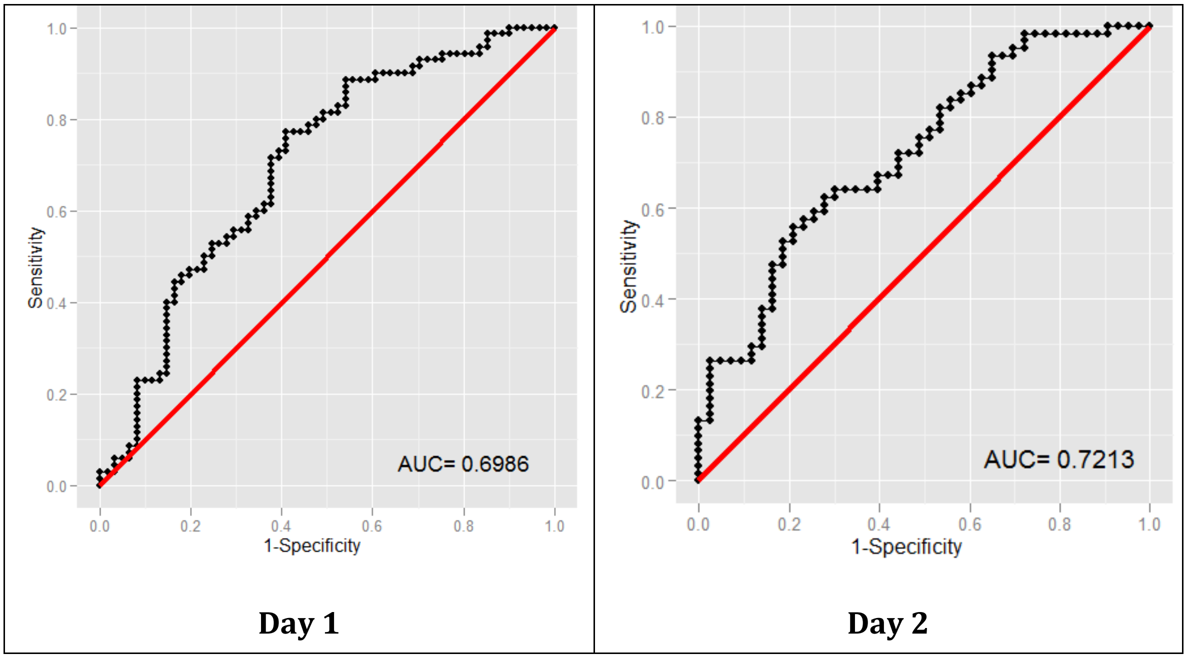

Figure 17. ROC Curves for combined oculomotor defensive response and response time.

Though the combined classification shows improvement over the response time or oculomotor defensive response models independently, analysis of deviance tests suggest that the improvements may not be significant for either Day 1 ($\chi2$ (1, N = 100) = 3.76, p = .44) or Day 2 ($\chi2$ (1, N = 100) = 5.32, p = .26).

7. DISCUSSION

The purpose of this study was to propose and investigate theory, protocol, and system design for the automated discovery of hidden knowledge. I chose to center the ASK system design on the CIT screening protocol because of its automation and interaction simplicity, and because it is considered to have the highest validity in previous research. I applied both orienting and defensive response theories to this novel area of research. This study proposed a method for using computer vision techniques to assess rigidity, and for the first time rigidity and oculomotor variables were investigated as potential cues to deception in a CIT. The table below provides a summary of the hypotheses in this study.

Table 10. Summary of Hypotheses Support

Hypothesis	Supported?
H1. Guilty knowledge increases the likelihood that an initial saccade will be directed toward the critical item in a collection of simultaneously presented CIT foil items.	Partially
H2. Over time, stimuli representing guilty knowledge will be less likely to attract the initial saccade.	No
H3: Guilty Knowledge will increase response latency when a critical item is presented.	Yes
H4. Guilty knowledge increases time spent gazing at the safety point.	Yes
H5. Guilty knowledge increases critical stimuli avoidance during repeat exposures.	Yes
H6: Guilty knowledge will increase body movement rigidity while a critical item is presented.	Yes

The results provide support for the potential for automated rigidity detection and oculomotor defensive behavior as alternative measurements in an ASK-like human screening system design. Because of the relative novelty of this line of research, there are many caveats, lessons learned, and areas for future research to consider. The results will first be discussed separately and then in combination.

7.1 Experiment 1 Discussion

The purposes of experiment 1 were two-fold: the first was to implement automated rigidity detection and glean insights from its use. The second was to provide empirical evidence of the rigidity effect during a CIT. These two factors inform the broader goal of investigating the potential of the proposed system design for automated human screening for concealed knowledge.

7.1.1 Experiment 1 Summary of Results

Several insights were gleaned from implementing the automated rigidity detection system. First, in spite of very little movement being present throughout the CIT, the system was able to detect movement variations that discriminated innocent from guilty participants in a realistic scenario. Second, greater ability to detect movement (i.e., through higher-resolution video), appears to strengthen the results.

The empirical evaluation of rigidity in a CIT produced very interesting results. An overall significant decrease in movement was found among those who had stolen a ring when an interviewer presented the CIT critical item as compared to when a non-critical item was presented. This significant decrease was present across all three CIT foils. When total movement was broken down into head, right hand, and left hand movement, head and right hand movement proved significant while left hand movement did not, though it trended in the same direction.

7.1.2 Experiment 1 Key Contributions

The first key contribution of experiment 1 is evidence that supports the proposition that rigidity is an active phenomenon in a CIT. Past research has identified rigidity in less controlled interviewing environments, but contextual variables such as question type effects may complicate or confound results. The CIT, however, provides a simple, standardized, reproducible method that is free of question type effects. This standardization and control provide a high level of internal and external validity which benefits both research and practice.

In the experiment 1 CIT, interviewees did not explicitly lie. But although no verbal deception occurred, the results showed that interviewees with guilty knowledge exhibited defensive responding in the form of rigidity. While prior research has confirmed that rigidity is associated with deception, most of these studies have focused on active deception that involves communicative gesturing. Kinesic cues to deception in the CIT have been largely ignored, presumably because of the lack of major movement. However, by applying defensive response theory, this study provides evidence for the potential of kinesic cues in a CIT context. Specifically, whenever a threat is perceived among alternative benign stimuli, rigidity in the CIT should increase.

Though rigidity appears to stem from defensive responding, this research stopped short of discovering the precise type of response(s) that caused the rigidity. However, one theoretical contribution of this work is to show that rigidity occurs even when very little fabrication is required. Some prior research has suggested that a higher demand on cognitive resources for fabrication may be a cause of rigidity, as there should be fewer cognitive resources available to allocate toward non-verbal messaging. But the results of this study indicate that rigidity can occur even when demands for fabrication are minimal.

A second major contribution of this study was the introduction of automatic detection of rigidity. Where in the past, the high cost of manual rigidity assessment in standardized interviewing has limited its use to interested researchers, this study shows that an automated, near-real time system for measuring and reporting rigidity is feasible. A post-hoc, qualitative viewing of a sample of the interviews from experiment 1 indicated that human coding of rigidity during a CIT may be very difficult, because for most cases, human observers seem to find it difficult to distinguish the relative amount of movement among foil items. This idea is further bolstered by the fact that higher resolution video of the interviewee's head was required to discover the rigidity effect—standard definition video using full-body frame was insufficient to detect the significant variation in head movement.

