A Systems Approach to Countermeasures in Credibility Assessment Interviews

Completed Research Paper

Nathan W. Twyman
University of Arizona
Tucson, AZ, USA
nathan.twyman@arizona.edu

Ryan Schuetzler
University of Arizona
Tucson, AZ, USA
rschuetzler@cmi.arizona.edu

Jeffrey Gainer Proudfoot
University of Arizona
Tucson, AZ, USA
jproudfoot@cmi.arizona.edu

Aaron C. Elkins
Imperial College, London
London, UK
a.elkins@imperial.ac.uk

Abstract

Countermeasures, or techniques for hiding guilt during a credibility assessment examination, have long been an important topic in cognitive psychology and criminal justice fields. With recent IS research on automated screening systems, understanding the potential for countermeasures in this new paradigm is of increasing importance. This paper reports on a large experiment examining countermeasures in an automated deception detection screening context. The effectiveness of traditional countermeasure types (mental and physical) are examined, as well as an exploratory approach of trying several countermeasures at once. The exploratory approach was tested to investigate a proposed novel systems-inspired solution to countermeasures—triangulating on deception likelihood using multiple sensors measuring multiple behavioral and psychophysiological anomalies. The findings give credence to the proposition that monitoring multiple heterogeneous cues to deception may be a viable solution for mitigating the effectiveness of countermeasures.

Keywords: Countermeasures Mitigation Systems, Deception Detection Systems, Automated Screening Systems, Automated Screening Kiosk

Introduction

For almost as long there have been systematic methods for detecting deception, there have been tools for mitigating their effectiveness. This is especially true for the polygraph, which is the most widely known and widely used technology designed to augment human interviewers in credibility assessment (Honts and Perry 1992). Research suggests that countermeasures can be effectively used against the traditional polygraph sensors for both the Control Question Technique (CQT) as well as the Concealed Information Test (CIT) method of interviewing (Ben-Shakhar and Dolev 1996; Elaad and Ben-Shakhar 1991; Honts and Kircher 1994; Honts et al. 1987).

The effectiveness of countermeasures have been of particular interest to the cognitive psychology and criminal justice fields, and now is of increasing interest to IS research investigating systems solutions to integrity screening and credibility assessment systems. New system designs are being evaluated and validated for the purpose of conducting more rapid, non-contact credibility assessment interviews, and to gain more insight into the underlying mechanics of credibility assessment (Derrick et al. 2011; Nunamaker et al. 2011; Twyman 2012; Twyman et al. 2011b). For example, eye tracking and vocalic measures of deception and concealed information have been a focus of recent automated interviewing research (Elkins et al. 2012b; Nunamaker Jr et al. 2012; Proudfoot et al. 2012; Proudfoot et al. 2013; Twyman et al. 2013), and have provided new insights into psychophysiological and behavioral correlates of deception. As non-contact, automated credibility assessments emerge as an enticing option for managing security and integrity, it is critical to identify and investigate possible countermeasures that may be employed.

This paper presents the findings of a study designed to evaluate the effectiveness of a variety of countermeasures against a system interaction designed to identify deception and concealed information. The remainder of the paper is organized as follows: (1) we discuss relevant deception detection and countermeasures literature, (2) we outline our research approach, which includes a high-level system design and laboratory experiment used to test countermeasures against an array of sensors, (3) we present the results of the experiment, and (4) we review the contributions, limitations, and conclusions of this work.

Literature Review

Technology has been used in credibility assessment interviews since at least 1895, when Cesare Lombroso, an Italian criminologist, used a medical device for measuring blood pressure changes during police interrogations (Trovillo 1939). In the 1920s and 30s, John Larson and Leonarde Keeler developed the now-widely-known polygraph machine which measures blood pressure, respiration, and skin conductance (a measure of arousal) (Alder 1998). Polygraph systems have since become the standard tool and decision support system for credibility assessment interviews. The polygraph and its associated control question technique of interviewing have not undergone major updates for many decades. Limitations to this current paradigm limit the potential application and usefulness of human credibility assessment. Research in several fields is seeking to update and improve upon existing knowledge and practices in this area of research, which has potential for high impact. Because of a general scientific consensus that the most common polygraph technique lacks validity (Iacono and Lykken 1997; National Research Council 2003), much academic research on credibility assessment focuses on alternative techniques such as the CIT that feature more control (Lykken 1998; MacLaren 2001; Rosenfeld and Labkovsky 2010b; Twyman et al. 2011b; Twyman et al. 2010; Vrij 2008).

Technology can clearly improve deception detection accuracy. Unaided human deception detection accuracy rates hover near chance levels (Bond and DePaulo 2006). Decision support or fully automated credibility assessment are an important area where IS research can make a substantial impact. Recently, additional technologies for human screening have been investigated, including non-contact technologies for measuring heart rate and blood pressure (Nunamaker et al. 2011), vocalic features (Elkins and Burgoon 2010; Gamer et al. 2006; Harnsberger et al. 2009), linguistic variables (Fuller et al. 2009; Vizer et al. 2009; Zhou and Zhang 2008), oculometric factors (Fukuda 2001; Osher 2007; Twyman 2012), thermal features (Pavlidis et al. 2002; Pavlidis and Levine 2002), and kinesic factors (Meservy et al. 2005; Twyman et al. 2011b). Beyond individual technologies, automated system designs are now beginning to be proposed, designs that cover the interview protocol, technologies, and interaction facilitation (Derrick et al. 2011; Nunamaker et al. 2011; Twyman et al. 2011a).

Just as no veracity cue has proven to be a "pinnochio's nose," or guaranteed sign of deception, no single technology seeking to measure a physiological or behavioral process will be foolproof. It is thus critical that research in this area be contextualized, clearly explaining boundaries and limitations for a given design or approach. For the current study, we focus on fully automated screening system interviews, and specifically automated systems employing a CIT-based interview.

The human-computer interaction (HCI) with automated screening systems is of particular importance, because the system must comprehend and adapt to signals from the human that are not traditionally monitored, such as body language or vocal patterns (Derrick et al., 2011). The best human signals to monitor need to be identified. Because of the lack of a single clear signal indicating deception, many signals will need to be monitored simultaneously. Perhaps most importantly, the interaction itself must be specially designed to elicit valid signals (Twyman et al., 2011a).

An important aim of this research is to gain a greater understanding of the robustness of the system interaction design when users attempt to manipulate the system in their favor. Unlike traditional human-computer interactions where the system supports the user or the system and user complement one another (Nass, Fogg, and Moon, 1996), the HCI of an automated screening system can at times consist of the system and human trying to reach opposite goals. In such cases, the way the individual interacts with the system will be strongly affected by their knowledge of the system's monitoring and their own *countermeasures*, or tactics employed to counter the system's screening effectiveness. Understanding such tactics and their effectiveness will provide insight into future iterations of the system and interaction design. For instance, a recent conceptualization of as HCI framework for automated screening systems mimics human interpersonal interactions, considering humans will interact with such systems as they would another person (Derrick et al., 2011), reflecting the Computers-are-Social-Actors (CASA) paradigm (Sundar and Nass, 2000; Reeves & Nass, 1996). However, when an individual is highly motivated to counter or subvert a system, he or she will more strongly consider the source of the interview, and therefore may not view the system as a social actor (Liang et al, 2013), but rather as a tool to be manipulated. The tactics taken by a user, then, would produce different observable behaviors than a manipulation seen in an interpersonal communication, especially if the manipulator understands how the automated screening system works.

Automated screening systems are a novel solution to novel contexts, and as such, much research is needed in this area. To the extent countermeasures are found to be effective, the system and interaction design will need to be specified for the detection of, adaptation to, or minimization of the effectiveness of countermeasures. In the remainder of this section we review deception theory generally and CIT-based screening systems in particular, then focus on countermeasures and how they relate to several promising non-contact technologies used in recent automated screening systems research.

