# EUROPEAN POLYGRAPH

**PUBLISHED QUARTERLY**

Volume 3 2009 Number 3–4 (9–10)


Andrzej Frycz Modrzewski Krakow University

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

# Contents



## Book reviews

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 at-

tached 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/l), 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 includ-

ed 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 discemable 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 inde-

pendent 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.

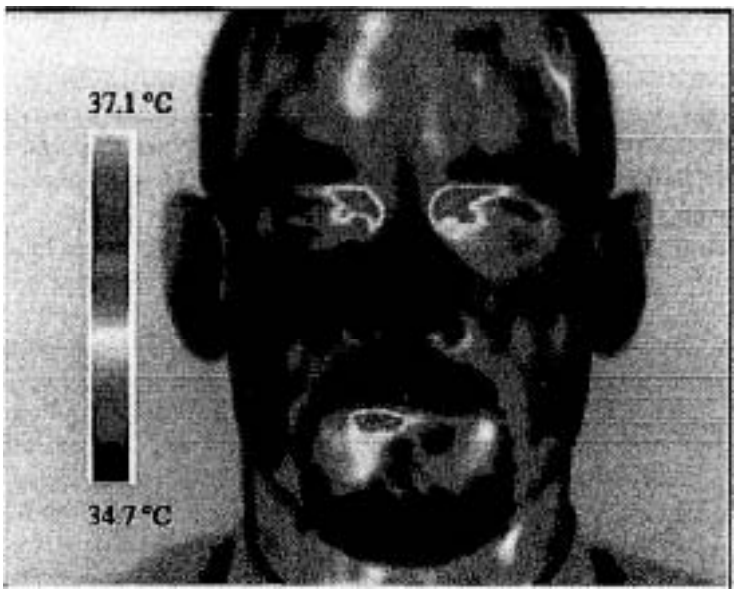



Figure 1. Thermal image showing the temperature distribution across a human face. The periorbital regions are typically hotter than other facial areas, even during resting conditions.

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 knowl-

edge 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 non-contact 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 de-

spite 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.

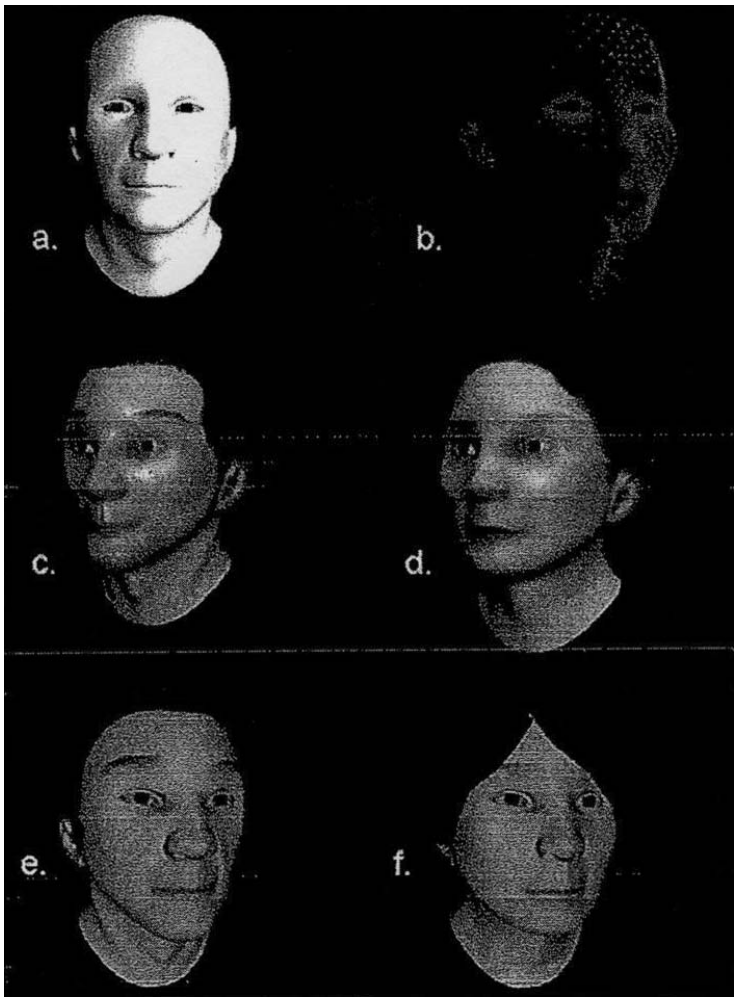



Figure 2. Template "base" mesh (a) from which all other avatars can be created. Wire-frame (b) of an avatar morphed from the base mesh showing its vertex positions in virtual 3D space. Male (c) and female (d) avatars created with the retail version of Battelle's avatar software. Male (e) and female (f)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