The left hand was similar to the head in that a post-hoc qualitative visual review indicated that in most cases, the left hand exhibited much less variation in movement. There was no video focused in on the left hand, so it is not possible with the current dataset to determine whether a rigidity effect was present in that hand.

The majority of movement throughout the CIT was in the right hand. No data were collected on the handedness of participants, but based on general population statistics one can estimate that about 90% of participants were right-handed. It seems likely that even minor, random movements are most likely to be performed by the dominant hand. Future research in this area may benefit from including handedness in the explanatory or predictive model.

7.1.3 Experiment 1 Limitations and Future Research

The ~77% prediction accuracy of the rigidity effect reflected the ~77% prediction accuracy of trained polygraph professionals. This was true in spite of the fact that the professionals had more data to work with (i.e., the pre-CIT interview). However, it is unclear if the portions of the interview prior to the CIT helped or hindered the interviewer's ability to make a judgment. Similarly, the interview questions prior to the activity could have had some preparatory or dampening effect on the CIT. Since the ultimate goal of this research is to move toward an automatic human screening system, future research will need to investigate these factors and determine how they might play a role.

In the same vein, the effects of the interviewer are unclear. Even though the interviewer plays a minimal role in the CIT, it is possible that a human may unconsciously send indicators to the examinee as to the critical item in a CIT foil, just as this study has shown that it is possible for the examinee to unconsciously give away their own guilty knowledge. Factors such as interviewer gender, dress, demeanor, timing, eye contact and other characteristics and behaviors could affect the level of anxiety or perceived threat of the interviewee. Interestingly, when the critical item in a foil was presented, increased rigidity in head movement was present even among *innocent* participants, though to a lesser degree than among guilty participants. Post-hoc analyses showed that this phenomenon was present during all three CIT foils, though strong enough to be statistically significant only in the third foil. Further exploration is needed to determine the source of this anomaly and if it can be remedied using an automated interviewer such as a disembodied voice or other conversational agent.

Finally, it is clear that technology and process improvements can and should be made on this initial investigation. However a 77% accuracy rate is an encouraging result for the first iteration. Future research should focus both on improving the prediction capability of automated rigidity measurement, and combining rigidity with other cues to concealed information for a more comprehensive understanding of the interaction.

7.2 Experiment 2 Discussion

Experiment 2 also had two main goals. The first was similar to experiment 1 in that it investigated and instantiated and alternative indicator of concealed knowledge, this time targeting oculomotor and response time indicators. Secondly it allowed an evaluation of the performance of an instantiated automated human screening system in the form of an ASK. This section will discuss the results of the oculometric analysis; the next section will discuss the performance of the ASK against the design criteria.

7.2.1 Experiment 2 Summary of Results

The results support the hypotheses that oculomotor defensive behavior would be apparent in participants who possessed guilty hidden knowledge. As predicted, participants carrying a mock IED tended to avoid gazing at foil items—choosing to spend more time gazing at the center of the screen (i.e., the "safety point") where the expected visual stimulus was unrelated to the test. This effect remained constant even on the second day of participation when participants were familiar with the ASK.

The orienting response, traditionally measured in practice via monitoring electrodermal activity, heart rate, and/or respiration, was somewhat effectively measured using eye-movement patterns, as is commonly seen in cognitive psychology research. Participants who carried the mock IED were more likely to orient their initial visual attention toward the critical foil item presented by ASK. However, this effect was seen only on the second day of participation.

Response time followed a slightly different pattern. All participants took more time to respond to foils that contained critical items. This finding may be confounded by an ordering effect: the non-threatening foil always came first. Participants may well have simply taken more time to respond during the first foil than to subsequent foils.

Response time decreased on the second day, presumably because participants were more familiar with the process and therefore required less high-level cognition, or were able to use cognitive resources more efficiently. However, guilty participants' response time did not decrease on the second day when threatening items were presented. This is somewhat consistent with previous research that shows an increased response time when critical items are presented (Seymour, et al., 2000). These results suggest promise for using this measure in automated screening. However, because of limitations in the experimental design and outstanding research questions regarding the easiness of countering this measure (Gronau, et al., 2005), further research should be done before drawing conclusions on the validity and reliability of response time for automated screening systems.

Overarching the hypotheses was the proposition that a CIT-based system like the ASK could be automated and extended to non-traditional domains for the discovery of valuable concealed knowledge. While further research is needed to refine the ASK design, the initial results are promising. The ASK operated automatically, with little need for manual intervention, and utilized the CIT framework to detect concealed information at a rate greater than chance and unaided human judgment.

7.2.2 Experiment 2 Contributions to Research and Practice

There are three major research contributions of this second study. This first is the application of orienting and defensive response theories to automated screening for concealed information. Second, oculomotor cues to concealed knowledge were identified, and response time was investigated in this new application. Third, these theories were successfully instantiated in an implemented automated system, where further insights were gleaned beyond the hypotheses specifically tested.

While the orienting reflex has been part of the traditional CIT since its inception, defensive responding has not been a major focus of CIT research. In this study orienting and defensive responding were measured simultaneously through oculomotor indicators of concealed knowledge. Oculometric measures have been used in research settings since 1902 and has recently found application in IS and deception detection research. However, this study is among the first to use eye movement as an indicator of concealed knowledge. The orienting reflex measure used in this study (the direction of the initial saccade) is common to cognitive psychology research in that eye focus is assumed to orient toward stimuli perceived to be novel or personally significant. Detecting defensive responding via oculomotor patterns, however, is an underexplored area. This study found strong effects of defensive responding in eye movement patterns. These effects could also have implications for eye tracking in human-computer interaction research, which tends to analyze what individuals do look at, when in fact what a person does not look at may have more signal value if it is threatening in nature.

7.2.3 Experiment 2 Limitations and Future Directions

Order effects and stimuli type are an important limitation to consider in light of the results of experiment 2. Some of the results were potentially confounded by the fact that non-threatening foils were always presented first, rather than being interspersed throughout the foils that contained threatening items. This seemed to have an effect especially on response time. Future research should investigate order effects further.

There were at least two areas for improvement regarding the modified CIT used in this study. First, images may perform better than words as stimuli in this system. Though the orienting reflex can occur prior to detection, it is unclear whether four words can be subconsciously identified prior to the initial saccade. Images may be more easily processed by parafoveal visual attention and thereby may produce improved results. However, the color, tone, and type of images can in and of themselves demand an orienting response, so careful selection and pretesting of images would be an important step to take.

The second area for improvement involves a "hurry up" instruction. Encouraging the participant to respond as quickly as possible should increase the likelihood that eye movements will be reflexive rather than overt. An example instruction that could be given at the beginning of the interaction is: "Please answer questions as quickly as possible. Extended hesitations may lead to additional screening." The orienting reflex is a reflex, and is probably more likely to influence eye movement when that eye movement is not being consciously controlled.

7.3 Evaluation of ASK Design

The ASK system and associated process marks the beginning of iterative research into human screening systems searching for concealed information. This section discusses the various aspects of ASK in light of the design principles inspired by the CIT.