Deception Theory

Deception occurs in many forms. For example, it can take the form of exaggeration, white lies, equivocation, complete fabrications, and impostership. We define deception as an intentional transmission of a message intended to foster false beliefs or perceptions in the recipient (Knapp & Comadena, 1979). This definition of deception emphasizes the involvement of both a sender and receiver during a deceptive interaction. During this interaction, deception theories explain and predict verbal, nonverbal, and physiological behavior differences between truthtellers and deceivers.

The earliest investigations into deception, from which modern theories have stemmed, were conducted by Ekman and Friesen (1969). They introduced their leakage hypothesis that predicted liars when experiencing arousal, negative affect, and stress would leak unintentional cues that would manifest particularly in the hands, legs, and feet. While experiencing these negative feelings, liars relieve their tension and discomfort through adaptors (i.e., foot tapping, touching of face). In addition to additional arousal induced behaviors, the leakage hypothesis predicts that liars will also inhibit certain behaviors and natural gesturing.

Later, Zuckerman, DePaulo, and Rosenthal (1981) extended this hypothesis with their Four-Factor theory to more directly explain the causes for the behavior differences between liars and truthtellers. This theory predicts leakage behavior is caused by 1) arousal, 2) negative affect, 3) cognitive effort, and 4) behavioral control. Liars not only are predicted to exhibit leakage cues or adaptors due to arousal and negative affect, but also experience more cognitive overload while managing their lie and appearance. This increased vigilance in appearance also causes over-control of normally automatic and natural interaction gesturing, causing rigid and inhibited behavior. If a liar does not experience fear or arousal, then cognitive or behavioral control cues would be more diagnostic.

During communication, people attempt to manage and create an impression of themselves (Jones & Pittman 1982). This self-presentational perspective was applied to deception theory by DePaulo (1992) to elaborate on how behavioral control affects deceivers. Liars attempt to present themselves as truthful and honest through their verbal and nonverbal behavior. Because they are consciously modifying their behavior, it appears unnatural or artificially inhibits gesturing. This perspective predicts that a liar's behavior will be moderated by their communication ability, motivation, and confidence.

Buller and Burgoon introduced Interpersonal Deception Theory (1996) to include the entire deceptive interaction. This theory emphasizes that deception is strategic and involves a complex interaction between at least two people. Depending on the skill of the liar and their relationship in the interaction (e.g., boss, parent, loved one), they are predicted to employ different strategies for managing a credible presentation. Time and the interplay between sender and receiver are important additions to this theory. A liar may start out feeling confident and reveal few behavioral cues to deception, but after sensing suspicion, begin over compensating and exhibiting unnatural gestures. As the conversation continues, they may have to think harder to keep their lie credible and consistent, leading to increased cognitive effort induced behavior.

CIT-based Automated Interviewing Systems

Traditional interviewing methods designed to identify deception often require a time consuming process during which preliminary interviews are conducted and interviewees are directly queried concerning their involvement in a given crime or incident of interest. In some cases, homogeneous physiological measures are recorded concurrently, providing the interviewer with data to scrutinize when making a credibility assessment. The interviewer then utilizes a combination of both objective data and subjective observations to render a decision. This process may yield inaccurate veracity judgments as data may be misinterpreted or subjective judgments may overrule objective data.

To remedy this approach, an alternative interviewing method was proposed several decades ago, which has since been heavily researched and validated in the scientific community. This method is the Concealed Information Test (CIT), also known as the Guilty Knowledge Test (GKT) (Lykken 1959; Lykken 1960). The CIT was developed with the purpose of identifying a person's knowledge of a given event or piece of information by judging whether he or she perceives event-relevant information to be personally significant. For example, a suspect accused of robbing a bank could be tested on his or her familiarity with the type of door on the bank vault or the amount of money that was stolen. Because this approach produces a strong individual baseline of comparison it has a significantly reduced rate of false positives.

During a CIT interview, an individual is presented with a question or statement together with groups of stimuli referred to as *foils*. Foils are comprised of two types of stimuli, target items and nontarget items. A *target item* is a stimulus that is related to the crime or information of interest, one that should elicit an atypical physiological or behavioral response from someone who recognizes it. *Nontarget items* are mixed in with the target items in each foil to serve as controls; they should not elicit any differences from truthful or deceptive individuals. The strength of the CIT is found in administering a test comprised of several foils as the chances are low of a truthful individual randomly exhibiting indications of concealed knowledge on several target items. Another strength is the more objective nature of both administering and interpreting results from the interview.

While commonly researched from the perspective of criminal justice or applied psychology, IS researchers are also leveraging the simplicity and control of the CIT for conducting automated credibility assessments. For example, CITs are now being evaluated in security screening scenarios (Proudfoot et al. 2012; Twyman

2012; Twyman et al. 2011a) and for use in survey tools while keystrokes are recorded and analyzed to identify deception and concealed information (Valacich et al. 2013).

Countermeasures

Because psychophysiological deception detection is based on detecting a physiological response to a stimulus, the employment of measures used to appear innocent is a threat to the validity of any test. Countermeasures have been shown to be effective in many deception detection tests, including the polygraph (e.g., Ben-Shakhar and Dolev 1996; Honts et al. 1996; Honts et al. 1985; Honts et al. 1987), P300-based neural activity detection tests (e.g., Rosenfeld et al. 2004; Sokolovsky et al. 2011), and fMRI-based tests (Ganis et al. 2011).

In the case of the CQT and the CIT, the psychophysiological detection of deception is based on responses differing from irrelevant to relevant items. In the case of the CQT, the expected difference in response is between control questions and relevant questions. For the CIT, the difference is between nontarget and target items. In either case, the response differing in the expected direction indicates the presence of deception.

Countermeasures function primarily by manipulating the participant's response to the irrelevant items to minimize the difference when the relevant item or question is asked. Countermeasures fall primarily into two categories: mental and physical. Physical countermeasures include finger movements (Ganis et al. 2011), pressing toes against the floor (Honts et al. 1996; Honts et al. 1985; Honts et al. 1987), or biting the tongue (Honts et al. 1985; Honts et al. 1987). These physical countermeasures are employed during a portion of the irrelevant items in either a CQT or CIT. They work by creating a similar physiological response to irrelevant items to what the guilty show during relevant items. Mental countermeasures include mental counting (Elaad and Ben-Shakhar 1991; Honts et al. 1996), recalling past emotional events (Ben-Shakhar and Dolev 1996), or simply saying your first and last name in your head (Sokolovsky et al. 2011).

Depending on the test, mental countermeasures are either employed during the irrelevant items only or during all items. For example, in a polygraph-based study Elaad and Ben-Shakhar (1991) found that item-specific mental countermeasures (counting to 10 during the relevant stimulus) increased the overall response and made detection easier. On the other hand, counting sheep throughout the length of the test decreased detection rates. In a P300-based deception detection experiment Sokolovsky et al. (2011) showed that participants who silently said their first and last name during two of four irrelevant stimuli were able to modify their responses enough to evade detection.

Where countermeasures are shown to be effective at manipulating deception detection results, cognitive psychology research has turned to detecting countermeasures. Physical countermeasures are especially vulnerable to detection. Honts et al. (Honts et al. 1987) showed that 90% of countermeasures users were identified using an electromyograph used to measure muscle activity in the legs and head. Similarly, Ganis et al. (2011) showed that while slight finger movements reduced detection accuracy in fMRI-based tests, they also increased activation of the motor cortex, the part of the brain responsible for movement. Increases in reaction time allowed countermeasure detection in P300 mental countermeasures (Rosenfeld et al. 2008). The identification of effective countermeasures allows researchers to discover ways to detect them.

