EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Dean A. Pollina[*]
Frank Horvath[**]
John W. Denver[***]
Andrew B. Dollins[****]
Troy E. Brown[*****]
Department of Defense Polygraph Institute
Fort Jackson, USA

Development of Technologies and Test Formats for Credibility Assessment

Introduction

Because Credibility Assessment is a relatively new and developing field there are different ways of describing it and what it encompasses. For ihe purposes of this paper Credibility Assessment is defined as the process of determining the reliability and validity of information, regardless of source. This informa-

[*] e-mail
[**] e-mail
[***] e-mail
[****] e-mail
[*****] e-mail

tion may include but is not limited to that which is collected by physiological and behavioral measures acquired overtly or covertly (Department of Defense Polygraph Institute, 2006).

In May, 2007 the Department of Defense Polygraph Institute (DoDPI) was officially renamed as the Defense Academy for Credibility Assessment (**daca**). This change was made, in part, because of changes in the field and, in part, because the interests of DoDPI had become, at least in the past two decades or so, focused more heavily on technologies and approaches to credibility assessment that were not technically based on polygraphs technology. Our primary purpose in this chapter is to describe these other approaches and to discuss the state-of-the-art regarding them. First, however, an overview of polygraph testing, the most widely known and perhaps most controversial approach to credibility assessment, often specified, inaccurately, as "lie detection" and more recently as "detection of deception," needs to be considered. Research on that issue, as will be seen, presents difficult methodological and other problems that must be considered when carrying out useful research in the field of credibility assessment whatever the approach or technology that is at issue.

Technologies of Interest

The Polygraph

The polygraph has been used by the police in the investigation of serious crimes since at least the early 1900's (Reid and Inbau, 1977; Trovillo, 1939; 1942). It is important to note that historical development in the field can be traced along two lines, one line involving instrumentation and the other testing techniques.

In 1887 Cesare Lombroso, an Italian criminologist, used a hydrosphygmograph and a "scientific cradle," for objective measurement of physiological changes associated with the detection of deception. Shortly after, an American psychologist, Hugo Munsterberg noted the effect of lying on breathing, cardiovascular activity and the galvanic skin response (GSR)— apparent changes in electrical resistance in the skin. In 1921 John Larson devised an instrument for making continuous recordings of both blood pressure and breathing. In 1930, Leonarde Keeler, generally credited with developing the prototype of the present-day polygraph, added a device for recording GSR. Modern computerized polygraphs are technical improvements over earlier devices, although the physiological activities recorded are essentially the same. The polygraph captures electrodermal activity (EDA) by means of two electrodes usually attached to the hand or fingertips. A standard blood pressure cuff is used to record relative blood pressure and pulse rate. Finally, breathing activity is recorded by "pneumograph" tubes which expand and contract with chest cavity movement. Activity in each of these physiological systems is usually converted from analog to digital form for display on a computer monitor. The "chart" display can be stored permanently on standard media for viewing and analysis.

Testing Approaches. In the formative years of polygraph testing ("lie detection") practitioners were few and the testing approaches that were used were highly individualized; that is, each of the early polygraph examiners carried out testing across situations in highly idiosyncratic ways (Alder, 2007; Reid and Inbau, 1977). Over time, however, testing approaches became more clearly defined through empirical observations and training protocols, though it is commonly understood that even today the approaches have not been standardized in ways that permit useful scientific assessments of some features of testing in applied settings. Importantly, all of these approaches are based on the same premise: There is no known physiological response that is unique to lying. Neither the polygraph nor any other device is capable of detecting a "lie." Lie detection, then, is at the present time an inferential process in which "lying" is inferred from comparisons of physiological responses to categories of questions that are asked during polygraph testing. There are three major families of testing procedures in use today: the relevant/irrelevant technique (R/I), the control [comparison] question technique (COT) and information recognition testing (IRT). Each of these procedures has its own advantages and disadvantages and each may be the procedure of choice, depending upon the application of interest (e.g., a criminal investigation, an employee "screening") and the circumstances confronting the examiner.

In its simplest form the R/I technique consists of asking a series of relevant questions, those pertinent to the crime at hand (e.g., "Did you shoot John Doe?"), among irrelevant questions that are not crime related (e.g., "Are you over eighteen years of age?"). The test questions, perhaps 10-12 items, are asked several times during the testing, typically once or twice within a test. A test consists of a single presentation of the complete question list by the examiner. An assumption implicit in the R/I technique is that truthful persons will not react differentially to a great degree to relevant and irrelevant questions, while people lying will. This assumption has been seriously challenged and is the primary reason that the CQT was developed by J. Reid (1947). Today, the CQT is the preferred method of "lie detection;" it is certainly the most widely used, and is also the most controversial testing procedure.

In the CQT the question list consists of irrelevant, relevant, and "control" [comparison] questions. Additionally, other types of questions may be included that test what are said to be individual differences in perceptions of the testing situation. For instance, in some variations a "sacrifice relevant" question is included; it serves as a buffer to the asking of the crime-relevant questions and is not "scored" during the data analysis process. The relevant and irrelevant questions are similar to those asked during R/I testing. The relevant questions are expressed clearly, succinctly and directly, without conjunctions, and typically relate only to a single event (a homicide, a robbery, etc.) under investigation. In some instances, the relevant questions may be re-phrasings of the same question. Comparison questions deal with matters similar to, but of presumed lesser significance than, the offense under investigation. The examiner interacts with the examinee in the pre-test interview to frame these questions properly so that they will be "probable lies." Generally, a single question list, depending on the variation of the CQT that is being applied, will consist of three or four relevant questions, two or three irrelevant questions, and two or three "comparison" questions that are often asked in a position immediately adjacent to the relevant questions. The question list is typically asked in three repetitions, though in some circumstances, when the data may be unclear, five repetitions are included in a single examination. Inferences of truthfulness and deception in the CQT are made by a systematic comparison of the responses from each of the physiological measures to the relevant and the comparison questions. Simply stated, more consistent and more pronounced physiological responses to relevant questions than to comparison questions indicate lying on the relevant issues. Conversely, consistently greater physiological responses to comparison than to relevant questions indicate truthfulness in the matter under investigation. In most circumstances, the CQT data are "scored" manually by assigning numeric values to the difference in response magnitudes between pairs of comparison and relevant questions. Within each pair and for each of the physiological measures, a "score" from 1 to 3 is assigned, with a "1" indicating a small difference, a "2" a moderate difference and a "3" to a pronounced difference; a "0" is assigned when there is no discernable difference. If the difference is greater to the relevant question in the pair, a negative sign is assigned; if to the comparison question, a positive sign. These values are algebraically accumulated across all presentations of the questions, that is across all of the "tests." A total "score" of+6 or greater or -6 or less typically serves as a cut-off for a decision, with the former score suggesting truthfulness to the relevant issue and the latter "deceptiveness." A score falling between those two values produces an "inconclusive" outcome, a result occurring in about 10% of the cases.

The use of "tests" in the IRT family of procedures may be included in the testing protocol when other procedures are applied, or they may be used independent of other approaches. For example, it is common to find in application of the CQT that an "acquaintance" test is carried out. Such a test is typically administered to demonstrate how the polygraph works. The test typically involves asking an examinee to conceal information, such as a chosen number in a series of numbers, and then "lie" about the chosen number as the list is presented while polygraphs data are collected. Recognition of the chosen number usually produces a greater physiological response than that produced by the other numbers. Less common but more important uses of IRT approaches are found in situations in which specific details of a criminal offense are known only to the police and the actual perpetrator(s). For example, assume that in a homicide investigation, the police know that a person was killed with a club; that and other pieces of information about the cause of death or other details of the offense have not been revealed to the public, In such a situation a "test" could be constructed consisting of the asking of a question stem and multiple options. For example, the stem might be: "Do you know if John Doe was killed with a?" The options might be the names of various, similar weapons, (e.g., gun, knife, club, etc.), including the one actually used, all asked in random order. The guilty person, recognizing the correct option (i.e., the club), would be expected to show a greater physiological response to that stimulus than to the others whereas an innocent person would not. Typically, a series of three or more such multiple choice tests would be carried out, provided that sufficient detailed information about the offense is known. Recognition of the correct option (i.e., the "key" item) in these tests would suggest that the examinee was concealing knowledge of the offense and thus an inference of "guilt" could be drawn. One can readily see that if there are more "tests," each including a critical "key" item, it would be possible to calculate the exact probability of chance responses to the "keys' by someone who does not possess "guilty knowledge." This procedure is referred to in the literature as a Concealed Information Test (CIT) or as a Guilty Knowledge (GKT) Test (Ben-Shakhar, Bar-Hillel, and Licblich, 1986; Lykken, 1959; 1960).