The first two design principles are repeated below:

1. *Identify appropriate stimuli that represent the concealed knowledge in question. Ensure there is reasonably strong certainty that the representation has relatively high personal significance for a person who is concealing such knowledge.*
2. *Identify irrelevant stimuli that each arouse the same baseline level of orienting and defensive responses.*

The ASK process identified appropriate stimuli by selecting first the target stimulus of interest (in this case, an explosive), then selecting and pretesting (via a survey) items that produced a similar orienting response. High personal significance of the critical item was ensured (manipulation checks supported this) because of the controlled nature of the experiment.

In a field implementation of ASK, selecting and pretesting items can be just as simple to do. There is no way to be 100% certain of personal significance in a field setting, but because the crime is in progress or planned, high personal significance should be even more likely than what one would expect to see in criminal investigations where the crime may be several months removed from the CIT being performed.

3. *Develop several foils consisting of relevant and irrelevant stimuli in a one-to-many ratio.*

The foils used in ASK were an adaptation of the standard CIT such as the one performed in experiment 1. The foils were presented visually and simultaneously. Visual, simultaneous foils decreased the time necessary to conduct a screening and seemed to elicit strong defensive responses, but further investigation is needed to determine the most effective way for an ASK to present the foils.

4. *Automatically present these foils in an environment free of potential distractions, including but not limited to human distractions.*

Distractions have potential to be problematic for a CIT because orienting and defensive responses are sensitive to alternative stimuli. The ASK used in this study was situated in a separate room where virtually no outside sound could be heard and no interruptions would occur. In a field environment, a similar setup would be ideal: a location free from alternative stimuli. This may mean a booth-like setup in a rapid screening application or a separated office in a more investigative application.

A weakness of the ASK was it was not height or depth adjustable. Even though individuals sat down for the short screening process, the monitor and chair often had to be manually adjusted to properly capture eye movement. An eye tracking system that allows more freedom in head movement and/or an ASK that has greater height and depth adjustability would be an important improvement to ensure the best capture of oculometric data.

5. *During foil presentation, automatically capture human indicators of the orienting and/or defensive responses.*

The ASK collected eye tracking, kinesic, and vocalic data, but required post-processing to translate this raw data into usable indicators. Field implementations of an ASK need indicators and risk assessments in near-real time. Foil segmentation is one of the challenges to this design principle. All captured data needs to be segmented temporally in order to understand which eye movement, body movement, and so on is associated with which foil or foil item. For experiment 1 and part of experiment 2, this segmentation process was done manually. The ASK in this study did automatically segment the eye tracking data by allowing a set amount of time for a response before moving to the next foil. The data for each succeeding foil was tagged according to which time segment it fell under. Other methods that could be used to segment data are through speech recognition or asking the examinee to press a button when his or her response is complete.

6. *Apply categorization algorithms to the collected data to assess concealed knowledge.*

The categorization algorithms performed on the data collected by the ASK exceeded chance levels, but require more sophistication and/or combination with other measures to be viable for a field context. Additional non-invasive, non-contact candidate measures may in include measuring uncertainty and stress via vocalic analysis and measuring the orienting response with via pupillometry algorithms applied to high-speed cameras.

7.4 Overall Key Contributions

This research shows that there is promise in extending the traditional applications of concealed knowledge testing. The traditional CIT is a powerful method for this type of knowledge discovery testing. In the traditional CIT, an individual's orienting response when presented with irrelevant stimuli is compared to his or her response when presented with relevant stimuli. Traditionally, the CIT protocol has not been widely adopted in practice, except in a small number of criminal investigations (Japan being a notable exception, where it is more commonly used). This study takes the first steps toward overcoming the barriers that have prevented more widespread use in both criminal investigations and alternative arenas. This is accomplished by designing the automated screening kiosk (ASK) system to leverage the core principles of the CIT, while introducing alternative measures that do not require human intervention.

Alternative or enhanced measures such as kinesic and oculomotor cues to concealed knowledge can improve the portability and simplicity of the detection of concealed knowledge, because they eliminate the need for specialized personnel or invasive equipment. The non-invasive, low-cost nature of an ASK design may be a welcome contrast to traditional techniques that use SCR, respiration, and heart rate monitors, as well as more recent CIT techniques using fMRI, all of which can be prohibitively expensive for widespread use in practice.

The positive results of this study suggest that a simple procedure could be developed to complement screening at high risk locations such as airports, public events, or border crossings. Unlike current screening systems which attempt to find the threat itself (e.g. by checking x-ray images for shapes that look like weapons), an ASK system may not identify the specific threat, but rather indicate that further investigation is warranted. Adding an abstract layer to the screening system might be comparable to an antivirus program that isn't limited to searching for threats that have signatures identified in the past, but also investigates program behaviors that might indicate a need for a closer look.

The search for concealed information need not be limited to municipal organizations. In locations where law allows, businesses can use an ASK system to improve internal security or help prevent insecure behavior. For instance, an ASK could determine which employees are most likely to be leaking sensitive company information. An ASK interaction could also become part of a regular security policy review course. Where an employee may not be willing to openly admit negligence or mistakes with regard to secure behavior, an ASK could help discretely determine which security protocols are likely to be a concern on an individual or aggregated level. An ASK could potentially apply to many situations where an organization needs to know about specific events but its members are motivated to hide what they know.

Even the traditional CIT as it is used today may also benefit from integrating these additional indicators. While traditional measures rely exclusively on the orienting response, the results of this study indicate that defensive responding can also distinguish concealed knowledge from lack of knowledge, when that the concealed knowledge in question is associated with a level of anxiety to the examinee. Drawing from more than one underlying mechanism can make countering a CIT more difficult. CIT practitioners and researchers can also use an automated approach to control for interviewer effects.

As this work is expanded and refined, this proactive threat-detection system can conceivably complement current physical security screening processes. Similar to behavioral virus detection in virtual environments, screeners can use an ASK-like system to detect threats or valuable concealed knowledge in a physical environment—even if the specifics of the threat are unknown.

Though physical security was the chosen field of interest for this study, the findings of this study can potentially extend to many similar applications, such as internal fraud investigations, policy compliance examinations, and uncovering B2B espionage. In each case, the ASK will need to be tailored to the situation at hand, relying on the design principles put forth in this paper.

8. LIMITATIONS AND FUTURE RESEARCH

Several limitations of this study provide important opportunities for future research. This original study is at the beginning of a stream of research that will further investigate concealed knowledge detection. Among the limitations that need to be addressed are the "witness" problem of the CIT, further refinement of the process and system, and the potential for defeating countermeasures.

Distinguishing between guilty persons and innocent witnesses is an inherent limitation of the CIT (Gamer, 2010). For instance, in a security screening scenario, it is feasible to witness or to know about an illegal or improper activity, while not having participated in the activity directly. For instance, an examinee may purposely conceal knowledge that a friend is smuggling drugs or weapons through a screening checkpoint, though the examinee may not be personally participating in the activity. Further investigation of such guilty knowledge may or may not be desirable in that situation. Though the current study did not address this limitation, future work should assess how this witness problem may be minimized.

Future research will improve on the design principles of automated human screening systems and the methods of measuring the orienting and defensive responses. There are at least two possible reasons why the orienting reflex was not as effective as expected. First, cognitive psychology research on the orienting reflex usually includes instructions to the participant to respond as quickly as possible. This additional instruction may facilitate the autonomic visual reflexes. A second possibility is that word-based cues require more peripheral attention processing than relatively lower-level visual cues such as an image of an object. Future research is needed to investigate these considerations.