This paper proposes a slightly different approach to mitigating countermeasures. In addition to investigating the effectiveness of traditional countermeasures on new, non-contact sensors, we propose increasing the number and heterogeneity of veracity cues tracked and used in decision making. If individuals can be simultaneously attentive to a limited number of activities, then they are less likely to be able to counter a credibility assessment system when the system tracks and measures many heterogeneous human factors.

Pupil Dilation

Changes in pupil dilation can be linked with a number of cognitive functions. Preliminary research discovered that changes in pupild dilation can be used to identify activation and arousal in autonomic activity (Goldwater 1972; Nunnally et al. 1967). Additionally, differences in pupillary responses have been linked with short-term and long-term memory retrieval (Beatty and Kahneman 1966; Gardner et al. 1975; Janisse 1977). A study investigating differences in pupil dilation associated with viewing novel and repeated stimuli revealed that pupils exhibit increased dilation when repeatedly exposed to a given stimulus (Heaver 2011). The authors referred to this as the Pupil Old/New Effect, or PONE. Additional work has replicated this (Maw and Pomplun 2004).

The orienting response is traditionally the gold-standard human factor of interest in a CIT, and tracking electrodermal activity (skin sweatiness) is the standard method for measuring the orienting response. However, the physiological activation triggered by the orienting response is not limited to electrodermal activity alone. Several physiological changes result, among them being pupil dilation. Within CIT research, evidence supports the notion that common mechanism triggers both pupil dilation and electrodermal activity in response to concealed knowledge (Bradley et al. 2008). Additionally, changes in pupil dilation during a CIT are attributable to simply having concealed knowledge and not necessarily to any concurrent deceptive behaviors, such as verbally lying (Janisse and Bradley 1980).

The use of pupil dilation as a cue to deception is dependent on the differential response to the target item. The arousal triggered by recognition of the target item causes the pupil to dilate to a much greater extent than when viewing other items. The employment of mental countermeasures to artificially increase pupillary response to non-target items could help artificially raise the baseline of comparison, resulting in a less accurate determination of deception. Because pupil dilation is related to cognitive processing (Kahneman 1973), taxing mental tasks can increase pupil dilation on demand. By performing mental arithmetic during non-target items, the pupillary response during the target item can thereby be masked. Thus, we predict the following:

- H1: Deceptive individuals will have a larger pupillary response to target items than to non-target items.
- H2: Deceptive individuals using mental countermeasures will exhibit reduced pupil dilation differential between target and non-target items compared to deceptive individuals using no countermeasures.

In addition to cognitive effort, pupil dilation can also be triggered by pain. Using electrical stimulation, Chapman et al. (1999) showed that pupil dilation increases nearly immediately at the onset of pain, and that this dilation increases with increasing pain intensity. Further research has shown that the pupil dilation is not only immediate, but lasts for the duration of the pain (Ellermeier and Westphal 1995). These results indicate that physical countermeasures such as biting the tongue, as employed in polygraph studies, may also work to artificially manipulate the pupil dilation baseline in an automated CIT screening paradigm. As with mental countermeasures, causing pain during non-target items should increase the pupil dilation, reducing the difference between target and non-target pupil dilation in a CIT.

- H3: Physical countermeasures will reduce the pupil dilation differential between target and non-target items.

Kinesic Rigidity

Kinesic rigidity is the constriction of body movement. Rigidity has been found in communication research to be an indicator of low veracity during open-ended or semi-structured interviewing techniques (Caso et al. 2006; Vrij and Mann 2001). When lying, participants tend to exhibit less overall movement, especially expressive or illustrative gestures, and the movement that does occur tends to be spatially constricted and appear forced rather than natural (Buller and Aune 1987; Vrij and Mann 2001; Vrij et al. 1996). Recent IS research has discovered that this phenomenon is also present in the more controlled CIT interview setting, and has developed a method for automatic detection of rigidity via comparison of body movement during baseline items to body movement during target items (Twyman et al. 2011b).

Likely because of the high cost of traditional measurement of rigidity, this cue to deception is not used in practice. It has also received almost no attention in countermeasures research. One psychology study determined that controlling rigidity is very difficult in semi-structured interviews, at least when trying to control it directly (Vrij et al. 1996). However, there are still many unknowns, including the effectiveness of traditional countermeasures, and how well rigidity can be overtly controlled in a highly controlled automated screening setting where many behaviors must be controlled simultaneously.

In semi-structured interviews, rigidity has been hypothesized to stem from cognitive overload, in that more cognitive effort is being placed on mentally constructing and relaying a plausible story, leaving fewer resources to allocate toward nonverbal presentation, creating less overall movement and more constricted movement (DePaulo et al. 1988; Ekman and Friesen 1972). A second theory suggests that rigidity itself may be a form of countermeasure, in that because people generally falsely believe that liars exhibit increased movement, they purposely minimize their own movement to appear truthful (DePaulo and Kirkendol 1989). A third possible explanation is the biologically-driven freeze response that all humans experience when confronted with something that is threatening (Gray 1988). Previous IS research discovered rigidity in a CIT, which requires no communicative or illustrative movement, so cognitive overload will not be a likely driver in the highly controlled, automated format employed by the system design we use. It is possible that the root cause is a combination of the freeze response and behavioral control.

The effectiveness of traditional countermeasures on kinesic rigidity is likely low. Physical countermeasures do not eliminate the threat or the desire to appear truthful. Mental countermeasures should be effective to the extent they not only cognitively distract the examinee, but also help them ignore the potential threat or the desire to appear truthful. However, because individuals have to verbally respond and be visually attentive, mental countermeasures are limited in how much they can distract. Much research is needed to understand these dynamics, but at this early stage it appears unlikely that traditional countermeasures will be particularly effective against the rigidity effect.

- H4: Deceptive individuals will exhibit less overall movement when viewing and responding to a target item.
- H5: Deceptive individuals employing mental or physical countermeasures will exhibit less overall movement when viewing and responding to a target item.

Vocal Pitch

Vocalics refer to the nonverbal elements of speech. To speak, the diaphragm pushes air out from the lungs, through the vibrating vocal folds (Titze and Martin 1998). The frequency of the air affected by the vibration of the vocal folds is perceived as the vocal pitch. The intensity of the air pushed by the diaphragm is perceived as the volume or loudness of the sound. The vocal fold vibrations are facilitated by muscles about the larynx in the vocal tract. Just like other muscles in the body, when an individual experiences stress or arousal the larynx exhibits tension. Tension around the larynx cause the vocal folds to increase the frequency of vibration, thereby increasing the overall variation of vocal pitch.

Increases in mean and range in vocal pitch have been predictive of deceptive speech (DePaulo et al. 2003; Elkins et al. 2012a; Rockwell et al. 1997) and heightened emotions and arousal (Bachorowski and Owren 1995; Scherer et al. 2003). Because vocal pitch provides primarily emotional arousal-based information, it has not traditionally been included in a CIT, which emphasizes orienting responses resulting from cognitive arousal. However, it is likely that emotional arousal may be present in the CIT even though it has not traditionally been measured. Participants may have relatively heightened feelings of fear or anxiety during a CIT.

- H6: Deceptive individuals will exhibit greater variation in voice pitch when responding to a target item.

Mental countermeasures are expected to be effective for the same reasons as stated previously, specifically its distracting qualities. However, physical countermeasures may not cause tension in the larynx, which would negate their potential effectiveness.

- H7: Deceptive individuals using mental countermeasures will exhibit reduced vocal pitch variation differential between target and non-target items compared to deceptive individuals using no countermeasures.

- H8: Deceptive individuals employing physical countermeasures will exhibit greater vocal pitch when viewing and responding to a target item.

It is especially important to determine whether individuals can be successful when they attempt to counter many factors at once. Thus, in addition to testing these eight hypotheses, we also investigate whether countermeasures are less effective when multiple countermeasures are employed at once.