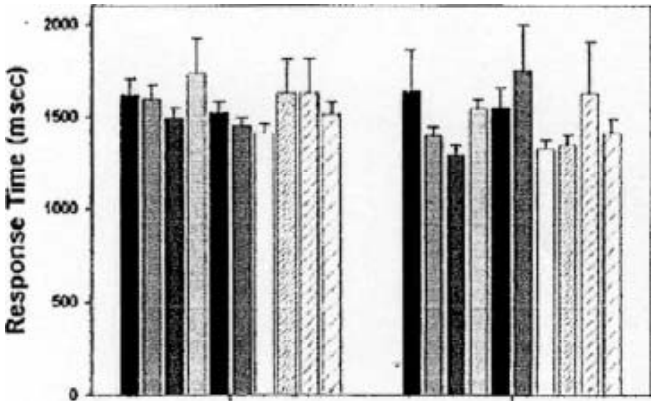
## Results

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).



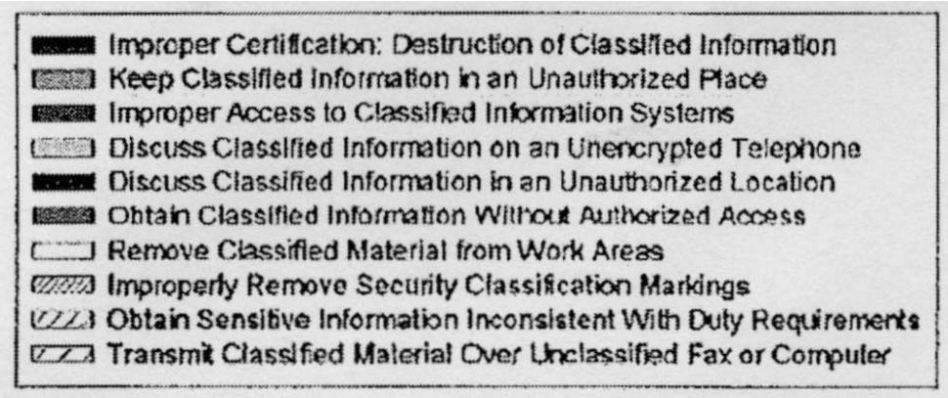



Figure 3. Mean (+S.E.M.) response times to security questions asked by a male (Group 1) or female (Group 2) avatar. Measurement of response time began at the end of the avatar's question and ended at the onset of the participant's verbal response.

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 us-

ing 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 DoDPl 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.

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Tuvia Shurany*
[uczelnia]
James Allan Matte**
[uczelnia]
Einat Stein***
[uczelnia]

# Influence of case facts on blind scorers of polygraph tests

The first author (Shurany) was asked by a client if the Quality Control Reviewer should be provided with the facts of the case before blind scoring the physiological data recorded on the polygraph charts related to that polygraph examination, and whether it would have an influence on the reviewer's judgment and evaluation. The question generated this research study, which was preceded by a field study by Dror & Rosenthal (2008) that employed meta-analytic procedures to determine the degree of reliability and bias ability of forensic experts. The sample sizes in Dror et al's study involving fingerprint identification were

* e-mail
** e-mail
*** e-mail

quite small, some with only five and six experts. However, its findings revealed a biasing effect that increased the proportion of incorrect
judgments from 30% to 70%. Because fingerprint analysis is so well established, it is expected that the results of Dror et al's study would apply equally to other forensic

## Method

### Participants and procedure

Three groups of polygraph examiners were selected for this study comprising a total of 82 polygraphists; from Mexico (n=35), United States (n=34) and Bulgaria (n=13). All Polygraphists were skilled and experienced examiners. They were instructed to evaluate and score the polygraph charts presented to them via PowerPoint slides in the manner in which they were trained, using the 3-position scale.