Future research will also investigate additional non-invasive methods for identifying concealed knowledge. Some promising potential technologies include face movement analysis for detecting emotion, vocalic analysis for measuring voice stress, pupillometry for measuring the orienting reflex, and linguistic analysis for detecting strategic message manipulation.

Using several methods simultaneously may be critical for overcoming countermeasures. *Countermeasures* are methods examinees use to “counter” or beat the test system. Countermeasures have been shown to be somewhat effective against the CIT and related veracity assessment techniques. There is some indirect evidence that the use of multiple heterogeneous sensors designed to detect different effects of concealed knowledge may help deter countermeasures (D. C. Derrick, et al., 2010; Nunamaker, Derrick, Elkins, Burgoon, & Patton, 2011). A combination of effective measures may make countermeasures much less effective, because an individual’s limited cognitive capacity should hinder the number and type of countermeasures that can be simultaneously employed.

The apparatus and sensors for a second instantiation of an ASK has already been assembled and will soon be developed into a revised ASK. This second iteration begins to address many of the limitations mentioned here. A description of the second ASK iteration is found in Appendix A.

Finally, it is important to note that a CIT screening system will likely never be foolproof. Thus, its main contributions will be as risk assessment and decision support, and as a means of automatically identifying extremes—those true who are extremely likely and unlikely to be a concern. These objective assessments can free up human resources to focus more time on those cases most likely to be a problem.

9. CONCLUSION

This research proposed, implemented, and evaluated a human screening system design for the discovery of the presence of valuable concealed knowledge. The ASK system design draws from psychology research on hidden information discovery, improving on existing CIT research by detecting the orienting reflex and defensive response non-invasively and automatically. To accomplish these goals, alternative automated methods for detecting these underlying mechanisms were tested and shown to have merit. A CIT screening system such as the ASK can indicate the presence of potential threats or malintent even if a specific threatening activity has not yet been tried before. The ASK system design also decreases the need for specialized training and lengthy setup times. This work adds to concealed knowledge detection research by introducing new oculometric and kinesic variables as indicators of detecting guilty knowledge. An ASK system could serve as an inexpensive first-level screening filter and also as a decision support system for investigative activities in municipal and business settings.

10. APPENDIX A: THE SECOND ASK ITERATION

A second instantiation of a screening system based on the ASK design principles has already been assembled. This ASK is pictured below.

Figure 18. The second iteration of an ASK system.

This second iteration overcomes limitations of real-time feature generation by using a Microsoft© Kinect™ sensor to collect body movement data. The Kinect improves on previous methods by collecting data in real time rather than using post-processing on traditional video. Rather than tracking positions of head and hands only, the Kinect generates frame-by-frame positions of 20 points on the body. The Kinect also tracks these points in three dimensions using a stereoscopic camera. Basic software for capturing this data has been developed and can be found in Appendix B along with sample data output. A force platform that independently measures movement on the left- and right-hand sides of the body will also collect kinesic data. These improvements will provide increased ability to measure rigidity and also investigate additional kinesic cues to concealed knowledge.

The Kinect does not track facial movement, so the ASK includes a high definition camera trained on the face and will incorporate facial movement tracking software. A microphone for voice recognition and vocalic analysis is affixed. An eye tracking system that collects blink, pupillometric, and eye movement data is also included. The ASK has a touchscreen for tactile interaction and measurement.

The ASK is also placed on a reticulating arm, allowing the apparatus to be adjusted to optimal height and depth for a given person and context. The ASK is easily mounted to a wall or a pole and can be easily switched back and forth. Alternate sensors can also be added or taken away with relative ease. The purpose of this ASK is not to serve as a finished model, but as a configurable platform for further iterations of human screening experimentation.

11. APPENDIX B: CODE AND DATA SAMPLES FOR REAL-TIME KINESIC DATA CAPTURE SOFTWARE

Below is some sample data that is generated by the ASK in real time using a Microsoft Kinect sensor.

Time	Millisecond	SkeletonID	JointID	JointX	JointY	JointZ
6:16:08	39	4	HipCenter	0.06166497	-0.1605196	2.747354
6:16:08	39	4	Spine	0.05519164	-0.099904	2.804271
6:16:08	39	4	ShoulderCenter	0.06627864	0.2650657	2.832123
6:16:08	39	4	Head	0.07043894	0.4582048	2.821789
6:16:08	39	4	ShoulderLeft	-0.1033907	0.140487	2.788962
6:16:08	39	4	ElbowLeft	-0.1819153	-0.07149	2.764786
6:16:08	39	4	WristLeft	-0.2609192	-0.2773067	2.702001
6:16:08	39	4	HandLeft	-0.3152553	-0.3335559	2.660625
6:16:08	39	4	ShoulderRight	0.2483098	0.142146	2.845469
6:16:08	39	4	ElbowRight	0.3048614	-0.07658	2.828591
6:16:08	39	4	WristRight	0.3707651	-0.2923496	2.80153
6:16:08	39	4	HandRight	0.4662179	-0.5522138	2.788885
6:16:08	39	4	HipLeft	-0.01831	-0.239352	2.714637
6:16:08	39	4	KneeLeft	-0.1087503	-0.5795925	2.655539
6:16:08	39	4	AnkleLeft	-0.1062556	-0.8309524	2.594287
6:16:08	39	4	FootLeft	-0.1400814	-0.7488239	2.603081
6:16:08	39	4	HipRight	0.139947	-0.2386055	2.744128
6:16:08	39	4	KneeRight	0.1758052	-0.5844672	2.648381
6:16:08	39	4	AnkleRight	0.1864027	-0.8381631	2.617846
6:16:08	39	4	FootRight	0.1557258	-0.7717767	2.610299
6:16:08	54	4	HipCenter	0.06172369	-0.1580266	2.750765
6:16:08	54	4	Spine	0.05535929	-0.0981434	2.807629
6:16:08	54	4	ShoulderCenter	0.06669749	0.2651726	2.83489
6:16:08	54	4	Head	0.07049168	0.458257	2.825631
6:16:08	54	4	ShoulderLeft	-0.1037694	0.1405163	2.792696
6:16:08	54	4	ElbowLeft	-0.2164899	-0.0815465	2.755363
6:16:08	54	4	WristLeft	-0.3423347	-0.2859718	2.674932

6:16:08	54	4	HandLeft	-0.4394781	-0.4457697	2.640919
6:16:08	54	4	ShoulderRight	0.2491306	0.1425115	2.84924
6:16:08	54	4	ElbowRight	0.3247877	-0.0949460	2.829608
6:16:08	54	4	WristRight	0.4464777	-0.3118598	2.799425
6:16:08	54	4	HandRight	0.5711275	-0.5212782	2.79853
6:16:08	54	4	HipLeft	-0.018705	-0.2374434	2.717973
6:16:08	54	4	KneeLeft	-0.110186	-0.5794815	2.658908
6:16:08	54	4	AnkleLeft	-0.105069	-0.8564035	2.591405
6:16:08	54	4	FootLeft	-0.1383674	-0.7525697	2.603314
6:16:08	54	4	HipRight	0.1401658	-0.2364716	2.747092
6:16:08	54	4	KneeRight	0.1786352	-0.5844484	2.649752
6:16:08	54	4	AnkleRight	0.2345826	-0.8585159	2.609676
6:16:08	54	4	FootRight	0.1733328	-0.8021312	2.610182