Research Approach

This study followed a design science research (DSR) approach, in which knowledge is discovered through the design, implementation, and evaluation of system prototypes (Gregor and Hevner 2013; Hevner et al. 2004; Nunamaker and Briggs 2011; Nunamaker et al. 1991; Nunamaker et al. 2013). In particular, it adheres to the broad-based, high-impact model of DSR, where multidisciplinary programs of research pursue solutions to major issues through sustained, interrelated projects that build on one another (Nunamaker et al. 2013). Projects iterate through prototyping, experimentation, field studies, and theoretical development in a nonlinear fashion. Persistent, sustained projects use the most expedient research methods in their quest to push solutions from initial ideation to 1) a proof-of-concept stage, where there is sufficient evidence that a concept can work, to 2) a proof-of-value stage, where there is sufficient evidence that a concept does work and adds value to real-world situations, to 3) a proof-of-use stage, where a concept is making a sustained, real-world impact independent of research (Nunamaker and Briggs 2011; Nunamaker et al. 2013). Because of the broad nature of the research approach, contributions stemming from the high-impact DSR framework may reference many disciplines, can affect multiple disciplines, and can have a broad range, including contributions such as practical field observations, new or refined theory, new or refined methods, new or refined design specifications, or a nascent understanding of a new problem space.

The current study represents one part of a larger program of research focused on systems approaches to credibility assessment. In line with the research approach, this program of research heavily pulls from psychology, communication, psychophysiology, cognitive science, and criminal justice in combination with information systems literature.

Because of the relative novelty of the problem space, a laboratory experiment was an appropriate method of investigation at this stage. An experiment was designed to evaluate the ability of deceivers to succesfully bypass the automated screening system through the use of countermeasures. The experimental task was patterned after a number of experiments designed to test the ability of noninvasive sensors to identify deception and concealed information. This was intentional as the findings of previous work could then be compared to the results of this study. The experient was comprised of 4 guilty and 1 control condition. Three of the four guilty treatments entailed learning countermeasures to defeat the system. Measures were repeated within subjects and within question for a total of 20 cases per individual. Detailed information about the data collection process and the experimental task is provided in the following subsections.

Participants

Participants were recruited from undergraduate and graduate business courses at a large southwestern university. While the ideal population would be individuals who regularly participate in illicit activities, such a population was not feasibly obtainable. Students were selected as a target population as they are comprised of individuals from a diverse set of ethnic and cultural backgrounds. Participants (N=175) consented to participate and subsequently completed the experiment. The mean age of participants was 21.86, with a median age of 21, a minimum age of 18, and a maximum age of 36. The sample was comprised of individuals with a diverse range of ethnic backgrounds, including: 33% Asian, 3% Black, 1% Hawaiian/Islander, 46% White, 13% Hispanic, 3% Middle Eastern, and 1% other. Males accounted for 59% of the total participants; 41% were female. Over half (126) of the participants spoke english as a second language. Of the initial 175 participants, 18 were disqualified because they either failed two manipulation check questions or they did not follow instructions (e.g., failed to construct the device). Participants received course credit for participation, as well as $20 if they successfully made it through the checkpoint and completed their task without arousing suspicion. This was a mild deception to encourage personal investment; in reality no participants were required to complete their task beyond the screening

Experiment Task

Each participant was required to provide written consent prior to beginning the experiment. The experimental task mirrored Twyman et al. (2011a), except there were additional conditions. After providing consent, each individual was randomly assigned to one of the following five conditions: guilty with no countermeasures, guilty with mental countermeasures, guilty with physical countermeasures, guilty with all (mental, physical, and additional countermeasures), and an innocent (control) group. The number of participants assigned to each group was roughly equal.

Participants in the guilty conditions were tasked with constructing an improvised explosive device, concealing it in a bag with standard travel items, and delivering the bag to a person in a secure portion of a building. Participants randomly assigned to one of the three countermeasures groups also constructed the improvised explosive device and packed it in a bag; however, they also completed training on how to use countermeasures to defeat the system. Members of the control group packed only innocuous household items in a bag prior to proceeding to the simulated screening; they did not receive any information about the system or countermeasures training. After packing the bag, they proceeded to the simulated screening environment, purportedly to security screening prior to entry.

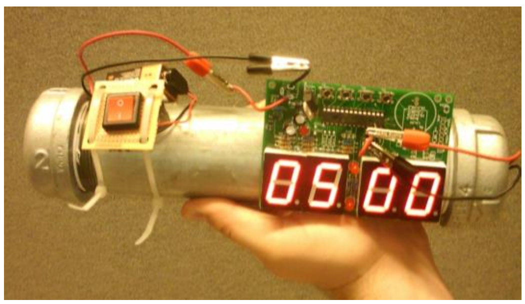

Figure 1. Improvised Explosive Device built by participants in Guilty conditions

Countermeasures training entailed watching a video describing the sensors that would be used during the screening interview to identify deception, as well as methods that could be used to thwart the effectiveness of the sensors. Participants in the mental countermeasures group were instructed to count backwards by seven from a three-digit number. This tactic is designed to occupy the interviewee's mental processes, rendering them cognitively "numb" to the interviewee that is taking place, minimizing any behavioral or physiological differences in behavior that could occur. Participants in the physical countermeasures group were instructed to moderately but carefully bite their tongue when control items were presented during the interview, resulting in physiological responses on control items that could mimic the expected response of a guilty person to an item of interest. This should result in a false negative classification as the guilty participant should appear more concered about control items than items of interest, and would thus appear innocent. The final countermeasures group was trained to use the mental and physical countermeasures described previously, as well as to directly controlt kinesic rigidity (not "freeze up"), control their voice pitch such that it is always the same, keep a neutral facial expression, and to look at all images naturally, not focusing on one item on the screen more than another.

All participants entered the screening environment, placed the packed bag on a counter, and stood in front of the automated screening system. The system was designed to identify the presence of an interviewee and begin the interview autonomously. The first step of the interview entailed calibration of the interviewee to the eye tracking device. During this process, the person must fixate their eye gaze on nine different locations on the screen, resulting in a more accurate and consistent oculometric data collection. After calibration, the screening interview commenced, which entailed displaying 5 slides containing images of banned items to the interviewee. Each slide contained four small images of banned items (Figure 2), and an embodied conversational agent (ECA) asked the interviewee "Are you carrying any of these items?" immediately prior to the display of each screen. The interviewee was given five seconds to respond before the question repeated and the next slide appeared. Participants were asked to vocally respond either yes or no. One of the five slides contained the target item, which was an image that resembled an explosive device. Participants in the guilty conditions had to verbally lie in response to this question to appear truthful. The five-slide (foil) was repeated four times, with the slide order and the spatial location of the images randomly changed each time. In total, each participant responded to 20 questions, four of which required an outright lie.

Figure 2. Example slide used in the automated screening

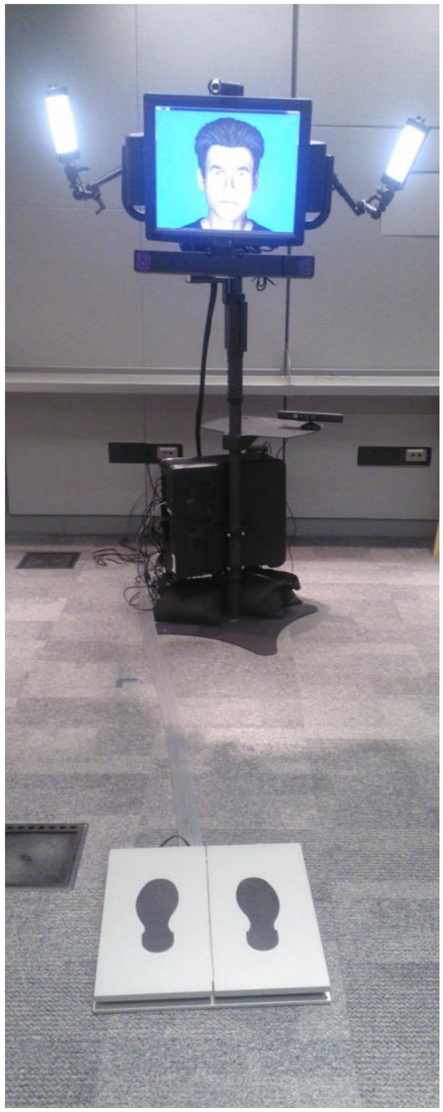

Figure 3. Automated screening kiosk used to question participants

After the screening, each participant completed a brief survey designed to check experimental manipulations and to collect information on their experience during the interview.