Another form of the IRT approach is what is referred to in the field literature as a "peak of tension" test (POT). This procedure can be used in circumstances similar to those described for the CIT. The difference here is that evaluation of the response data typically involves not only a specific response to the "key", as in the CIT, but also an anticipatory "response" which dissipates subsequent to the presentation of the "key" item, usually not placed randomly in the list but, by design, in the middle of the series of options. The POT approach can also be used when there is interest in searching for information which is not known to the examiner (or the police) but is assumed to be within the examinee's knowledge. For instance, assume that an examinee denies knowing where, within a defined location, such as a square mile of land, a body has been buried. In such a situation an examiner may construct a POT asking about specific points of interest within the suspected location to determine if the examinee "recognizes" one of those points; that is, if the examinee produces a greater physiological response to one item than to the others.

In all of the IRT approaches it is assumed that an examinee who is concealing knowledge about the matter under investigation, that is, about the "key" items, will produce physiological responses to them distinguishable from those to the non-key items. If the differential between those two categories of questions is not apparent, then an inference of "guilt" or "deception" would not be warranted. The IRT procedures, especially the "CIT" method, have been the predominant mode of laboratory "lie detection" studies. There is stronger scientific support for this method than for the more commonly used CQT (National Research Council, 2003). However, because most criminal investigations do not lend themselves to the use of the CIT and because that method is not suitable for testing in screening contexts, it is not often applied in real-life situations. Researchers, however, often ignore this fact and frequently approach "lie detection" using a CIT-related methodology; for that reason, their findings do not generalize to situations of interest in most applications.

The Examination Process. Regardless of the testing approach that is used, all polygraph examinations in applied settings involve a complex clinical process. While some approaches may be less dependent on this process, that is, on examiner and examinee interaction, than others, an understanding of the effect of the clinical component has not been the focus of most research. It is important though to consider why, in applied settings, such interaction is necessary and how it might influence testing outcomes.

Polygraph examinations are often said to involve three stages of processing, a pre-test interview, an "in-test" phase, and, in some literature, a post-test discussion. What takes place in each of these stages varies somewhat depending upon the procedure to be applied. Because the most common procedure is the CQT, the description here is specific to that approach. During the pre-test interview the examiner explains the instrumentation, the "theory" of the testing process (usually fight/flight response), and the purpose of the testing, that is, the known facts relating to the reason the testing is being carried out. The examinee is requested to provide his or her understanding of the issue and based on that, the examiner prepares the questions to be asked during the in-test phase. Each of the questions is reviewed with the examinee verbatim and, if necessary, the wording of the questions is modified to ensure that there is no misunderstanding. When the examinee agrees that all of the questions are clear and can be answered with either a "yes" or a "no", the in-test phase can begin. It is common to find that an acquaintance test is conducted immediately following the pre-test interview. This is done to demonstrate to the examinee the nature of the testing process and, by some accounts, the efficacy of the testing. It is generally accepted that a poorly conducted pre-test interview, one that is not conducted properly, that is, objectively and impartially, may yield examination results that are "inconclusive" or, perhaps, incorrect.

During the in-test phase the reviewed questions are asked at about 20 second intervals while the polygraphs data are collected. Usually the same question list is presented at least 3 times. Each "chart" (or "test," the asking of the question list one time) accounts for approximately five minutes of time. The time limitation is due to the discomfort caused by inflation of the blood pressure cuff. After the data collection in the in-test phase, the examiner evaluates the examinee's physiological responses and determines if the examinee responded more dramatically and more consistently to one category of question than the other (e.g., comparison or relevant). This is typically done using the numerical scoring system discussed previously; in some instances, this manual "scoring" is supported with the use of algorithms specifically developed for the scoring of COT polygraphs data collected digitally (Applied Physics Laboratory, 1993).

Subsequent to the evaluation of the polygraphs data, the examiner proceeds to the post-test phase of the examination. This discussion period takes various forms depending on the outcome of the scoring of the data and the type of testing procedure that was administered. In some instances, an explanation or clarification of the examinee's position regarding the testing issue results in new questions. If so, additional testing may be carried out at either that time or on a subsequent day.

Broadly speaking, polygraph testing can be categorized as involving either a specific incident or a screening situation. The former type addresses a specific, known event about which the examinee's involvement is under investigation. For instance, polygraph examinations inquiring into an examinee's involvement in a robbery, a theft, a rape, or a murder are common specific incident investigations. In these situations it can be seen that there is an identifiable event and the examinee cither participated in it or didn't. The relevant test questions pertain to that particular event and are direct and unambiguous (e.g., "Did you shoot John Doe?"). Screening examinations are generally related to employment matters; usually the examinees are persons who wish to gain employment in an intelligence or law enforcement agency, or to continue to hold a position of trust, such as those in which security clearances are required. Screening examinations may address security issues (e.g., sabotage, espionage, mishandling classified material) or lifestyle matters (e.g., drug use, falsifying information).

In screening examinations the relevant questions are somewhat broad in scope and are necessarily more ambiguous than in a specific incident examination. For example, the relevant questions in a screening examination may be: "Did you provide any classified information to an unauthorized person?" or "Did you use any illegal drugs in the past five years?" In each of these instances, the examiner docs not know of a specific incident in which the examinee might have been involved; nor does the examinee always know precisely what behavior is truly significant. Providing classified information to a spouse and providing such information to a foreign agent, for instance, may both be unauthorized disclosures. In both instances an examinee may be reluctant to disclose such conduct. In the latter instance, the reason for that reluctance is obvious. In the former, however, an examinee may fail to disclose to the examiner in the pre-test phase what is in some cases a minor transgression; if so, the relevant question cannot be appropriately modified. That problem, among other things, may complicate the testing process in ways not usually seen in specific-incident testing. It is partly for that reason that the recent review of the research on "lie detection" by the National Research Council (2003) led to the conclusion, that screening tests might be less accurate than specific-incident tests.

In this brief overview it can be seen that polygraph testing is applied in a variety of ways across a range of quite different situations. In that sense it has great utility. However, all observers recognize that whatever the value of polygraph testing, there are considerable scientific and practical limitations to the technology, which, by the way, has remained essentially unchanged for over fifty years. The need for new approaches, new technologies, and an enhanced understanding of the theoretical underpinnings of "lie detection" and the broader field of credibility assessment has never been greater. The sense of direction that is given by what is now underway is apparent in the following paragraphs in which some of the newer approaches are discussed.

Infrared (IR) Thermography

Thermography provides a potential non-contact technology to enhance credibility assessment procedures (Figure 1). Dynamic IR thermography is a type of thermal imaging involving the detection of infrared radiance in real-time. Thermographic cameras detect and produce images of radiation in the infrared range of the electromagnetic spectrum (roughly 0.9-14 μm). Because infrared radiation is emitted by all objects, thermography makes it possible to "sec" the environment regardless of the presence or absence of visible light. The amount of radiation emitted by an object increases with temperature; therefore thermography allows one to see variations in temperature. Thermographic camera technology has advanced to the point where even relatively small changes in temperature (i.e., a change of .005 °C or less) are detectable.

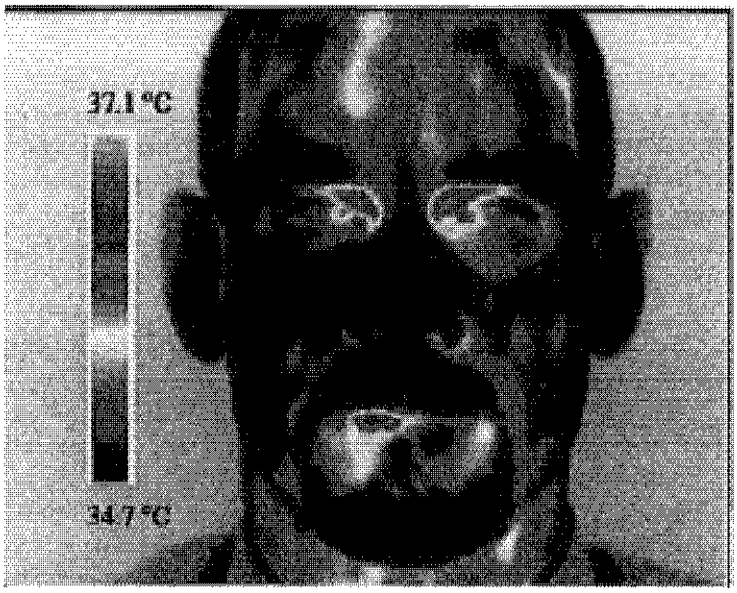


Figure 1. Thermal image showing the temperature distribution across a human face. The periorbital regions are typically hotter than other facial areas, even during resting conditions.