Figure 19. Sample Joint output (2 frames) from the Kinesic Real Time Data Capture Software

Time	Millisecond	SkeletonID	JointX	JointY	JointZ
6:16:08	39	4	0.042094	-0.24195	2.662193
6:16:08	54	4	0.046264	-0.2435	2.664862
6:16:08	86	4	-0.15867	-0.27755	2.455804
6:16:08	117	4	-0.14502	-0.27214	2.438134
6:16:08	148	4	-0.12311	-0.2574	2.463184
6:16:08	195	4	-0.09961	-0.25025	2.464329
6:16:08	210	4	0.193052	-0.17576	2.686492
6:16:08	242	4	0.222623	-0.17232	2.689159
6:16:08	273	4	0.235105	-0.16601	2.694462
6:16:08	320	4	0.252239	-0.16588	2.697867
6:16:08	351	4	0.272886	-0.15219	2.703254
6:16:08	382	4	0.28863	-0.149	2.707058
6:16:08	429	4	0.313742	-0.14663	2.711317
6:16:08	460	4	0.340276	-0.14295	2.718948
6:16:08	491	4	0.377089	-0.13771	2.725929

Figure 20. Sample Skeleton Output (15 frames) from the Kinesic Real Time Data Capture Software

The sample code below shows how the above Microsoft Kinect data can be collected for analysis.

Main Program File

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading;
using System.Diagnostics;
using Microsoft.Research.Kinect.Nui;
using System.Windows.Forms;
using System.Xml;

namespace KinectGetData
{
    class Program
    {


        static string jointDataOutputPath = "";
        static string skeletonDataOutputPath = "";

        static Runtime nui = new Runtime();
        static void Main(string[] args)
        {
            //get config settings
            XmlTextReader reader = new
XmlTextReader("KinectGetData.exe.config");
            XmlDocument xmlDoc = new XmlDocument();
            xmlDoc.Load(reader);
            jointDataOutputPath =
xmlDoc.SelectSingleNode("configuration/appSettings/JointDataOutputP
ath/text()").Value;
            skeletonDataOutputPath =
xmlDoc.SelectSingleNode("configuration/appSettings/SkeletonDataOutp
utPath/text()").Value;

            //start recording kinect data
            Thread kinectDataCaptureThread = new Thread(new
ThreadStart(captureJointPoints));
            try
            {


                using (System.IO.StreamWriter file = new
System.IO.StreamWriter(@jointDataOutputPath, true))
                {
                    //write to file: quality measure (joint.W)
```

```
currently does not work properly; so says Microsoft. I left it out.
                    file.WriteLine("Time, Millisecond, SkeletonID, 
JointID, JointX, JointY, JointZ");
                }
                using (System.IO.StreamWriter file = new 
System.IO.StreamWriter(@skeletonDataOutputPath, true))
                {
                    //write to file: quality measure (joint.W) 
currently does not work properly; so says Microsoft. I left it out.
                    file.WriteLine("Time, Millisecond, SkeletonID, 
JointX, JointY, JointZ");
                }
            }
            catch (System.IO.IOException)
            {
                System.Windows.MessageBox.Show("Error opening the 
file for Kinect Data export.\nMake sure the output files are not 
currently in use and restart the program.");
            }
            kinectDataCaptureThread.Start();


            Thread.Sleep(1000 * 180);//3 minutes: change to match 
experimental protocol

            //TODO: shut down kinect data collection; save to file

            kinectDataCaptureThread.Abort();


        }

        static void startApplication(string directory, string 
filename)
        {
            System.IO.Directory.SetCurrentDirectory(directory);
            Process p = new Process();
            p.StartInfo.FileName = filename;
            //p.StartInfo.Arguments = "";
            p.Start();
        }


        public static void captureJointPoints()
        {

            try
            {
                nui.Initialize(RuntimeOptions.UseSkeletalTracking);
            }
            catch (InvalidOperationException)
            {
                System.Windows.MessageBox.Show("Runtime 
```

```
initialization failed. Please make sure Kinect device is plugged
in.");
                return;
            }

            nui.SkeletonFrameReady += new
EventHandler<SkeletonFrameReadyEventArgs>(nui_SkeletonFrameReady);
            Thread.Sleep(1000);//needs time to initialize or it
crashes for some reason. Hokey way to deal with this for now.
        }

        static void nui_SkeletonFrameReady(object sender,
SkeletonFrameReadyEventArgs e)
        {
            SkeletonFrame skeletonFrame = e.SkeletonFrame;

            int iSkeleton = 0;


            foreach (SkeletonData data in skeletonFrame.Skeletons)
            {

                if (SkeletonTrackingState.Tracked ==
data.TrackingState)
                {
                //capture overall center of mass for the skeleton
                using (System.IO.StreamWriter file = new
System.IO.StreamWriter(@skeletonDataOutputPath, true))
                    file.WriteLine(DateTime.Now.ToLongTimeString()
+ ","+ DateTime.Now.Millisecond  + "," + data.TrackingID +
                        "," + data.Position.X + "," +
data.Position.Y + "," + data.Position.Z);

                    // Save joints


                    foreach (Joint joint in data.Joints)
                    {
                        using (System.IO.StreamWriter file = new
System.IO.StreamWriter(@jointDataOutputPath, true))

file.WriteLine(DateTime.Now.ToLongTimeString() + "," +
DateTime.Now.Millisecond + "," + data.TrackingID + ","
                                 + joint.ID + "," +
joint.Position.X + "," + joint.Position.Y + "," + joint.Position.Z
                                 );

                      //  Console.Out.Write(joint.ID + "," +
joint.Position.X + "," + joint.Position.Y + "," +
joint.Position.Z);


                    }
```

```
                }
                iSkeleton++;
            } // for each skeleton


        }
    }
}
```

C# project settings file

```
<?xml version="1.0" encoding="utf-8"?>
<Project ToolsVersion="4.0" DefaultTargets="Build"
xmlns="http://schemas.microsoft.com/developer/msbuild/2003">
  <PropertyGroup>
    <Configuration Condition=" '$(Configuration)' == ''
">Debug</Configuration>
    <Platform Condition=" '$(Platform)' == '' ">x86</Platform>
    <ProductVersion>8.0.30703</ProductVersion>
    <SchemaVersion>2.0</SchemaVersion>
    <ProjectGuid>E10B9616-3C49-4399-81F6-157F524CB0E5</ProjectGuid>
    <OutputType>Exe</OutputType>
    <AppDesignerFolder>Properties</AppDesignerFolder>
    <RootNamespace>KinectGetData</RootNamespace>
    <AssemblyName>KinectGetData</AssemblyName>
    <TargetFrameworkVersion>v4.0</TargetFrameworkVersion>
    <TargetFrameworkProfile>Client</TargetFrameworkProfile>
    <FileAlignment>512</FileAlignment>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' ==
'Debug|x86' ">
    <PlatformTarget>x86</PlatformTarget>
    <DebugSymbols>true</DebugSymbols>
    <DebugType>full</DebugType>
    <Optimize>false</Optimize>
    <OutputPath>bin\Debug\</OutputPath>
    <DefineConstants>DEBUG;TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' ==
'Release|x86' ">
    <PlatformTarget>x86</PlatformTarget>
    <DebugType>pdbonly</DebugType>
    <Optimize>true</Optimize>
    <OutputPath>bin\Release\</OutputPath>
    <DefineConstants>TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
```