Measures

Kinesic data was captured using a Microsoft Kinect sensor, by capturing the 3-dimensional cartesian coordinates of 20 distinct body points in real time at approximately 30 Hz. This raw data was used to calculate overall movement via the same procedure detailed in previous work on automated rigidity detection (Twyman 2012; Twyman et al. 2011b). This raw data was averaged by participant (N=157) and slide (N=20) for each body point. There was one case of lost data due to malfunction, leaving 62,780 data points.

Pupillometry was measured via an EyeTech TM3 eye tracking sensor. The TM3 captured pupil diameter at approximately 30 Hz. The mean pupil diameter for both eyes was calculated for each measurement, then this data was likewise summarized by participant and slide. Technical malfunctions occurred on 2.6% of slides, leaving 3,058 usable data points.

Raw vocal data was captured using an array microphone at 48 kHz. For each response, the maximum, mean, and standard deviation of vocal pitch from the beginning to the end of an utterance of a "no" response, were extracted from the raw vocal data. Technical malfunctions occurred on 5.6% of slides, leaving 2,994 usable data points.

To control for effects stemming from highly variable interpersonal differences such wide variance in nervousness, stillness, eye size, and vocal range, the data points from each of these indicators were standardized using within-subject z-scores (Ben-Shakhar 1985), meaning each subject's observations were representative of a personal baseline as opposed to a population baseline. All observations were also standardized within-foil, to take advantage of the question-specific baseline. In the case of body movement, movement was also standardized for each body point separately to account for natural differences in movement patterns between body points.

Analysis and Results

Because of the relative novelty of vocal measurements in a CIT context, vocal pitch variation underwent a preliminary analysis to explore the three possible measures of variation. Then, separate multilevel regression analyses were performed for each veracity cue.

The vocal measures investigated were: mean pitch, pitch standard deviation, and max pitch. As with body movement and pupil dilation cues, each of these variables were normalized within subject and foil before being submitted to repeated measures ANOVA (Condition X Target Item). The interaction of Condition and Target Item was not significant for mean pitch (p=.28) or pitch standard deviation (p=.065). Max pitch, a measurement of high-end pitch range, was significant for the Condition and Target interaction, F(4, 2989)=2.32, p=.05.

For each target deception indicator of interest, a multilevel regression model was specified with the indicator as the dependent variable. The dependent variables are standardized scores and thus represent standard deviations from an individual's baseline. In each case, the independent variables include Target Item (a binary variable indicating whether the stimuli slide included an IED image), Time (a value between 1 and 4 representing the temporal order of the four foils), and the Condition (the four guilty conditions were dummy coded using the Innocent condition as a baseline). Interaction effects between Condition and Target Item were included to test hypotheses. The results of the separate multilevel models are in Table 1.

Table 1. Overall Movement: Multilevel Regression Model Results

Fixed Effects	Overall Movement β (S. E.)	Pupil Diameter β (S. E.)	Max Vocal Pitch β (S. E.)
(Intercept)	0.007	**0.090***	0.077
	(0.010)	(0.045)	(0.047)
Target Item	-0.026	-0.074	**-0.247****
	(0.019)	(0.082)	(0.086)
Time	0.000	**-0.008****	-0.003
	(0.001)	(0.003)	(0.003)
Guilt	0.015	**-0.123***	-0.061
	(0.012)	(0.053)	(0.056)
Mental Countermeasures (MC)	0.019	**-0.167****	-0.049
	(0.013)	(0.054)	(0.057)
Physical Countermeasures (PC)	**0.041****	**-0.128***	-0.070
	(0.013)	(0.054)	(0.057)
All Countermeasures (AC)	0.013	**-0.204*****	-0.040
	(0.012)	(0.054)	(0.056)
Guilt X Target Item	**-0.076****	**0.611*****	**0.304***
	(0.027)	(0.119)	(0.124)

MC X Target Item	**-0.097*****	**0.835*****	0.243
	(0.028)	(0.121)	(0.130)
PC X Target Item	**-0.203*****	**0.639*****	**0.348****
	(0.028)	(0.121)	(0.128)
AC X Target Item	**-0.063***	**1.022*****	0.198
	(0.028)	(0.121)	(0.126)
N	62780	3058	2994

* p < .05; ** p < .01; *** p < .001; models fit using maximum likelihood

Each model was compared to an unconditional multilevel regression model that excluded fixed effects, and each explained significantly more variance. The unconditional model partitions the variance across participants unconditioned by predictor variables. Comparing the unconditional models against each of the models allows for testing if the inclusion of the predictors significantly improves the fit of the model to the data (Singer and Willett 2003) before examining the fixed effects.

Kinesic Rigidity

Kinesic rigidity was detected among all groups who were smuggling the IED. When the target item was present on the screen, Guilty (b=-0.076, p=.006), physical countermeasures (b=-.203, p=.000), mental countermeasures (b=-.097, p=.001), and all countermeasures (b=-.063, p=.024) groups exhibited rigidity. Those performing only physical countermeasures showed the greatest amount of rigidity, large enough to create a significant net positive main effect in that condition because of the standardized scoring method.

Pupil Diameter

When the target item (IED image) was displayed on the screen while an examinee responded to a question, pupil dilation was significantly larger for participants in the Guilty (b=.611, p<.001), mental countermeasures (b=.835, p<.001), physical countermeasures (b=.639, p<.001) and all countermeasures (b=1.022, p< .001) group, which showed the largest effect. There were relatively narrow pupil main effects of each condition resulting from the pronounced pupil dilation that occurred when target items were present, since all items were standardized within foil.

Vocal Pitch Variation

Max pitch was selected as the dependent variable for vocal pitch variation in the multilevel regression model detailed in Table 1. Both physical (b=.348, p=.007) and guilty without countermeasures (b=.304, p=.014) conditions demonstrated increases in max pitch when responding to target items. The mental countermeasures condition did not achieve significance, b=.24, p=.06. The condition using several countermeasures was not significantly different from the innocent condition, b= .19, p=.12.

Discussion

This study was a novel examination of countermeasures in automated interviewing systems, an emerging paradigm in IS research. This is also the first study exploring a systems solution to countermeasures in the field of automated credibility assessment. Traditional countermeasures had limited effectiveness against automated interviewing systems in CIT-driven interviews. The results also suggest that there is potential in the proposition that tracking multiple heterogeneous veracity cues will decrease the effectiveness of countermeasures.

Summary of Results

Overall body movement and pupil dilation had been investigated in CIT interviews previously, and the results replicated prior work showing kinesic rigidity (H4) and dilated pupils (H1) during presentation of target items. Traditional countermeasures were not effective at countering these behavioral and physiological responses, in agreement with rigidity predictions (H5) but contrary to expectations for pupil dilation (H2, H3). Pupil dilation was the strongest effect among those investigated and appeared to be the most resilient to countermeasures. The pupil dilation resulting from the orienting response was strong, and there was no decrease in this effect when mental distraction or pain was used.