Thermal infrared imagers convert the energy in the infrared wavelength into a visible light video display. All objects above 0 degrees Kelvin emit thermal infrared energy so thermal imagers can passively see all objects regardless of ambient light. The spectrum and amount of thermal radiation depend strongly on an object's surface temperature. This makes it possible for a thermal camera to display an object's temperature. However, other factors also influence the radiation, which globally limits the accuracy of this technique. For example, the radiation depends not only on the temperature of (he object, but is also a function of the emissivity of the object. Emissivity can be thought of as the percent of energy radiated from an object's surface versus the total radiation hitting the surface of that object. Fortunately, human skin has one of the highest emissivity values (approximately 98% of received energy is emitted from skin), making thermography a highly accurate technology for extracting human temperature shifts when confounding variables are limited. When viewed by a thermal camera, warm objects stand out well against cooler backgrounds; humans and other warmblooded animals become easily visible against the environment, day or night. As a result, thermography's use can historically be attributed to the military and security services' need to observe activity under low-light conditions.

The potential utility of 1R thermography in the field of credibility assessment is two-fold. First, 1R thermography may be useful in detecting deception via thermal reactivity specific to deceptive responses. Second, 1R thermography may augment or potentially displace the more traditional contact methods of monitoring physiological activity during polygraph examinations (i.e., cardiac, respiratory and blood pressure transducers).

In 2001, DACA researchers and collaborators at Honeywell conducted a pilot study using non-invasive IR thermography for the extraction of deception information (Pollina and Ryan 2002; Pavlidis, Eberhardt and Levine 2002). Using thermography in conjunction with traditional polygraph measures, the researchers tested for deception-related thermal reactivity in a periorbital region directly below the pupil. The researchers used a zone comparison test (ZCT) based on a mock crime scenario involving simulated theft and assault (Backster, 1963). The researchers (Pavlidis, et. al., 2002) provided promising data suggesting that thermal imaging, in and of itself, was more accurate than traditional polygraph measures in detecting non-deceptive examinees (11/12 vs. 8/12 respectively). They also reported equivalent accuracy on deceptive examinees (6/8 for each method). Pollina and Ryan (2002), using a different algorithm and the same dataset, reported that IR alone had a lower sensitivity (.70) than did traditional FDD measures (.88). However, by combining data from both IR (i.e., data from regions on the face) and traditional polygraph measures there was a slight enhancement of overall accuracy, relative to the use of traditional polygraph data alone. The potential of IR thermography is further supported by a more recent study by Pollina et al. (2006). In this study data are presented from a CIT polygraph exam that followed the mock crime procedure as described in the earlier articles (Pavlidis et al., 2002; Pollina and Ryan, 2002). Results from the CIT paradigm demonstrated that IR alone was more accurate at correctly classifying groups (91.7%) than in the ZCT test paradigm. These two studies have to be viewed as pilot projects that help frame the potential for IR thermography in the field of credibility assessment.

Although IR thermography has shown promise, little is known about the sensitivity of this technology for extracting more traditional variables such as respiration and heart rate. While the reports described thus far have discussed the application of thermal imaging to detect facial thermal reactivity during PDD procedures there are currently no published reports describing the use of thermal imaging to extract the traditional physiological variables used in standard FDD procedures. However, the potential for obtaining physiological measures from 1R data has been presented in several conference papers (see for instance Sun, Garbey, Merla, Pavlidis 2005; Murthy and Pavlidis 2005; Garbey, Sun, Merla and Pavlidis 2005; Murthy, Pavlidis and Tsiamyrtzis 2005).

If movement tracking algoritltms can be further developed and refined, the potential for defining additional physiologically-relevant regions of interest exists. For example, pilot data has shown that the extraction of physiological variables (e.g., respiration, blink rate, and heart rate) from the IR data stream is feasible. If thermography can provide an accurate and sensitive non-contact measure of physiology then it will complement or supplant the use of the traditional contact measurement technologies that monitor cardiovascular function and respiration. This may greatly enhance the capabilities of credibility assessment researchers and investigators and expand the opportunities to monitor individuals. Thus, infrared technology could potentially evolve into a real-time "pre-screening" tool in assessment situations (i.e., customs, border patrol, personnel screening). However, before this technology can be translated into security-related procedures, several important and identifiable problems in dynamic IR imaging need to be investigated and solved. Specifically, research needs to address problems related to movement restrictions, standardization, validation through convergent and synchronous monitoring with other physiological variables, and applicability within existing credibility assessment protocols and paradigms.

Methodological Issues in Thermography

1. Movement distortion of the thermal data stream is a current problem with attempts to apply IR technology in field applications, such as during interview protocols. Existing studies have described the difficulties in accurately extracting reactivity measures from specific regions due to subject movement (e.g., Pavlidis et at., 2002; Pollina et at., 2006). As a participant moves, so do the relevant regions of interest and the corresponding radiometric pixels within the thermal image. Movement confounds the tracking of specific pixel clusters defining each region, making it a difficult task that currently requires extensive and time-consuming off-line analyses. Due to the complex nature of head movement and the inherent vulnerability of IR data extraction to movement confounds, the continued development of technologies that can accurately track movements in real-time while simultaneously extracting relevant temperature data is necessary before this technology will be suitable for field use.
2. There are very few standardized procedures, technologies, or methodologies currently available that would allow for replication or extension of the existing credibility-related research across laboratories. For example, camera technologies are rapidly changing and researchers have not used cameras with similar technical specifications (i.e., spectral ranges and sensitivities). The technologies for thermal cameras are rapidly improving and

these technological changes have resulted in difficulties in replication and the possibility that reported findings are camera specific. There is also a lack of standardized, open-source formats for storing thermal data streams, which further complicates the comparison of research data across laboratories.

3. Although there have been reports describing the application of thermal imaging to detect facial thermal reactivity during PDD procedures (Pollina and Ryan, 2002; Pollina et al. 2006), there are no published reports specifically describing the use of thermal imaging to extract the traditional physiological variables used in standard credibility assessment procedures (e.g., respiration and heart rate signals). In light of these current issues, and until they can be resolved, the applicability of 1R thermography for credibility assessment remains unclear. Only through research, method and system validation and report dissemination can the methods and methodologies crucial to accurate credibility assessment be standardized and therefore incorporated into credibility assessment procedures. There is still a great deal of work to be done for the integration of IR technology into the existing credibility assessment field.

4. The conclusions of the IR studies cited here suggest that a great potential exists for IR technology to enhance the capabilities of current credibility assessment techniques. If the tracking algorithms provide a means for the accurate extraction of IR data, the potential for systematically defining additional physiologically-relevant regions of interest exists. This will allow for the tracking of the periorbital region currently described in the literature, as well as regions such as the nasal passage (for the extraction of respiration measures) and the carotid region (for extracting heart rate data). Additionally, this may allow for the accurate tracking of facial muscle regions (for emotion-related data).

fMRI

Functional magnetic resonance imaging (fMRJ) is the use of MRI to measure the hemodynamic response related to neural activity in the brain and is one of the most recently developed forms of neuroimaging. Blood circulation maintains the supply of oxygen to neurons in the brain. Functional MRI is a procedure for recording these blood flow changes in the brain. Increases in neuronal activity in specific regions of the brain occur when people perform specific cognitive tasks, and these activities create changes in blood flow near these neurons. Using fMRI, scientists can visualize these local blood flow changes in specific brain regions over time, and infer that the brain areas where blood flow changes occur are responsible for cognitive tasks performed during the fMRI recordings.

During fMRI testing, the examinee enters an fMRI scanner - a tube that surrounds the person's head. The person answers questions, performs calculations or other types of cognitive tasks while technicians record blood flow changes inside the brain. The data output from the fMRI process are digital images of two dimensional "slices" through parts of the brain. Statistical analysis of these images can determine which brain areas are more active during specific types of tasks. One cognitive task that has been investigated using fMRI technology is deception (Ganis and Kosslyn, 2007; Ganis, Kosslyn, Stose, Thompson, and Yurgelun-Todd, 2003; Kozel et al, 2004; 2005; Langleben, Loughead, and Bilker, 2005). Researchers have found that when people attempt to deceive others during fMRI testing, specific areas inside the brain are active[1]. Although this technology shows great promise for research purposes, there is not enough credibility-related systematic research at this time to incorporate 1MRI into existing credibility assessment procedures or protocols. Furthermore, utilization of an fMRI system is limited due to the cost and size constraints inherent in the technology. However, these restrictions do not preclude the importance of continued research and development of fMRI technology for credibility assessment.