```
  <ItemGroup>
    <Reference Include="Microsoft.Research.Kinect, Version=1.0.0.0, Culture=neutral, PublicKeyToken=31bf3856ad364e35, processorArchitecture=MSIL" />
    <Reference Include="PresentationFramework" />
    <Reference Include="System" />
    <Reference Include="System.Core" />
    <Reference Include="System.Windows.Forms" />
    <Reference Include="System.Xml.Linq" />
    <Reference Include="System.Data.DataSetExtensions" />
    <Reference Include="Microsoft.CSharp" />
    <Reference Include="System.Data" />
    <Reference Include="System.Xml" />
  </ItemGroup>
  <ItemGroup>
    <Compile Include="Program.cs" />
    <Compile Include="Properties\AssemblyInfo.cs" />
  </ItemGroup>
  <ItemGroup>
    <None Include="KinectGetData.exe.config" />
  </ItemGroup>
  <Import Project="$(MSBuildToolsPath)\Microsoft.CSharp.targets" />
  <!-- To modify your build process, add your task inside one of the targets below and uncomment it.
       Other similar extension points exist, see Microsoft.Common.targets.
  <Target Name="BeforeBuild">
  </Target>
  <Target Name="AfterBuild">
  </Target>
  -->
</Project>
```

Configuration File

```
<?xml version="1.0" encoding="utf-8" ?>
<configuration>

  <appSettings>

<JointDataOutputPath>C:\ScreeningExperiment2\KinectData\JointData.csv</JointDataOutputPath>

<SkeletonDataOutputPath>C:\ScreeningExperiment2\KinectData\SkeletonData.csv</SkeletonDataOutputPath>
  </appSettings>