Vocalic features had been explored in alternative interviewing protocols, and those studies indicated variation in vocal pitch stemming from increased anxiety or tension may be exhibited in CIT responding. Participants in the guilty with no countermeasures condition exhibited greater variation in max vocal pitch, as did those in the physical countermeasures condition, as hypothesized (H6, H8). The prediction that vocalic variance changes could be countered by mental countermeasures (H7) was supported.

The vocal findings from this study reveal that max vocal pitch could be used to discriminate between guilty and innocent speakers during a traditional or automated CIT. When countermeasures are employed, the effective method against vocal detection was using mental countermeasures. Mental countermeasures were successful even when employing many countermeasures at the same time, possibly suggesting that individuals find vocal manipulation easier than other countermeasures such as controlling body movement or pupil dilation.

Attempting many countermeasures at the same time proved difficult. When individuals tried to control many things at once, the pupil dilation effect was strongest. Physical countermeasures produced the strongest levels of rigidity. These findings may be used used to detect specific countermeasures.

Contributions

By investigating countermeasures in a new domain (automated screening systems using non-contact sensors), this IS study is breaking new ground. However, we should learn from related work that has sought to approach the countermeasures problem. Frequently, when new tests or protocols are developed, they are initially claimed to be resistant to countermeasures (e.g., Lykken 1960; Rosenfeld and Labkovsky 2010a; Rosenfeld et al. 2008). In many cases, the matter is not that countermeasures will not work, but simply that the same countermeasures previously employed in other deception detection tests do not apply in the new test. Lykken (1960) claimed that the newly developed CIT (then called the Guilty Knowledge Test) was resistant to countermeasures. Later, several countermeasures were shown to reduce the accuracy of Lykken's CIT method (Honts et al. 1996) using the polygraph. The same phenomenon occurred in the P300 based test with the advent of the Complex Trial Protocol (Rosenfeld et al. 2008). The countermeasures previously used on the P300 were shown to be easily detected using the new protocol. However, Sokolovsky et al. (2011) and Hu et al. (2011) both showed that a slightly modified version of the countermeasure was able to reduce accuracy substantially.

The approach to countermeasures in this study took a similar approach, but with an added systems-inspired proposition—triangulating on deception through measurement of multiple behavioral and psychophysiological anomalies simultaneously. The findings presented in this paper suggest that this approach may be effective in some areas (e.g., pupil dilation and body movement), but not others (e.g., vocal pitch). The results are promising enough to justify additional research investigating countermeasure combinations at a more granular level.

The study presented in this paper makes an important contribution to IS credibility assessment literature as it contains a seminal evaluation of the effectiveness of countermeasures against systems designed to identify deception and concealed information in an automated credibility assessment context. Technologies that can identify deception and concealed information rapidly and without contact have the potential to be used in a variety of interviewing and screening contexts, changing how integrity and security are managed. As with the polygraph, these new credibility assessment systems will encounter some individuals who will attempt to mitigate their effectiveness through the use of countermeasures. Ultimately, the results of this study will feed back into the system design, driving revised detection algorithms, refined interactions, and key procedural modifications. Thus, this study represents one step in a much larger effort to create system-driven solutions to credibility assessment.

Limitations and Future Directions

One limitation of this work is the set of countermeasures used by participants in the countermeasures treatments. The countermeasures taught to participants had been identified in previous deception literature, however, these countermeasures were traditionally employed to thwart polygraph examinations. While the concept of using countermeasures to create bogus physiological and behavioral responses certainly applies to the sensors used in this study, there may exist additional countermeasures more suited for use against sensors not requiring contact with the interviewee. As such, an area of future research is the identification of new types of countermeasures that may be employed specifically for use against the sensors evaluated in this research. Future work should also assess the effectiveness of these novel countermeasures.

Training is a particular limitation of this study, since participants had but a few minutes to learn countermeasures and practice them on their own. It is reasonable to believe that countermeasure effectiveness may improve with greater training and practice, and future studies should examine the effects of various levels of training. However, relatively small amounts of practice represents the large majority of real-world attempts to counter credibility assessment interviews. Individuals rarely train for extended periods before taking a polygraph exam or smuggling illicit items into a sports arena or secure building. Nevertheless, highly motivated individuals will do extensive training, and the effectiveness of such training is an open area of research. Future research can examine this using an experimental design similar to the current study, but including varying levels of training and practice.

Another type of countermeasures less commonly studied in academic literature is the chemical countermeasure. Bradley and Ainsworth (1984) studied the effects of intoxication during the crime on the detection of deception during both the CIT and CQT following the crime. Intoxication during the commission of the crime reduced the detection accuracy of both tests as compared to sober participants. There are significant ethical challenges to conducting examinations of chemical-based countermeasures which prevented our including them in this round. However, future research will need to address this issue, perhaps via a standoff intoxication identification function.

Finally, another area of future research is investigating the interrelationships between disparate cues of deception, correlating newly identified cues in automated screening with validated deception detection cues such as electrodermal variation. Simultaneously collecting behavioral and psychophysiological data using a variety of sensors should yield new insights regarding the underlying relationships of the mechanisms triggering certain responses associated with deception. Insights gleaned in this area will yield important findings not only for improving detection accuracy but also illuminating theoretical understanding of deception and its accompanying cues.

Conclusion

This study examined the effectiveness of countermeasures in the newly emerging field of automated human interviewing systems. These new systems employ non-traditional measurement and new techniques, and the effectiveness of countermeasures was unknown. The results show that traditional countermeasures were largely ineffective, as were attempts to also employ direct countermeasures alongside traditional ones. Future research will seek to further confirm the notion that an effective system for combating countermeasures is one that tracks multiple, distince psychophysiological and behavioral indicators of deception simultaneously.

Acknowledgments

The Department of Homeland Security's (DHS) National Center for Border Security and Immigration (BORDERS) and the Center for Identification Technology Research (CITeR), a National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC), provided funding for this research. Any statement provided herein does not necessarily represent the opinion of the funding organizations.

References

Alder, K. 1998. "To Tell the Truth: The Polygraph Exam and the Marketing of American Expertise," *Historical Reflections* (24:3), pp. 487-525.

Bachorowski, J.-A., and Owren, M.J. 1995. "Vocal Expression of Emotion: Acoustic Properties of Speech Are Associated with Emotional Intensity and Context," *Psychological Science* (6:4), pp. 219-224.

Beatty, J., and Kahneman, D. 1966. "Pupillary Changes in Two Memory Tasks," *Psychonomic Science* (5), pp. 371-372.

Ben-Shakhar, G. 1985. "Standardization within Individuals: A Simple Method to Neutralize Individual Differences in Skin Conductance," *Psychophysiology* (22:3), pp. 292-299.

Ben-Shakhar, G., and Dolev, K. 1996. "Psychophysiological Detection through the Guilty Knowledge Technique: Effect of Mental Countermeasures," *Journal of Applied Psychology* (81:3), pp. 273-281.

Bond, C.F., and DePaulo, B.M. 2006. "Accuracy of Deception Judgments," *Personality and Social Psychology Review* (10:3), pp. 214-234.

Bradley, M., and Ainsworth, D. 1984. "Alcohol and the Psychophysiological Detection of Deception," *Psychophysiology* (21:1), pp. 63-71.

Bradley, M.M., Miccoli, L., Escrig, M.A., and Lang, P.J. 2008. "The Pupil as a Measure of Emotional Arousal and Autonomic Activation," *Psychophysiology* (45:4), pp. 602-607.

Buller, D.B., and Aune, R.K. 1987. "Nonverbal Cues to Deception among Intimates, Friends, and Strangers," *Journal of Nonverbal Behavior* (11:4), pp. 269-290.

Buller, D.B., and Burgoon, J.K. 1996. "Interpersonal Deception Theory," *Communication Theory* (6:3), pp. 203-242.

Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., and Mann, S. 2006. "The Impact of Deception and Suspicion on Different Hand Movements," *Journal of Nonverbal Behavior* (30:1), pp. 1-19.

Chapman, C.R., Oka, S., Bradshaw, D.H., Jacobson, R.C., and Donaldson, G.W. 1999. "Phasic Pupil Dilation Response to Noxious Stimulation in Normal Volunteers: Relationship to Brain Evoked Potentials and Pain Report," *Psychophysiology* (36:1), pp. 44-52.

DePaulo, B.M. 1992. "Nonverbal Behavior and Self-Presentation," *Psychological bulletin* (111:2), p. 203.

DePaulo, B.M., and Kirkendol, S.E. (eds.). 1989. *The Motivational Impairment Effect in the Communication of Deception*. Dordrecht, Netherlands: Kluwer.

DePaulo, B.M., Kirkendol, S.E., Tang, J., and O'Brien, T.P. 1988. "The Motivational Impairment Effect in the Communication of Deception: Replications and Extensions," *Journal of Nonverbal Behavior* (12:3), pp. 177-201.

DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., and Cooper, H. 2003. "Cues to Deception," *Psychological Bulletin* (129:1), pp. 74-118.

Derrick, D., Jenkins, J.L., and Nunamaker, J.F. 2011. "Design Principles for Special Purpose, Embodied, Conversational Intelligence with Environmental Sensors (SPECIES)," *AIS Transactions on Human-Computer Interaction* (3:2), pp. 62-81.

Ekman, P., and Friesen, W.V. 1969. "Nonverbal Leakage and Clues to Deception," *Psychiatry* (32:1), pp. 88-&.

Ekman, P., and Friesen, W.V. 1972. "Hand Movements," *Journal of Communication* (22), pp. 353-374.

Elaad, E., and Ben-Shakhar, G. 1991. "Effects of Mental Countermeasures on Psychophysiological Detection in the Guilty Knowledge Test," *International Journal of Psychophysiology* (11:2), pp. 99-108.

Elkins, A.C., and Burgoon, J.K. 2010. "Validating Vocal Analysis Software to Assess Credibility in Interpersonal Interaction: A Multilevel Factor Analytic Approach," *National Communication Association 96th Annual Convention*, San Francisco, CA, USA.

Elkins, A.C., Derrick, D.C., and Gariup, M. 2012a. "The Voice and Eye Gaze Behavior of an Imposter: Automated Interviewing and Detection for Rapid Screening at the Border," *Proceedings of the Workshop on Computational Approaches to Deception Detection*: Association for Computational Linguistics, pp. 49-54.

Elkins, A.C., Derrick, D.C., and Gariup, M. 2012b. "The Voice and Eye Gaze Behavior of an Imposter: Automated Interviewing and Detection for Rapid Screening at the Border," *Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France.

Ellermeier, W., and Westphal, W. 1995. "Gender Differences in Pain Ratings and Pupil Reactions to Painful Pressure Stimuli," *Pain* (61:3), pp. 435-439.

Fukuda, K. 2001. "Eye Blinks: New Indices for the Detection of Deception," *International Journal of Psychophysiology* (40:3), pp. 239-245.

Fuller, C.M., Biros, D.P., and Wilson, R.L. 2009. "Decision Support for Determining Veracity Via Linguistic-Based Cues," *Decision Support Systems* (46:3), pp. 695-703.

Gamer, M., Rill, H.G., Vossel, G., and Godert, H.W. 2006. "Psychophysiological and Vocal Measures in the Detection of Guilty Knowledge," *International Journal of Psychophysiology* (60:1), pp. 76-87.
Ganis, G., Rosenfeld, J., Meixner, J., Kievit, R., and Schendan, H. 2011. "Lying in the Scanner: Covert Countermeasures Disrupt Deception Detection by Functional Magnetic Resonance Imaging," *NeuroImage* (55:1), pp. 312-319.
Gardner, R.M., Beltramo, J.S., and Krinsky, R. 1975. "Pupillary Changes During Encoding, Storing, and Retrieval of Information," *Perceptual and Motor Skills* (41:3), pp. 951-955.
Goldwater, B.C. 1972. "Psychological Significance of Pupillary Movements," *Psychological Bulletin* (77:5), pp. 340-355.
Gray, J.A. 1988. *The Psychology of Fear and Stress*, (2nd ed.). Cambridge, UK: Cambridge University Press.
Gregor, S., and Hevner, A.R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.
Harnsberger, J.D., Hollien, H., Martin, C.A., and Hollien, K.A. 2009. "Stress and Deception in Speech: Evaluating Layered Voice Analysis," *Journal of Forensic Sciences* (54:3), pp. 642-650.
Heaver, B. 2011. "Psychophysiological Indices of Recognition Memory." Brighton, England: University of Sussex.
Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.
Honts, C.R., Devitt, M.K., Winbush, M., and Kircher, J.C. 1996. "Mental and Physical Countermeasures Reduce the Accuracy of the Concealed Knowledge Test," *Psychophysiology* (33:1), pp. 84-92.
Honts, C.R., Hodes, R.L., and Raskin, D.C. 1985. "Effects of Physical Countermeasures on the Physiological Detection of Deception," *Journal of Applied Psychology* (70:1), pp. 177-187.
Honts, C.R., and Kircher, J.C. 1994. "Mental and Physical Countermeasures Reduce the Accuracy of Polygraph Tests," *Journal of Applied Psychology* (79:2), pp. 252-259.
Honts, C.R., and Perry, M.V. 1992. "Polygraph Admissibility," *Law and Human Behavior* (16:3), pp. 357-379.
Honts, C.R., Raskin, D.C., and Kircher, J.C. 1987. "Effects of Physical Countermeasures and Their Electromyographic Detection During Polygraph Tests for Deception," *Journal of Psychophysiology* (1:3), pp. 241-247.
Hu, X., Hegeman, D., Landry, E., and Rosenfeld, J. 2011. "Increasing the Number of Irrelevant Stimuli Increases Ability to Detect Countermeasures to the P300-Based Complex Trial Protocol for Concealed Information Detection," *Psychophysiology* (49:1), pp. 85-95.
Iacono, W.G., and Lykken, D.T. 1997. "The Validity of the Lie Detector: Two Surveys of Scientific Opinion," *Journal of Applied Psychology* (82:3), pp. 426-433.
Janisse, M.P. 1977. *Pupillometry: The Psychology of the Pupillary Response*. Washington D.C.: Hemisphere Publishing Corporation.
Janisse, M.P., and Bradley, M.T. 1980. "Deception, Information and the Pupillary Response," *Perceptual and Motor Skills* (50:3), pp. 748-750.
Jones, E.E., and Pittman, T.S. 1982. "Toward a General Theory of Strategic Self-Presentation," *Psychological perspectives on the self* (1), pp. 231-262.
Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall Inc.
Liang, Y., Lee, S.A., and Jang, J.. 2013. "Mindlessness and gaining compliance in Computer-Human Interaction," *Computers in Human Behavior* (29:4), pp. 1572-1579.
Lykken, D.T. 1959. "The Gsr in the Detection of Guilt," *Journal of Applied Psychology* (43:6), pp. 385-388.
Lykken, D.T. 1960. "The Validity of the Guilty Knowledge Technique: The Effects of Faking," *Journal of Applied Psychology* (44:4), pp. 258-262.
Lykken, D.T. 1998. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York, NY: Plenum Trade.
MacLaren, V.V. 2001. "A Quantitative Review of the Guilty Knowledge Test," *Journal of Applied Psychology* (86:4), pp. 674-683.
Maw, N.N., and Pomplun, M. 2004. "Studying Human Face Recognition with the Gaze-Contingent Window Technique," *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*, Chicago, Illinois: Cognitive Science Society, pp. 927-932.