Laser Doppler Vibrometry

Laser Doppler Vibrometers (LDVs) are non-contact optical instruments used for the accurate measurement of velocity and displacement of vibrating structures. LD V is based on the detection of the Doppler shift of light that is scattered from a small area on a test object due to that object's vibration. The object scatters or reflects light from the laser beam, and the Doppler frequency shift caused by the vibration of the object is used to measure the velocity of vibrations which lie along the axis of the laser beam.

For credibility assessment, the utilization of the LDV system currently focuses on the novel application of this technology for the assessment of physiological activity (Rohrbaugh, Sirevaag, and Ryan, 2006). The LDV assessment method works on the principle that by detecting minute vibrations at the skin surface, the mechanical properties of underlying physiological activity can be recorded. The use of this metric shows promising face-validity as this mechanical activity is often visually observable during various states of physical activity, stress or emotion. For instance, the pulsing arteries present at the temple are often observable when an individual is engaged in physical activity, and can similarly be visually observed in states of high stress. These biological vibra-

[1] This activity appears to be a matter of degree, with specific regions in the anterior cingulate gyrus and frontal lobes being more active when lying, relative to telling the truth. At present, there does not appear to be evidence for a "lie center" in the brain (But See Saxe, Carey, & Kanwis her, 2004 for some interesting related work).

tions are the basis of utilizing LDV technologies for the extraction of physiological activity. Using the LDV method much more subtle forms of activity can be reliably detected and quantified. For example, LDV systems can easily identify the minute vibrations (i.e., those present at the carotid artery region) relating to cardiac pulses, which in turn can be translated into heart rate and other cardiac-related physiological signals. With each heart pulse (and other concomitant vibrations within the biological system), vibrations are present within the circulatory system. These vibrations are easily captured at the carotid artery due to its accessibility and proximity to the heart. The extraction of the velocity and displacement signals from the carotid theoretically allows for the assessment of several physiological signals. The simplest example is the extraction of heart rate. As such, the method has substantial potential for extensive applications in the detection of deception, and for the assessment of credibility in a broader context.

Eye Movement-based Memory Assessment

Physiological response mechanisms have long been used as correlates to credibility. If physiological changes occur, it is logical to assume that changes in other mechanisms, such as cognition and perception, occur as well. Tracking eye movements during the presentation of familiar and novel stimuli has been used to characterize the nature of information processing and how familiarity affects that process. This processing of information occurs independently of consciously mediated control (Cohen and Eichenbaum, 1993).

The technique of eye movement-based memory assessment operates by determining the probability of an individual's prior exposure or familiarity to stimuli based on the eye movement patterns during visual processing of digital images of faces, scenes, and possibly objects. In faces, this effect has been attributed to the underlying cognitive processes involved in perception and shown to differentiate between images of familiar and novel faces (Althoff and Cohen, 1999). Previously viewed items have statistically fewer eye fixations to fewer regions in an image and lower levels of statistical dependency in the patterns of eye movement transitions between regions. Ryan, Althoff, Whitlow, and Cohen (2000) used eye movements to indirectly assess prior exposure to scenes and found a relational manipulation effect indicated by increased viewing of manipulated scene elements for subjects who had viewed the original scenes versus those who had not. Preliminary research indicates similar effects occur for images of objects as well (F. M. Marchak, personal communication, August 15, 2004).

While eye movement-based memory assessment is not a test of credibility, it has potential to become a powerful tool for the assessment of prior knowledge of faces, scenes, and possibly objects. This could be a valuable tool for law enforcement. Current eye movement-based memory assessment systems (TRACKER, Veridical Research and Design, Bozeman, MT) perform noncontact assessments with a specialized monitor embedded with a low-level infrared camera and lights to track eye movements (saccades) and fixations.

Combining Approaches

Unfortunately it is not possible to review the entire literature on emerging credibility assessment technologies in this chapter. A complete review would have to also include the recent advances in behavioral and neurophysiological tests related to credibility assessment (Horvath, Jayne, and Buckley, 1994; Masip, Sporer, Garrido, and Herrero, 2005; Pollina and Squires, 1998; Rosenfeld, Soskins, Bosh, and Ryan, 2004; Verschuere, 2007). Another area that deserves notice uses tracking of facial muscle regions to extract emotion-related data. This approach began with Charles Darwin in his seminal work examining emotion expression across animal species (Darwin, 1997). Paul Ekman and his colleagues have extended this work and attempted to apply it to the credibility assessment field (Ekman, 1992, Ekman, Friesen, and Ancoli, 1980. Ekman, Friesen, and Ellswork, 1972). More recently Cohn and his research group have demonstrated that the extraction of emotion from standard video streams is feasible according to recent research that uses facial tracking algorithms to extract action-unit movement and the corresponding underlying emotion (Cohn, Zlochower, Lien, and Kanade, 1999; Schmidt and Cohn, 2001). Therefore, the potential to extract this information from the IR video stream also exists if the resolution of the IR data is sufficient for the tracking algorithms.

Vetting Credibility Assessment Tests

Laboratory Studies

There is, at present, no unified theory of FDD. However, most researchers in the field assume that the physiological changes are caused, at least in pail, by emotions[2] such as fear and guilt (Dufek, 1970). Because of this, low accuracy rates obtained in the laboratory are often seen as a consequence of the lack of psychological stress experienced by study participants, relative to field condi-

[2] Some researchers have suggested that central nervous system measures such as fMRI measure brain activity associated with deception rather than its emotional correlates (Kozel, Padgett, and George, 2004).

tions in which examinees are suspected of committing actual crimes and the fear of incarceration and loss of personal freedom is usually present (Ginton, Daie, Elaad, and Ben-Shakhar, 1982). Laboratory studies are often designed to be non-threatening, and rewards are sometimes used as incentives for those participants who can pass a PDD examination (Kircher, 1984). Because of ethical concerns, it is very difficult to design a laboratory mock-crime study that can adequately generalize to the field (Barlaud and Raskin, 1975; Podlesney and Raskin, 1977, Kircher and Raskin, 1988). However, certain procedures such as the use of more motivated study participants and more realistic mock crime scenarios will result in physiological data that are more similar to those obtained under field conditions (Pollina, Dollins, Senter, Krapohl, and Ryan, 2004).

Field Vetting

One of the most difficult challenges facing credibility assessment researchers studying these techniques in actual field settings concerns the lack of standardization of either test construction or response scoring in current field practice. This often leads to discrepant results which make it difficult to obtain a preponderance of evidence for or against a particular test or measure. However, the lack of standardization is often not due to sloppiness on the part of researchers. As is the case with field studies in other areas of social/behavioral psychology, several useful techniques that are standard practice when conducting an experiment are simply not available to researchers in the credibility assessment field. Simple and easily reproducible stimuli (such as pure tones or standard emotion-evoking sounds or pictures) are not easily incorporated into traditional test formats, and are therefore seldom used. Random assignment to treatment group is not possible when field data are used, greatly complicating (or perhaps rendering impossible) the delineation of cause-and-effect relationships. Additionally, the nature of the interpersonal interactions involved in actual criminal investigations is extremely complicated and this makes the development of simple, highly predictive theories very difficult.

Studying the process of suspect interviewing is therefore not unlike studying the efficacy of clinical treatments by medical professionals (Crewson, 2001). In clinical trials, often there are several subpopulations of study volunteers who respond to treatments in various ways. Inclusion/exclusion criteria used in clinical trials are rarely perfect and it is often the case that some individuals assigned to the treatment group are not suffering from the medical condition being treated. Similarly, in suspect interviewing every case is different; each with a unique set of case facts and circumstances leading to the criminal act, and ground truth is almost never known with anything approaching certainty. This makes assignment to (deceptive/nondeceptive) group very difficult despite rigorous inclusion/exclusion criteria. There is another important similarity between clinical trials and credibility assessment research. In both cases, statistical tests conducted on group data, such as M/ANOVA, are often not as important as tests of the proportion of individuals correctly identified. Significant (and therefore presumably real) treatment effects with small effect sizes, while of scientific interest, are not practically useful in either field. Further, the definition of "practically useful" is a moving target. In credibility assessment, there are a variety of considerations that are relevant when considering the utility of each technology of interest - including cost, ease of use, and relative effectiveness.