</configuration>
```

12. APPENDIX C: EXPERIMENT 2 FACES CIT ANALYSIS

Four of the five visual CIT questions for experiment two presented words as stimuli. As an exploratory measure a fifth CIT foil was added. This final foil was more visual in nature: four faces were presented on the screen, one of which would have been significant only to a participant in the Guilty condition. There was no individual baseline to compare with, but between subjects effects were determined for each day using logistic regression models that specified the presence of hidden guilty knowledge as the response variable. The logistic regression models for day 1 and day 2 both produce a Nagelkerke R^2 of .16 and a Cox & Snell R^2 of .12. Time spent gazing at the center of the screen was significant on day 1 but not day 2. The direction of the initial saccade was not significant. Response time was consistently predictive across both days. The results of these models are shown in the table below.

Table 11. Results of Logistic Regression Models for Faces Visual CIT Foil

Fixed Effects	**Day 1 Model**	**Day 2 Model**
	β (S. E.)	**β (S. E.)**
Intercept	2.580** (.907)	2.937** (.901)
Time Viewing Center of Screen	**-.076**** (.028)	-.027 (.018)
Initial Saccade Toward Center of Screen	.028 (.445)	-.003 (.499)
Response Time	**-.638*** (.294)	**-.914**** (.325)

Notes: ** $p < .01$; * $p < .05$

The model for day 1 produced an AUC of .70; the model for day produced an AUC of .72. ROC Curves for these models are displayed below.

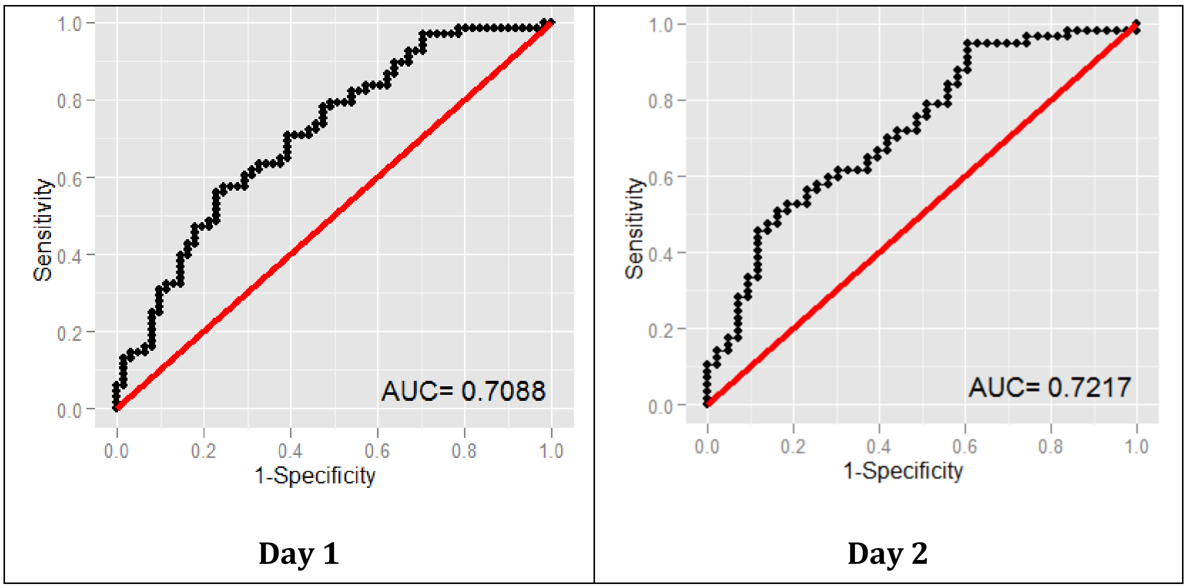

Figure 21. ROC Curves for the CIT foil that used faces.

REFERENCES

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). Separating deceptive and orienting components in a Concealed Information Test. *International Journal of Psychophysiology, 70*(2), 95-104.

Bandyk, M. (2010). What Airport Security Costs You. *U.S. News,* (20 May 2011). Retrieved from http://money.usnews.com/money/business-economy/articles/2010/01/11/what-airport-security-costs-you?PageNr=1

Basuchoudhary, A., & Razzolini, L. (2006). Hiding in plain sight - using signals to detect terrorists. *Public Choice, 128*(1-2), 245-255.

Ben-Shakhar, G. (2002). A Critical Review of the Control Question Test (CQT). In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 103-126). San Diego, CA, USA: Academic Press.

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

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., & Kugelmas.S. (1970). Guilty Knowledge Technique - Application of Signal Detection Measures. [Article]. *Journal of Applied Psychology, 54*(5), 409-413.

Bernstein, A. S. (1979). The orienting response as a novelty and significance detector: Reply to O'Gorman. *Psychophysiology, 16*(3), 263-273.

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

Bracha, H. S., Ralston, T. C., Matsukawa, J. M., Williams, A. E., & Bracha, A. S. (2004). Does "Fight or Flight" Need Updating? *Psychosomatics, 45*, 448-449.

Bradley, M. T., MacLaren, V. V., & Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. [Article]. *Journal of Applied Psychology, 81*(2), 153-160.

Buller, D. B., & Aune, R. K. (1987). Nonverbal cues to deception among intimates, friends, and strangers. *Journal of Nonverbal Behavior, 11*(4), 269-290.

Bureau of Transportation Statistics. (2008). Border Crossing/Entry Data. Retrieved August 24, 2011, from http://www.bts.gov/programs/international/transborder/TBDR_BC/TBDR_BCQ.html

Burgoon, J., Nunamaker, J. F., & Metaxas, D. (2010). *Noninvasive measurement of multimodal indicators of deception and credibility* (Final Report to the Defense Academy for Credibility Assessment (Grant No. IIP-0701519)). Tucson, AZ: University of Arizona.

Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal Deception 5: Accuracy in Deception Detection. *Communication Monographs, 61*(4), 303-325.

Burgoon, J. K., Jensen, M., Twyman, N. W., Meservy, T. O., Metaxas, D. N., Michael, N., et al. (2010). *Automated Kinesic Analysis for Deception Detection*. Paper presented at the HICSS-43 Symposium on Credibility Assessment and Information Quality in Government and Business.

Cannon, W. B. (1914). The emergency function of the adrenal medulla in pain and the major emotions. *American Journal of Physiology, 33*(2), 356-372.

Cannon, W. B. (1929). *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Research Into the Function of Emotional Excitement* (2nd ed.). New York, NY, USA: 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: The external validity of mock crime studies. [Article]. *Journal of Experimental Psychology-Applied, 9*(4), 261-269.

Cavusoglu, H., Koh, B., & Raghunathan, S. (2010). An Analysis of the Impact of Passenger Profiling for Transportation Security. [Article]. *Operations Research, 58*(5), 1287-1302.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. [Article]. *Decision Support Systems, 50*(1), 164-175.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustic Society of America, 25*(5), 975-979.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding, 61*(1), 38-59.

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. [Review]. *MIS Quarterly, 33*(3), 539-566.

Department of Homeland Security. (2009). *Department of Homeland Security Annual Performance Report for Fiscal Years 2008--2010* Retrieved from http://www.dhs.gov/xabout/budget/editorial_0430.shtm.

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), 177-201.

Derrick, D. C., Elkins, A. C., Burgoon, J. K., Nunamaker, J. F., & Zeng, D. D. (2010). Border Security Credibility Assessments via Heterogeneous Sensor Fusion. *IEEE Intelligent Systems, 25*(3), 41-49.

Derrick, D. C., Moffitt, K., & Nunamaker, J. F. (2011, Jan 4-7). *Eye Gaze Behavior as a Guilty Knowledge Test: Initial Exploration for Use in Automated, Kiosk-based Screening.* Paper presented at the HICSS-44 Symposium on Credibility Assessment and Information Quality in Government and Business, 44th Annual Hawaii International Conference on System Sciences, Koloa, HI, USA.

Djamasbi, S., Siegel, M., Skorinko, J., & Tullis, T. (2011). Online Viewing and Aesthetic Preferences of Generation Y and the Baby Boom Generation: Testing User Web Site Experience Through Eye Tracking. [Article]. *International Journal of Electronic Commerce, 15*(4), 121-157.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd ed.). New York, NY, USA: Springer-Verlag.

Elaad, E. (1997). Polygraph examiner awareness of crime-relevant information and the guilty knowledge test. *Law and Human Behavior, 21*(1), 107-120.

Elaad, E., & Ben-Shakhar, G. (2009). Countering Countermeasures in the Concealed Information Test Using Covert Respiration Measures. *Applied Psychophysiology and Biofeedback, 34*(3), 197-208.

Fainberg, A. (1992). Explosives Detection for Aviation Security. *Science, 255*(5051), 1531-1537.

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.

Fukuda, K. (2001). Eye blinks: New indices for the detection of deception. *International Journal of Psychophysiology, 40*(3), 239-245.

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

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), 1287-1301.

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

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

Gati, I., & Ben-Shakhar, G. (1990). Novelty and Significance in Orientation and Habituation - a Feature-Matching Approach. [Article]. *Journal of Experimental Psychology-General, 119*(3), 251-263.

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems, 50*(3), 595-601.

Gray, J. A. (1988). *The Psychology of Fear and Stress* (2nd ed.). Cambridge, UK: Cambridge University Press.

Graybow, M. (2009, August 11, 2011). Madoff mysteries remain as he nears guilty plea. Retrieved 9 August 2011, from http://www.reuters.com/article/2009/03/11/us-madoff-idUSTRE52A5JK20090311?pageNumber=2&virtualBrandChannel=0&sp=true

Green, I., Raz, T., & Zviran, M. (2007). Analysis of active intrusion prevention data for predicting hostile activity in computer networks - Using a generic and reliable model to anticipate future attack scenarios. *Communications of the ACM, 50*(4), 63-68.

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.

Hahm, J., Ji, H. K., Jeong, J. Y., Oh, D. H., Kim, S. H., Sim, K. B., et al. (2009). Detection of Concealed Information: Combining a Virtual Mock Crime with a P300-based Guilty Knowledge Test. [Article]. *Cyberpsychology & Behavior, 12*(3), 269-275.

Hansen, J. V., Lowry, P. B., Meservy, R., & McDonald, D. (2007). Genetic Programming for Prevention of Cyberterrorism through dynamic and evolving intrusion detection. *Decision Support Systems, 43*(4), 1362-1374.

Hartwig, M., Granhag, P. A., Stromwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior, 30*(5), 603-619.

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.

Hollingworth, A. (2007). Object-Position Binding in Visual Memory for Natural Scenes and Object Arrays. *Journal of Experimental Psychology, 33*(1), 31-47.

Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems, 46*(4), 853-864.

Honts, C. R. (2004). The psychophysiological detection of deception. In P. A. Granhag & L. A. Stromwall (Eds.), *Deception detection in forensic contexts* (pp. 103-123). Cambridge, UK: Cambridge University Press.

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 & Management, 10*(1), 101-118.