Meservy, T.O., Jensen, M.L., Kruse, J., Burgoon, J.K., Nunamaker, J.F., Jr., Twitchell, D.P., Tsechpenakis, G., and Metaxas, D.N. 2005. "Deception Detection through Automatic, Unobtrusive Analysis of Nonverbal Behavior," *IEEE Intelligent Systems* (20:5), pp. 36-43.

Nass, C., Fogg, B. and Moon, Y. 1996. "Can computers be teammates?" *International Journal of Human-Computer Studies*, (*45*:6), pp. 669–678.

National Research Council. 2003. "The Polygraph and Lie Detection," Committee to Review the Scientific Evidence on the Polygraph (ed.). Washington, D.C., USA: The National Academies Press.

Nunamaker, J.F., and Briggs, R.O. 2011. "Toward a Broader Vision for Information Systems," *ACM Transactions on Management Information Systems* (2:4), pp. 20: 01-12.

Nunamaker, J.F., Chen, M., and Purdin, T.D.M. 1991. "Systems Development in Information Systems Research," *Journal of Management Information Systems* (7:3), pp. 89-106.

Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., and Patton, M.W. 2011. "Embodied Conversational Agent–Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), pp. 17-48.

Nunamaker, J.F., Twyman, N.W., and Giboney, J.S. 2013. "Breaking out of the Design Science Box: High-Value Impact through Multidisciplinary Design Science Programs of Research " in: *Proceedings of the Americas Conference on Information Systems*. Chicago, Illinois, USA.

Nunamaker Jr, J.F., Burgoon, J.K., Twyman, N.W., Proudfoot, J.G., Schuetzler, R., and Giboney, J.S. 2012. "Establishing a Foundation for Automated Human Credibility Screening," *IEEE International Conference on Intelligence and Security Informatics (ISI)*, Washington D.C.

Nunnally, J.C., Knott, P.D., Duchnowski, A., and Parker, R. 1967. "Pupillary Response as a General Measure of Activation," *Attention, Perception, & Psychophysics* (2:4), pp. 149-155.

Osher, D.B. 2007. "Multimethod Assessment of Deception: Oculomotor Movement, Pupil Size, and Response Time Measures.," in: *Department of Educational Psychology*. Salt Lake City, Utah, USA: University of Utah.

Pavlidis, I., Eberhardt, N.L., and Levine, J.A. 2002. "Seeing through the Face of Deception," *Nature* (415:6867), pp. 35-35.

Pavlidis, I., and Levine, J. 2002. "Thermal Facial Screening for Deception Detection," *24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society*, pp. 1143-1144.

Proudfoot, J.G., Twyman, N.W., and Burgoon, J.K. 2012. "A Comparison of Invasive and Noninvasive Sensors in the Concealed Information Test," in: *European Intelligence and Security Informatics Conference*. Odense, Denmark.

Proudfoot, J.G., Twyman, N.W., and Burgoon, J.K. 2013. "Eye Tracking and the CIT: Utilizing Oculometric Cues to Identify Familiarity with Wanted Persons," *46th Annual Hawaii International Conference on System Sciences*, Maui, Hawaii.

Reeves, B. and Nass, C. I. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rockwell, P., Buller, D.B., and Burgoon, J.K. 1997. "Measurement of Deceptive Voices: Comparing Acoustic and Perceptual Data," *Applied Psycholinguistics* (18), pp. 471-484.

Rosenfeld, J., and Labkovsky, E. 2010a. "New P300-Based Protocol to Detect Concealed Information: Resistance to Mental Countermeasures against Only Half the Irrelevant Stimuli and a Possible Erp Indicator of Countermeasures," *Psychophysiology* (47:6), pp. 1002-1010.

Rosenfeld, J., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., and Chedid, E. 2008. "The Complex Trial Protocol (Ctp): A New, Countermeasure-Resistant, Accurate, P300-Based Method for Detection of Concealed Information," *Psychophysiology* (45:6), pp. 906-919.

Rosenfeld, J., Soskins, M., Bosh, G., and Ryan, A. 2004. "Simple, Effective Countermeasures to P300-Based Tests of Detection of Concealed Information," *Psychophysiology* (41:2), pp. 205-219.

Rosenfeld, J.P., and Labkovsky, E. 2010b. "New P300-Based Protocol to Detect Concealed Information: Resistance to Mental Countermeasures against Only Half the Irrelevant Stimuli and a Possible Erp Indicator of Countermeasures," *Psychophysiology* (47:6), pp. 1002-1010.

Singer, J., and Willett, J. 2003. *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, USA: Oxford Press.

Scherer, K.R., Johnstone, T., and Klasmeyer, G. 2003. "Vocal Expression of Emotion," *Handbook of affective sciences*), pp. 433-456.

Sokolovsky, A., Rothenberg, J., Labkovsky, E., Meixner, J., and Rosenfeld, J. 2011. "A Novel Countermeasure against the Reaction Time Index of Countermeasure Use in the P300-Based

Complex Trial Protocol for Detection of Concealed Information," *International Journal of Psychophysiology* (81:1), pp. 60-63.

Sundar, S. S. and Nass, C. 2000. "Source Orientation in Human-Computer Interaction Programmer, Networker, or Independent Social Actor." *Communication Research*, (*27*:6), pp. 683–703.

Titze, I.R., and Martin, D.W. 1998. "Principles of Voice Production," *The Journal of the Acoustical Society of America* (104), p. 1148.

Trovillo, P.V. 1939. "A History of Lie Detection," *Journal of Criminal Law and Criminology* (29), pp. 848-881.

Twyman, N.W. 2012. "Automated Human Screening for Detecting Concealed Information," in: *Department of Management Information Systems*. Tucson, AZ: University of Arizona.

Twyman, N.W., Burgoon, J.K., Elkins, A.C., and Proudfoot, J.G. 2013. "Alternative Cues in Concealed Information Testing," *46th Annual Hawaii International Conference on System Sciences*, Maui, Hawaii.

Twyman, N.W., Burgoon, J.K., and Nunamaker, J.F. 2011a. "Non-Invasive Screening for Concealed Information," *Proceedings of the Symposium on Credibility Assessment and Information Quality in Government and Business at the 44th Annual Hawaii International Conference on System Sciences (HICSS)*, Koloa, Hawaii.

Twyman, N.W., Elkins, A., and Burgoon, J.K. 2011b. "A Rigidity Detection System for the Guilty Knowledge Test," *HICSS-44 Symposium on Credibility Assessment and Information Quality in Government and Business*, Koloa, HI, USA: IEEE.

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

Valacich, J.S., Jenkins, J.L., Nunamaker, J.F., Hariri, S., and Howie, J. 2013. "Identifying Insider Threats through Monitoring Mouse Movements in Concealed Information Tests," in: *HICSS-46 Symposium on Credibility Assessment and Information Quality in Government and Business*. Maui, HI, USA.

Vizer, L.M., Zhou, L., and Sears, A. 2009. "Automated Stress Detection Using Keystroke and Linguistic Features: An Exploratory Study," *International Journal of Human-Computer Studies* (67:10), pp. 870-886.

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

Vrij, A., and Mann, S. 2001. "Telling and Detecting Lies in a High-Stake Situation: The Case of a Convicted Murderer," *Applied Cognitive Psychology* (15), pp. 187-203.

Vrij, A., Semin, G.R., and Bull, R. 1996. "Insight into Behavior Displayed During Deception," *Human Communication Research* (22:4), pp. 544-562.

Zhou, L., and Zhang, D. 2008. "Following Linguistic Footprints: Automatic Deception Detection in Online Communication," *Communications of the ACM* (51:9), pp. 119-122.

Zuckerman, M., DePaulo, B.M., and Rosenthal, R. 1981. "Verbal and Nonverbal Communication of Deception," *Advances in Experimental Social Psychology* (14:1), pp. 1-59.