Developing New Test Formats

In our experience, many researchers who decide to do work in credibility assessment have become experts in the use of a specific technology that they believe could be uniquely suited to the assessment of an individual's credibility, and recognize that it might have certain advantages over traditional polygraph. This is useful and has led to several new discoveries. However, many of these researchers are at a distinct disadvantage when it comes to devising mock-test scenarios to vet these new technologies. In our opinion, this is extremely unfortunate and speaks to the paucity of much-needed interdisciplinary research in the field of credibility assessment. For example, the engineers who are able to build new types of credibility assessment devices might benefit greatly from the help of a team of social psychologists with experience in designing experiments that simulate the conditions of an actual criminal investigation.

The "Avatar for Credibility Assessment" Software

The complicated nature of interpersonal interactions during the credibility assessment process has made it extremely challenging to study verbal and nonverbal exchanges in the laboratory setting. Nevertheless, we believe that it is necessary to do so if systematic advances are to be made to this process. One of the ways that researchers in our laboratory are studying these interactions is through the use of computer-generated (CG) three dimensional (3D) avatars that resemble humans as realistically as possible. The avatar program that was designed and created for our laboratory research is programmable to enable the user to define items such as avatars' spoken text, voice characteristics, facial features and facial expression changes. These avatars can also "understand" (using voice recognition software) the "yes" or "no" verbal responses that human examinees make to their own questions, and then respond with specific follow-on questions. In this way, relatively involved avatar-human interviews are possible.

Figure 2(a) shows the "base mesh" that the avatar software uses to create a new CG character, as well as renderings of specific CG character "meshes" created from the base mesh. This mesh is a mathematical description of a set of points in 3D virtual space.

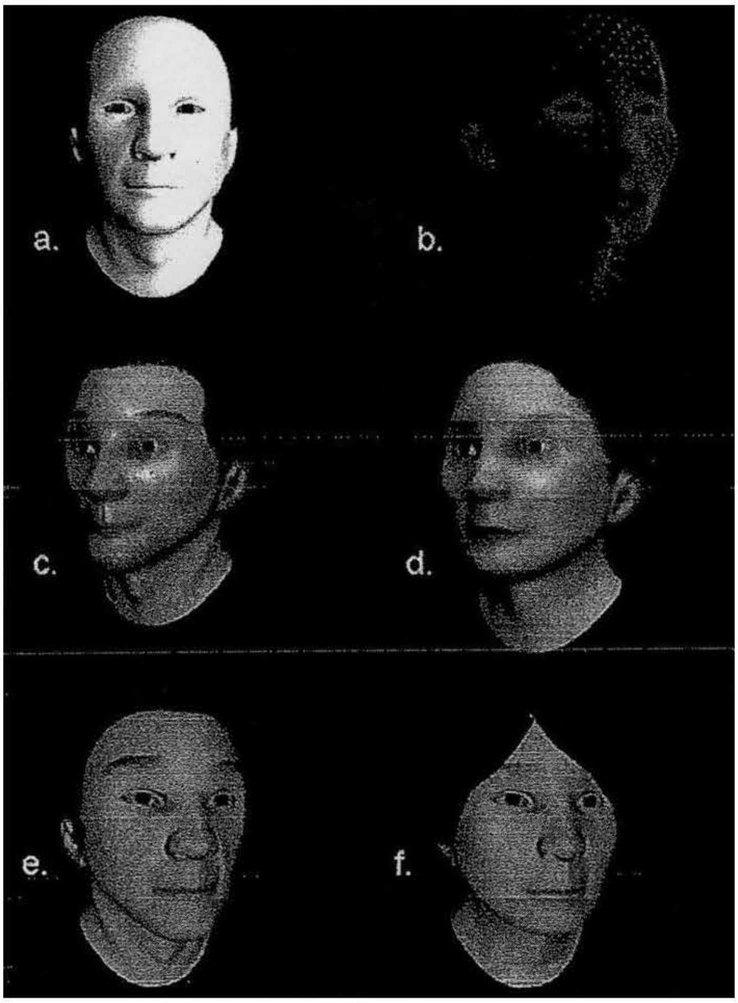


Figure 2. Template "base" mesh (a) from which all other avatars can be created. Wireframe (b) of an avatar morphed from the base mesh showing its vertex positions in virtual 3D space. Male (c) and female (d) avatars created with the retail version of Battelle's avatar software. Male (e) and female (f) avatars used during the DACA pilot study and beta testing.

Transformations of this mesh can then be performed to "morph" the base mesh into specific CG character meshes. The user can then configure the avatar with specific predefined textures, skin tint and hair, and voice. Although the number of character meshes that can be created within the program is virtually unlimited, the same base mesh is always used. In this way, the user can combine morphs of the base mesh with combinations of secondary features (e.g., hair and skin texture) in specific, mathematically definable ways to create a unique set of physical characteristics for each new character. This allows different users to reproduce the same character from the set of mathematical parameters saved in that character's file, as well as change specific parameters in systematic ways to create new characters. After a character is created, the user then has the option of choosing other features of the character's surroundings, such as distance from the CG character to the virtual camera, and background color.

A separate interview editor enables the user to configure a new interview script or to modify an existing one. Each interview is comprised of separate tracks which contain the audio files generated by the text-to-speech engine from text files created by the user. These audio files are passed through the program's lip sync subroutines to produce the avatar's spoken text when the interview is later executed. If the text within a track contains a question, branch points that determine the next track to be played are also specified. Voice recognition subroutines within the program are able to classify the interviewee's verbal responses during the interview. In this way, the avatar can respond to the interviewee's "yes' or 'no' answers to its own questions with specific follow-on questions to produce an automated interview between the CG character and a human interviewee.

The avatar is also capable of simulating specific human facial expressions of basic emotions. These expressions are created by morphing the mesh in specific regions of the face to simulate the human emotions of anger, fear, surprise, happiness, sadness, and disgust. Each emotional expression produced by the avatar is a combination of "action units" derived from Ekman's facial action coding system (Ekman and Friesen, 1978). In Ekman's system, action units represent the activity of specific facial muscles. In the Avatar software, action units are created using mesh morphs that were designed to resemble action units created by muscle activity as closely as possible. However, no attempt was made to simulate the underlying muscular physiology responsible for each action units because this would have been very computationally intensive and complicate the process of creating an interview. In the latest version of the software, the time course and intensity of each facial expression is set by the user from within the Interview Editor and becomes a part of the scripted interview.

Avatar Pilot Study

We conducted a preliminary test of a beta version of the Avatar software to determine the feasibility of using it for credibility assessment. Our primary research question concerned whether the software would be suitable for field use. We reasoned that one of the most valuable applications of this technology in the near term would be to automate the process of interviewing applicants for Federal security clearances. This process is normally very time consuming, labor intensive, and costly to the Federal Government. Typically, applicants fill out a security questionnaire that includes questions about their personal history, previous illegal activity, and foreign contacts. This information is then reviewed by a security officer in the presence of the applicant. This process takes time and also requires that the security officer and applicant be in the same place at the same time, which often leads to scheduling conflicts. Automating any portion of this process could therefore potentially save time and allow[7] human interviewers to use their time more effectively.

Several recent studies have shown that humans can respond socially to "computer-controlled entities" of various sorts (Gaggioli, Mantovani, Castelnuovo, Wiederhold, and Riva, 2003; Ku, et a!., 2005; Rizzo, Neumann, Enciso, Fidaleo, and Noh, 2001). However, prior to the present study it was not known how human interviewees would respond to being interviewed by a computer over several minutes, how well the humans would understand the avatar's speech (generated from the interview text), or how effective the software would be at responding to the human interviewees' verbal responses to its own questions. It was also important to obtain feedback from the interviewees about their experiences. Even if the mechanics of the process produced acceptable results, it was not known whether humans would make statements against self-interest to a computer or admit to any wrongdoing during the course of the interviews. If not, then although technically feasible, the project might be of little practical use.

Because so many of the avatar's features can be changed in systematic ways, a great deal of experimental control is possible and the effects of systematic changes in the avatar's appearance on human interviewees can be explored. In this study, we focused on gender effects. By investigating how male and female human interviewees respond to either a male or a female avatar conducting the same interview, it was hoped that new insights could be obtained concerning the attitudes and behaviors of male and female interviewees during a credibility assessment interview. Additionally, the use of computer-generated characters controlled for observer effects that can contaminate data when human interrogators are used. Essentially, this observer bias is created when law enforcement officers or other (human) interviewers change their behavior, either knowingly or unknowingly, because they are aware that they are being observed by an experimenter (Leo, 1996).