Horvath, F., Jayne, B. P., & Buckley, J. P. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal of Forensic Sciences, 39*(3), 793-807.

Hsu, I. (2012). Response Time Analysis in an Automated CIT [working title]. University of Arizona.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585-594.

Iacono, W. G. (2000). The detection of deception. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd ed., pp. 772-793). Cambridge, UK: Cambridge University Press.

Iacono, W. G. (2008). Effective Policing--Understanding how Polygraph Tests Work and are Used. *Criminal Justice and Behavior, 35*(10), 1295-1308.

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

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal Interrogation and Confessions* (4th ed.). Gaithersburg, MD, USA: Aspen.

Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Nunamaker, J. F. (2010). Technology dominance in complex decision making: The case of aided credibility assessment. *Journal of Management Information Systems, 27*(1), 181-207.

Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Nunamaker Jr., J. F. (2011). Effects of Automated and Participative Decision Support in Computer-Aided Credibility Assessment. *Journal of Management Information Systems, 28*(1), 201-233.

John E. Reid & Associates. (2011). Interviewing & Interrogation. Retrieved 9 August, 2011, from http://reid.com/training_programs/interview_overview.html

Kanaujia, A., Huang, Y., & Metaxas, D. (2006). Tracking Facial Features Using Mixture of Point Distribution Models. *Computer Vision, Graphics and Image Processing, 4338*, 492-503.

Kim, H., Shin, K. G., & Pillai, P. (2011). MODELZ: Monitoring, Detection, and Analysis of Energy-Greedy Anomalies in Mobile Handsets. *IEEE Transactions on Mobile Computing, 10*(7), 968-981.

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., et al. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *Neuroimage, 15*(3), 727-732.

Lee, A. J., & Jacobson, S. H. (2011). The impact of aviation checkpoint queues on optimizing security screening effectiveness. *Reliability Engineering & System Safety, 96*(8), 900-911.

Levine, T. R., Shaw, A., & Shulman, H. C. (2010). Increasing Deception Detection Accuracy with Strategic Questioning. *Human Communication Research, 36*(2), 216-231.

Lorigo, L., Haridasan, M., Brynjarsdottir, H., Xia, L., Joachims, T., Gay, G., et al. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology, 59*(7), 1041-1052.

Lykken, D. T. (1959). The GSR in the Detection of Guilt. *Journal of Applied Psychology, 43*(6), 385-388.

Lykken, D. T. (1974). Psychology and Lie Detector Industry. [Article]. *American Psychologist, 29*(10), 725-739.

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York, NY, USA: Plenum Trade.

MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology, 86*(4), 674-683.
Maltzman, I. (1979). Orienting reflexes and significance: A reply to O'Gorman. *Psychophysiology, 16*(3), 274-281.
Meijer, E. H., & Verschuere, B. (2010). The Polygraph and the Detection of Deception. [Article]. *Journal of Forensic Psychology Practice, 10*(4), 325-338.
Menneer, T., Cave, K. R., & Donnelly, N. (2009). The Cost of Search for Multiple Targets: Effects of Practice and Target Similarity. *Journal of Experimental Psychology-Applied, 15*(2), 125-139.
Meservy, T. O., Jensen, M. L., Kruse, J., Burgoon, J. K., & Nunamaker, J. F., Jr. (2005, May 19-20). *Automatic extraction of deceptive behavioral cues from video.* Paper presented at the The IEEE International Conference on Intelligence and Security Informatics (ISI 2005), Atlanta, GA.
Meservy, T. O., Jensen, M. L., Kruse, J., Burgoon, J. K., Nunamaker, J. F., Jr., Twitchell, D. P., et al. (2005). Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems, 20*(5), 36-43.
Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers, 36*(4), 630-633.
Moffitt, K. C. (2011). *Toward Enhancing Automated Credibility Assessment: A Model for Question Type Classification and Tools for Linguistic Analysis.* Unpublished Dissertation, University of Arizona, Tucson, AZ, USA.
Morales, J. A. (2008). A Behavior Based Approach to Virus Detection. *FIU Electronic Theses and Dissertations, Paper 41*(http://digitalcommons.fiu.edu/etd/41).
Mosk, M., Hill, A., & Fleming, T. (2010). Gaping Holes in Airline Security: Loaded Gun Slips Past TSA Screeners. Retrieved from http://abcnews.go.com/Blotter/loaded-gun-slips-past-tsa-screeners/story?id=12412458
Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 315-330.
Münsterberg, H. (1908). *On the witness stand*. New York: The McClure Company.
Nakanishi, Y. (2008). Aviation Security Management. *Journal of Transportation Security, 1*(4), 257-265.
Nakayama, M. (Ed.). (2002). *Practical Use of the Concealed Information Test for Criminal Investigation in Japan*. San Diego, CA, USA: Academic Press.
National Research Council. (2003). *The Polygraph and lie detection*.
Nunamaker, 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.
O'Kane, P., Sezer, S., & McLaughlin, K. (2011). Obfuscation: The Hidden Malware. *IEEE Security & Privacy, 9*(5), 41-47.

Osher, D. B. (2007). *Multimethod assessment of deception: Oculomotor movement, pupil size, and response time measures.*, University of Utah, Salt Lake City, Utah, USA.
Park, G., & DeShon, R. P. (2010). A Multilevel Model of Minority Opinion Expression and Team Decision-Making Effectiveness. *Journal of Applied Psychology, 95*(5), 824-833.
Park, S.-W., Yuk, S., Ryu, J. S., Kim, M. S., & Yi, Y. (2006). Development of the body scan system with backscattered X-rays. *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 568*(1), 369-374.
Pavlov, I. P. (1927). *Condition Reflex*. Oxford, England: Clarendon Press.
Podlesny, J. A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest, 20*, 57-61.
Podlesny, J. A. (2003). A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations. *Forensic Science Communications, 5*(3).
Posner, M. I. (1980). Orienting of Attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3-25.
Raskin, D. C., & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 1-47). San Diego, CA, USA: Academic Press.
Ryu, Y. U., & Rhee, H. S. (2008). Improving intrusion prevention models: Dual-threshold and dual-filter approaches. [Article]. *Informs Journal on Computing, 20*(3), 356-367.
Seymour, T. L., & Fraynt, B. R. (2009). Time and Encoding Effects in the Concealed Knowledge Test. [Article]. *Applied Psychophysiology and Biofeedback, 34*(3), 177-187.
Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". [Article]. *Journal of Applied Psychology, 85*(1), 30-37.
Sokolov, E. N. (1963). Higher Nervous Functions - Orienting Reflex. [Article]. *Annual Review of Physiology, 25*(1), 545-580.
Steptoe, W., Steed, A., Rovira, A., & Rae, J. (2010). *Lie tracking: social presence, truth and deception in avatar-mediated telecommunication*. Paper presented at the Proceedings of the 28th international conference on Human factors in computing systems.
Twyman, N. W., Elkins, A., & Burgoon, J. K. (2011, January 4-7). *A Rigidity Detection System for the Guilty Knowledge Test.* Paper presented at the HICSS-44 Symposium on Credibility Assessment and Information Quality in Government and Business, Koloa, HI, USA.
Twyman, N. W., Jenkins, J. L., Carl, K., & Nunamaker Jr., J. F. (2011). *Knowledge Sourcing and Knowledge Consumption in Computer-Mediated Complex*

Decision Making. Paper presented at the 44th Annual Hawaii International Conference on System Sciences (HICSS).

Twyman, N. W., Moffitt, K., Burgoon, J. K., & Marchak, F. (2010, January 5-8). *Using Eye Tracking Technology as a Concealed Information Test* Paper presented at the HICSS-43 Symposium on Credibility Assessment and Information Quality in Government and Business, 43rd Annual Hawaii International Conference on System Sciences, Koloa, HI, USA.

USA vs. Umar Farouk Abdulmutallab, 2:10-cr-20005-NGE-DAS (United States District Court Eastern District of Michigan Southern Division 2010)

Vassiliades, V., Evans, P., Kaufman, L., Chan, J. W., & Downes, D. (2008). Chapter 20: A Human-Centered Approach for Improving Airport Security Screening. In P. D. Bust (Ed.), *Contemporary Ergonomics 2008: Proceedings of the International Conference on Contemporary Ergonomics (CE2008), 1-3 April 2008, Nottingham, UK* (pp. 127–132): Taylor & Francis.

Verschuere, B., Crombez, G., Clercq, A. D., & Koster, E. H. W. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology, 41*(3), 461-466.

Viola, P., & Jones, M. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision, 57*(2), 137-154.

Vrij, A. (1995). Behavioral Correlates of Deception in a Simulated Police Interview. *Journal of Psychology, 129*(1), 15-28.

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

Vrij, A., Mann, S., & Fisher, R. P. (2006). An Empirical Evaluation of the Behaviour Analysis Interview. *Law and Human Behavior, 30*(3), 329-345.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. [Article]. *Law and Human Behavior, 32*(3), 253-265.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research, 22*(4), 544-562.

Wang, X. F. W. X. F., & Zhuang, J. (2011). Balancing congestion and security in the presence of strategic applicants with private information. *European Journal of Operational Research, 212*(1), 100-111.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., Jr., & Twitchell, D. P. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation, 13*(1), 81-106.