Methods

Participants. Thirty six participants (12 Female) between the ages of 18 and 42 (Mean = 24.2) were recruited from a sample of U.S. Army basic trainees stationed at Fort Jackson, South Carolina and assigned to duty at DACA. Participants' self-reported years of education ranged from 12 to 16 (Mean • 12.8). The percentage of female and male participants was based on the population of basic trainees at Fort Jackson selected by military personnel for assignment at the Defense Academy. Informed consent was obtained and documented for all participants.

Stimuli. The avatars used during this study were created using software designed specifically for use in automated credibility assessment interviews (Battelle, 2007; See Above). In an attempt to keep as many extraneous variables as constant as possible only two avatars were created for use in this study (Figure 2). The first used the generic "Asian Male" settings (base mesh with South Asian male weight = 1.0 and all others weight = 0.0; Male 'ITS with Pitch = 0.0 and Rate = 0.0) and the second used the generic "Asian Female" settings (base mesh with South Asian female weight = 1.0 and all others weight = 0.0; Female TTS with Pitch - 0.0 and Rate - 0.0).

Interview Script. A single script was used to conduct all computer-generated interviews during this study. Appendix 1 shows the text used to create the avatar's questions and statements during each interview. Each track was linked to previous and successive tracks using branch points. The flow of the interview was controlled in two ways. When the avatar produced a statement, the successive track was determined automatically by the program. When the avatar produced a question, the successive track was determined by the examinee's verbal response (Appendix 1).

Procedures. Each participant was assigned to either the "Male Avatar" or the "Female Avatar" condition. Group assignment was counterbalanced based on the participant's gender in batches of six. Prior to the interview, each participant filled out a questionnaire relating to security issues (questions 16-30 of the Standard Form 86, u.s. Office of Personnel management, 1995). Next, each participant was seated in a sound attenuating chamber, told that an interview with a computer-generated avatar would begin shortly, and asked to respond to each of the avatar's questions with either a 'yes' or 'no' answer as soon as the avatar had completed each question. Following the instructions, the interview was conducted according to the script in Appendix I via a computer monitor placed approximately 180 cm in front of the participant. At the completion of each of the avatar's questions speech recognition subroutines waited for an audible 'yes' or 'no' response from the participant. If the software did not register a response within 6 sec following the completion of the question, the track was repeated again until a response was detected. While the speech recognition subroutines were active, a small yellow question mark visible on the bottom left of the screen served as an additional visual cue that a response was expected. The interval between each of the avatar's statements and successive utterances was held constant at 2 sec.

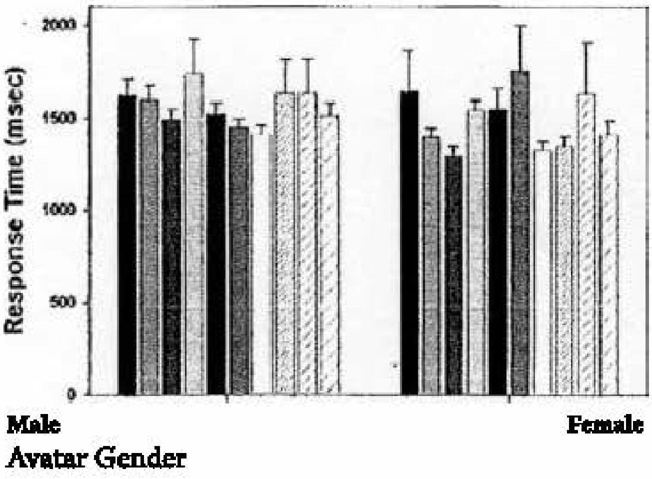
At the conclusion of the interview, each participant answered a series of questions about their experience, including the following: "Do you think the computer avatar was as effective as a human interviewer would be at conducting the interview? Did the avatar's questions make you feel any specific emotions during the interview? Did you purposely leave out any information when filling out the security clearance form? Did you purposely answer any of the avatar's questions incorrectly? How realistic did the avatar seem to you? Did you understand all of the avatar's statements?" If the participant admitted to prior illegal activity or security violations during the interview, they were also asked about these admissions during this debrief session.

Results

Questionnaire Data. Twenty-one of the 36 participants (58%) reported that they thought that the computer avatar was as effective as a human interviewer would be at conducting the interview, eight participants (22%) were not sure whether the avatar was as effective as a human, and seven participants (19%) believed the avatar to be less effective than a human would have been. Six participants (17%) reported feeling surprised at some point during the interview, two (6%) reported feeling disgusted, and one reported feelings of happiness. Two participants (6%) stated that they had purposely left out information when filling out the security clearance form, but there were no reports of any deliberate attempts to answer any of the avatar's questions incorrectly. Participants' (n = 36) ratings of how realistic the avatar seemed to them ranged from very (33%), to somewhat (64%) to not at all (3%). Thirty-four (94%) of 36 participants reported that they understood all of the avatar's statements.

Response Time Data. We used a stepwise linear regression procedure to determine the extent to which participant or avatar gender affected human interviewees' response time to the security questions asked by the avatar. We conducted three separate regression analyses. The first analysis used human gender (male humans in one group and female humans placed in a second group) as a dichotomous dependent variable, and the second used avatar gender (participants interviewed by the male avatar in one group and participants interviewed by the female avatar placed in the second group) as the dependent variable. The third analysis used four levels of dependent variable (Female Participant\Female Avatar; Female Participant\Male Avatar; Male Participant\Female Avatar; Male Participant\Male Avatar). In each of the analyses, participants' response times to the security questions were used as predictor variables, entered into the regression in blocks according to the category of question asked by the avatar. In the first block, response times for questions concerning mishandling of classified information were entered. This was followed by response times to questions concerning unauthorized foreign contact, use of recording or surveillance devices (Block 2), past illegal activity (Block 3), espionage, sabotage, or terrorist activity (Block 4), and willingness to answer questions asked by the avatar (Block 5).

Significant results were obtained in both the human gender and the avatar gender analyses. In the first (human gender) analysis, after step 3, with illegal activity in the equation, R - .23, $F_{Change} = 4.06$, $p < .04$. This significant finding was the result of greater mean response times to this question by female participants than males. The addition of variables within Blocks 4-5 did not reliably improve R^2. In the second (avatar gender) analysis, after step l, with the questions about mishandling of classified information in the equation, $R^2 = .17$, $F_{Change} = 7.14$, $p < .02$. This finding likely resulted from the trend, visible across responses to most of these questions, of decreased response times to questions asked by the female avatar (Figure 3).



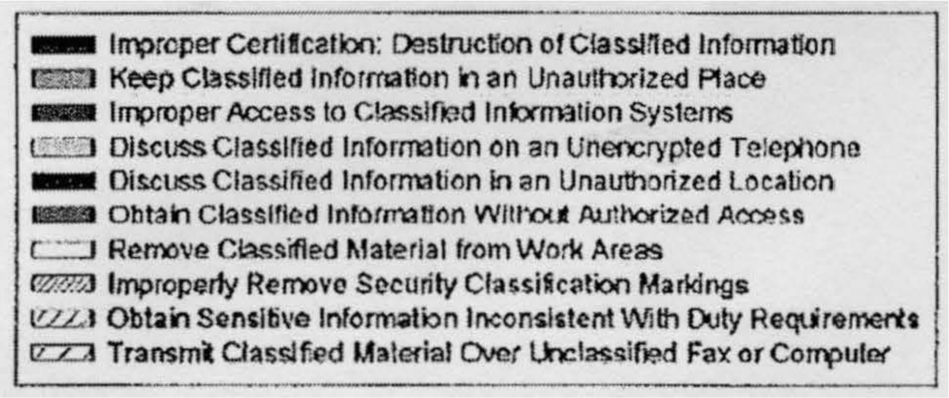


Figure 3. Mean (+S.E.M.) response times to security questions asked by a male (Group 1) or female (Group 2) avatar. Measurement of response time began at the end of the avatar's question and ended at the onset of the participant's verbal response.

The addition of variables within Blocks 2-5 did not reliably improve R^2. The third analysis, which examined interactions between human and avatar gender effects, failed to reach statistical significance even after all predictor variables were entered into the regression equation, suggesting that there were no interactions between human and avatar gender effects on response times.

Verbal Responses. Participants requested clarification regarding the meaning of specific questions asked by the avatar a total of 12 times during their interviews. In each case, the participant responded 'no' when the avatar asked them whether they understood what it meant when it asked them about classified information (n = 1), espionage (n = 2), illegal activity (n = 5), sabotage (n = 2), and committing a terrorist act (n = 2). Participants made admissions during their interviews a total of eight times. The majority of these admissions were to prior illegal acts (n = 4). One participant also admitted to keeping classified material at home or other unauthorized place, discussing classified information in an unauthorized location, removing classified information from work areas without authorization, and unauthorized use of listening devices in sensitive areas. A further discussion with the participant who admitted to questions about mishandling classified information revealed that he was a non-native English speaker, and that his answers were most likely due to his misunderstanding these questions. One participant's admission to prior illegal activity were verified during the debrief session as petty theft. Three other participants made admissions to prior illegal activity that were verified during the debrief session. Specific illegal acts admitted to at this time included possession/use of marijuana and driving under the influence.

Discussion

The results of this study suggest that computer avatars can effectively conduct interviews with humans. On several occasions, the human interviewees admitted to behaviors, such as prior illegal acts, that would be of interest to adjudicators conducting a background investigation for the process of granting a Federal security clearance. The majority of study participants reported that they thought the computer avatar was as effective as a human interviewer would have been at conducting the interview and almost all (94%) of the participants reported that they understood all of the avatar's statements. These findings support the continued study of avatars for use in credibility assessment interviews, with the eventual goal of using them in the field. There are at least three benefits to the use of computer avatars, including standardization of the interview process, precise synchronization of the interviewees' physiological responses to specific questions of interest, and mitigation of gender and cultural biases that might exist when humans conduct the interviews. In the experimental context, the use of avatars also allows for precise manipulation of variables of interest while at the same time controlling for factors extraneous to the research questions being studied.

Another research question in this study concerned the effects of interviewer gender on the response times of interviewees. Researchers in several previous studies failed to report significant gender effects on the likelihood of successful interrogation outcomes (Leo, 1996; Reed, 1999), and so it was interesting to examine whether any gender differences, even subtle effects, could be documented in the credibility assessment context. Effects of both interviewee and interviewer gender on response times were found in this study. The interviewee effect was due to significantly longer mean response times to the question about prior illegal activity for female participants. The interviewer effect appeared to be due to longer response times for all study participants when the male avatar asked the security-related questions. These effects were quite small and replication will be necessary before the findings can be considered verified. However, we argue that the use of avatars reduces variability inherent in the use of human interviewers in studies of this type, and might explain why small but nevertheless real gender effects were obtained in this study.

Conclusion

Although several new technologies are showing promise in the field of credibility assessment, the most studied physiological measures for this purpose are the cardiovascular, electrodermal, and respiratory responses recorded using the polygraph. The challenge for credibility assessment researchers in the coming years will be to improve the usefulness of new technologies in field settings. Most likely, a "one size fits all" approach will be less effective than one that works to the strengths of a particular test or technology. For example, thermal imaging and laser Doppler vibrometry both have the advantage of being non-contact, with no inherent limitations on the length of a recording session. FMRI and brainwave technologies, as central nervous system measures, hold out the promise of being direct measures of deception, though this has not yet been conclusively shown. Whichever technology one chooses to explore for the purpose of credibility assessment, appropriate psychological tests will have to be developed. Again, each test should be optimized for the technology being developed. In this chapter, work on the use of computer-generated avatars to interview humans as part of the U.S. federal security clearance process was presented. It seems clear that the magnitude of the problems faced by credibility assessment researchers, and the variety of the skills necessary for the development of an effective, field-usable technology necessitate an interdisciplinary approach. At a minimum, this will include the expertise of engineers, computer scientists, physiologists, and psychologists.

Acknowledgments

The authors would like to thank Betty Dillon and Rose Swinford of the DACA Research Staff for their assistance with data collection procedures. This project was funded by the Defense Academy for Credibility Assessment as project number DoDPI06-P-0020. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Defense or the U.S. Government. The United States Government retains a royalty-free, nonexclusive and irrevocable license to reproduce, translate, publish, and use this manuscript (32 CFR 32.36).

Appendix 1: Avatar Interview Script

Track Type	Track No.	Track Text	Next Track	Previous Track
S	1	Hello. Thank you for agreeing to talk with me today. I would like to ask you some questions about the questionnaire that you filled out earlier today.	Next: 2	
Q	2	Will you answer my questions?	Yes: 3; No: 30	1
Q	3	Have you ever had unauthorized contact with an individual who is known or suspected of being associated with a foreign intelligence, security, or terrorist organization?	Yes: 31; No: 4	2
Q	4	A past history of illegal activity could make a person susceptible to blackmail. Have you ever engaged in any illegal activity that might make you susceptible to committing a security violation?	Yes: 32; No: 5	3,31
Q	5	Do you understand what I mean when I ask you about illegal activity?	Yes: 7; No: 6	4
S	6	Illegal activity is any activity that is prohibited by law in the United States.	Next: 48	5
Q	7	Have you ever read, or discussed classified information in an unauthorized location?	Yes: 33; No: 8	5,32,48
Q	8	Do you understand what I mean when I ask you about classified information?	Yes: 10; No: 9	7
S	9	Classified information is any information that is available to authorized persons only, for reasons of national security.	Next: 49	8
Q	10	Have you ever attempted to obtain classified information for which you do not have authorized access or an official "need to know" this information?	Yes: 34; No: I t	8,33,49
Q	11	Have you ever asked other people for their signatures certifying that classified information was destroyed when these people did not actually observe the destruction?	Yes: 35; No: 12	10,34
Q	12	Did you ever use unauthorized cameras, recording devices, computers, or modems in areas where classified information is stored, discussed, or processed?	Yes; 36; No: 13	11,35
Q	13	Did you ever <partofsp part = "Verb"> use </partofsp> unauthorized listening or surveillance devices in sensitive or secure areas?	Yes: 37; No: 14	12,36
Q	14	Did you ever keep classified material at home or any other unauthorized place?	Yes: 38; No: 15	13,37
Q	15	Did you ever attempt to acquire access to classified information systems and computers without proper authorization?	Yes: 39; No: 16	14,38
0	16	Did you ever transmit classified material over unclassified FAX or computer?	Yes: 40; No: 17	15,39
Q	17	Did you ever try to obtain access to sensitive information that is inconsistent with your present duty requirements?	Yes: 41; No: 18	16,40

Q	18	Did you ever remove classified material from your work areas without appropriate authorization?	Yes: 42; No: 19	17,41
Q	19	Did you ever improperly remove security classification markings from documents?	Yes: 43; No: 20	18,42
Q	20	Did you ever discuss classified information on a non-secure, unencrypted telephone?	Yes: 44; No: 21	19,43
Q	21	Have you ever committed espionage against the United States?	Yes: 45; No: 22	20,44
Q	22	Do you understand what I mean when I ask you about espionage against the United States?	Yes: 24; No: 23	21
s	23	Espionage means spying, or using other people as spies, to obtain secret information about the United States Government or its interests.	Next: 50	22
Q	24	Have you ever committed sabotage against the United States?	Yes: 46; No: 25	22,45,50
Q	25	Do you understand what 1 mean when i ask you about sabotage against the United States?	Yes: 27; No: 26	24
S	26	Sabotage means destruction of United States Government property or obstruction of United States Government Operations.	Next: 51	25
Q	27	Have you ever committed a terrorist act against the United States?	Yes: 47; No: 28	25,46,51
Q	28	Do you understand what I mean when I ask you about, committing a terrorist act against the United States?	Yes: 30; No: 29	27
s	29	Terrorism is the unlawful use of force or violence against people or property with the intention of intimidating individuals or governments.	Next: 52	28
s	30	Thank you for coming. The interview is now over. The experimenter will arrive to talk to you shortly.	Next:	2,28,47,5 2
s	31-47	The experimenter will ask you about your response to this question after the Interview.	Next: Next Question	"Yes" answered Security Violation
Q	48	A past history of illegal activity could make a person susceptible to blackmail. Have you ever engaged in any illegal activity that might make you susceptible to committing a security violation?	Yes: 32; No: 7	6
Q	49	Have you ever read, or discussed, classified information in an unauthorized location?	Yes: 33; No: 10	9
Q	50	Have you ever committed espionage against the United States?	Yes: 45; No: 2<)	23
Q	51	Have you ever committed sabotage against the United States?	Yes: 46; No: 27	26
Q	52	Have you ever committed a terrorist act against the United States?	Yes: 47; No: 30	29

Note. Track Type S = Statement; Q = Question.

References

Alder K. (2007), *The lie detectors: The history of an American obsession*, New York: Free Press.

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

Applied Physics Laboratory (1993), *Polygraph automated scoring system*, Laurel, MD: The Johns Hopkins University.

Backster C. (1963), *Standardized polygraph notepack and technique guide: Backster zone comparison technique*, New York: Backster School of Lie Detection.

Barland G.H. and Raskin D.C. (1975), *An evaluation of field techniques in detection of deception*, Psychophysiology, 12, 321–330.

Battelle Inc. (2007), *Operator's manual for avatar for credibility assessment implementation*, Columbus, Ohio: Author.

Ben-Shakhar G., Bar-Hillel M. and Lieblich I. (1986), *Trial by polygraph: Scientific and juridical issues in lie detection*, Behavioral Science and the Law, 4, 459–479.

Crewson P. (2001), *Comparative analysis of polygraph with other screening and diagnostic tools*, Department of Defense Polygraph Institute. Ft. Jackson, SC. DoDIT01-R-0003.

Cohen N.J. and Eichenbaum II. (1993), *Memory, amnesia, and the hippocampal system*, Cambridge, MA: MIT Press.

Cohn J.F., Zlochower A., Lien J., Kanade T. (1999), *Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding*, Psychophysiology, 36, 35–43.

Darwin C. and Ekman P. (1997), *The expression of the emotions in man and animals*, Oxford: Oxford University Press (Original work published 1872).

Department of Defense Polygraph Institute (2006), Meeting of the DoDPI Scientific Review Committee (Interna! Memorandum). Fort Jackson, SC: Author.

Dufek M. (1970), *Emocje a polygraf [Emotion and the polygraph]*, Prokuratura, 3, 103–106.

Ekman P. (1992), *Telling lies: Clues to deceit in the marketplace, politics, and marriage*, New York: W. W. Norton and Company.

Ekman P. and Friesen W. (1978), *The facial action coding system*, Palo Alto: Consulting Psychologists Press.

Ekman P., Friesen W. and Ancoli S. (1980), *Facial signs of emotional experience*, Journal of Personality and Social Psychology, 39, 1125–1134.

Ekman P., Friesen W. and Ellsworth P. (1972), *Emotion in the human face*, New York: Elmsford.

Gaggioli A., Mantovani F., Castelnuovo G., Wiederhold B. and Riva G. (2003), *Avatars in clinical psychology: A framework for the clinical use of virtual humans*, Cyberpsychology and Behavior, 6, 117–124.

Ganis G. and Kosslyn S.M. (2007), *FMRI studies of different types of deception*, Psychophysiology, 44 (Supplement 1), S4–S5.

Ganis G., Kosslyn S.M., Stose S., Thompson W.L. and Yurgelun-Todd D.A. (2003), *Neural correlates of different types of deception: An fMRI investigation*, Cerebral Cortex, 13, 830–836.

Garbey M., Sun R., Merla A., Pavlidis I. (2005), *Contact-free measurement of cardiac pulse based on analysis of thermal imagery*, Abstracts of the 22nd Annual Houston Conference on Biomedical Engineering Research, Houston, TX. Feb. 10–11.

Ginton A., Daie N., Elaad E. and Ben-Shakhar G. (1982), *A method for evaluating the use of the polygraph in a real-life situation*, Journal of Applied Psychology, 67, 131–137.

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

Keeler L. (1930), ***A method for detecting deception***, The American Journal of Police Science, 1, 38–51.

Kircher J.C. (1984), ***Uses and abuses of the mock crime paradigm in research on field polygraph techniques***, Psychophysiology, 21, 566.

Kircher J.C. and Raskin D.C. (1988), ***Human versus computerized evaluations of polygraph data in a laboratory setting***, Journal of Applied Psychology, 73, 291–302.

Kozel F., Johnson K.A., Mu Q., Grenesko E.L., Laken S.J. and George M.S. (2005), ***Detecting Deception Using Functional Magnetic Resonance Imaging***, Biological Psychiatry, 58, 605–613.

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

Ku J., Jang H.J., Kim K.U., Kim J.H., Park S.H., Lee J.H. et al. (2005), ***Experimental results of affective valence and arousal to avatar's facial expressions***, Cyberpsychology and Behavior, 8, 493–503.

Langleben D.D., Loughead J.W. and Bilker W.B. (2005), ***Telling truth from lie in individual subjects with fast event-related fMRI***, Human Brain Mapping, 26, 262–272.

Larson J.A. (1923), ***The cardio-pneumo-psychogram in deception***, Journal of Experimental Psychology, 6, 420–454.

Leo R.A. (1996), ***Inside the interrogation room***, Journal of Criminal Law and Criminology, 86, 266–303.

Lombroso C. (1887), ***L'Homme Criminel; Criminel-Né, Fou Moral, Épileptique; Étude Anthropologique et médico-légale***, Paris: Alcan.

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

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

Masip J., Sporer S.L., Garrido E. and Herrero C. (2005), *The detection of deception with the reality monitoring approach: a review of the empirical evidence*, Psychology, Crime and Law, 11, 99–122.

Murthy R., Pavlidis I. (2005), *Non-contact monitoring of breathing function using infrared imaging. Technical Report UH-CS-05-09*, Computer Science Department, University of Houston. Accessed via web at: www.cbl.uh.edu/~pavlidis/index.html.

Murthy R., Pavlidis I., Tsiamyrtziz P. (2005), *Touchless monitoring of breath function*, Abstracts of the 22nd Annual Houston Conference on Biomedical Engineering Research, Houston, TX. Feb. 10–11.

National Research Council (2003), *The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. Division of behavioral and social sciences and education*, Washington, D.C: The National Academies Press.

Pavlidis I., Eberhardt N.L. and Levine J. (2002), *Human behavior: Seeing through the face of deception*, Nature, 415, 35.

Podlesney J.A. and Raskin D.C. (1977), *Physiological measures and the detection of deception*, Psychological Bulletin, 84, 782–799.

Pollina D.A., Dollins A.B., Senter S.M., Brown T.E., Pavlidis I., Levine J.A. and Ryan A.H. (2006), *Facial skin surface temperature changes during a concealed information test*, Annals of Biomedical Engineering, 34, 1182–1189.

Pollina D.A., Dollins A.B., Senter S.M., Krapohl D.J. and Ryan A.H. (2004), *Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations*, Journal of Applied Psychology, 89, 1099–1105.

Pollina D.A., Ryan A. (2002), *The relationship between facial skin surface temperature reactivity and traditional polygraph measures used in the psychophysiological detection of deception: A preliminary investigation*, U.S. Department of Defense Polygraph Institute. Ft. Jackson, SC. DoDPI02-R-0007.

Pollina D.A. and Squires N.K. (1998), *Many-valued logic and event-related potentials*, Brain and Language, 63, 321–345.

Reed S. (1999), *Effect of demographic variables on psychophysiological detection of deception examination outcome accuracies*, Polygraph, 28, 310–331.

Reid J.E. (1947), *A revised questioning technique in lie-detection tests*, Journal of Criminal Law and Criminology, 37, 542–547.

Reid J.E. and Inbau F.E. (1977), *Truth and deception*, Baltimore: Williams and Wilkins.

Rizzo A.A., Neumann U., Enciso R., Fidaleo D. and Noh J.Y. (2001), *Performance-driven facial animation: Basic research on human judgments of emotional state in facial avatars*, Cyberpsychology and Behavior, 4, 471–487.

Rohrbaugh J.W., Sirevaag E.J. and Ryan A.H. (2006), *The physiology of stress and emotion: Remote sensing using laser Doppler vibrometry*, Paper presented at the Association for Psychological Science, New York, NY. May 25–26.

Rosenfeld J.P., Soskins M., Bosh G. and Ryan A. (2004), *Simple, effective countermeasures to P300-based tests of detection of concealed information*, Psychophysiology, 41, 205–219.

Ryan J.D., Althoff R.R., Whitlow S. and Cohen N.J. (2000), *Amnesia is a deficit in relational memory*, Psychological Science, 11 (6), 454–461.

Saxe R., Carey S. and Kanwisher N. (2004), *Understanding other minds: Linking developmental psychology and functional neuroimaging*, Annual Review of Psychology, 55, 87–124.

Schmidt K.L. and Cohn J.F. (2001), *Human facial expressions as adaptations: Evolutionary questions in facial expression research*, Yearbook of Physical Anthropology, 44, 3–24.

Sun M., Garbey M., Merla A., Pavlidis I. (2005), *Imaging the cardiovascular pulse*, Paper presented at the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, C.A. June 20–25.

Trovillo P.V. (1939), *A history of lie detection*, Journal of Criminal Law and Criminology, 29, 848–881, 30, 104–119.

Trovillo P.V. (1942), *Deception test criteria: How one can determine truth and falsehood from polygraph records*, American Journal of Police Science, 33, 338–358.

Verschuere B. (2007), *Slow down and start sweating: Reaction-times and skin conductance for the detection of concealed information*, Psychophysiology, 44 (Supplement 1), S4–S5.