Commencing in the early 1970's, the Quality Control Review Section of the Department of the Army Criminal Investigation Division Command (CID) which reviewed all polygraph examinations conducted by the CID world-wide, used a 3-position scale to score the polygraph charts and whenever the reviewer's score disagreed with the score of the original polygraphist, the 7-position scoring scale was used to arrive at a final determination. (Brisentine, 1974, 1995). Several research studies have been conducted and reported regarding the effectiveness of the 3-position versus the 7-position scoring scales in polygraph chart analysis. Blackwell's (1999) field study found "the PDD examiners mean level of accuracy was 75.7% and 66.3% for the 7- and 3-position scoring scales, respectively." Blackwell stated that "without exception, the overall level of accuracy generated by the examiners when using the 7-position scoring scale was higher than when using the 3-position scoring scale. The same was true when looking at the overall percentages for either the innocent examinations or the guilty examinations." Krapohl (1998) found that the 3-position scale with a cutoff (fixed threshold) of +/–4 was statistically equivalent to the widely accepted 7-position scale with the +/–6 cutoff score (fixed threshold). However, Krapohl also found that "the highly experienced raters in this study rarely used the full range of available values in the 7-position scale, employing the narrower range of the 3-position scale for about 90% of the question comparisons." Capps and Ansley (1992) and Van Herk (1991), like Krapohl, found that the accuracy of the 7- and 3-position scales depended on the threshold used. The Backster and Matte Comparison Techniques (Matte, Backster 2007) use

an *increasing* threshold, whereas other Zone Comparison Technique modifications (DoDPI, Utah) employ a fixed threshold. (Matte, Backster 2000; Matte 2002).

The charts were selected from cases that used the Backster Zone Comparison Technique. All the Polygraphists were presented with 8 separate examinations each containing 4 charts. The first and last polygraph examination consisted of the same confirmed No Deception Indicated (NDI) charts. The second through the seventh examination consisted of unconfirmed NDI and DI (Deception Indicated) results which were used as a buffer to prevent recognition of examination #8 as being the same as examination #1.

Before the presentation of examination #1, no case facts were given to the polygraphists. Conversely, immediately prior to the presentation of examination #2 through #7, they were given fictitious facts of the case. Thus, they knew these facts prior to their evaluation and scoring of the charts. At examination #8, the polygraphists were presented with fictitious case facts and information supposedly provided by the police that fingerprints of the examinee found at the scene of the crime matched with 90% probability.

The first group of 35 polygraphists from Mexico evaluated examinations #1 through #8 on the same day. One of them asked the first author (Shurany) whether they had seen the charts of examination #8 before. He was instructed to simply evaluate the charts without providing him an answer to his question, and all of the polygraphists were able to hear the question and answer.

The above mentioned procedure was applied with another group of 34 polygraphists from the USA and 13 polygraphists from Bulgaria, with the exception that examinations #1 through #4 were presented on one day and examinations #5 through #8 were presented 48 hours later. This change in procedure was necessitated by the experience of the first group where one of the polygraphists recognized examination #8 as being identical to examination #1.

The score sheets from examinations #1 and #8 from all participants were entered into an Excel sheet. The names of the polygraphists were omitted from the score sheets which were identified only by a random number. The score sheets of examination #1 were compared with the score sheets of examination #8.

## Results

In order to examine whether there are differences between the Charts' scores with prior knowledge vs. without prior knowledge, and in order to explore whether these differences are similar among the Polygraphists from the different countries, a 2 (knowledge) x 3 (country) MANOVA with repeated measures on knowledge (without vs. with) was performed.

The Manova revealed significant differences between the Polygraphists' scores to the same chart with vs. without prior knowledge ($F_{(2, 78)} = 17.9$, $p<.001$, $Eta^2=.32$).

Univariate ANOVAs revealed that the scores following previous knowledge were significantly lower as compared to the scores of the same chart without previous knowledge(see Table 1), for each compared question (Q5, Q7).

For Q5 the scores decreased by 37% on average, with a maximum decrease of 8 points, and for Q7 the average decrease was 26%, with a decrease of as much as 10 points in one case.

Table 1. Means and Standard Deviations of Charts' scores with vs. without prior knowledge

| | Prior Knowledge | | | | | |
|---|---|---|---|---|---|---|
| | Without | | With | | | |
| Measures | M | SD | M | SD | $F_{(1,79)}$ | $Eta^2$ |
| Q5 | 3.44 | 2.73 | 2.17 | 2.60 | 27.24*** | .26 |
| Q7 | 5.04 | 2.56 | 3.70 | 3.15 | 19.90*** | .20 |

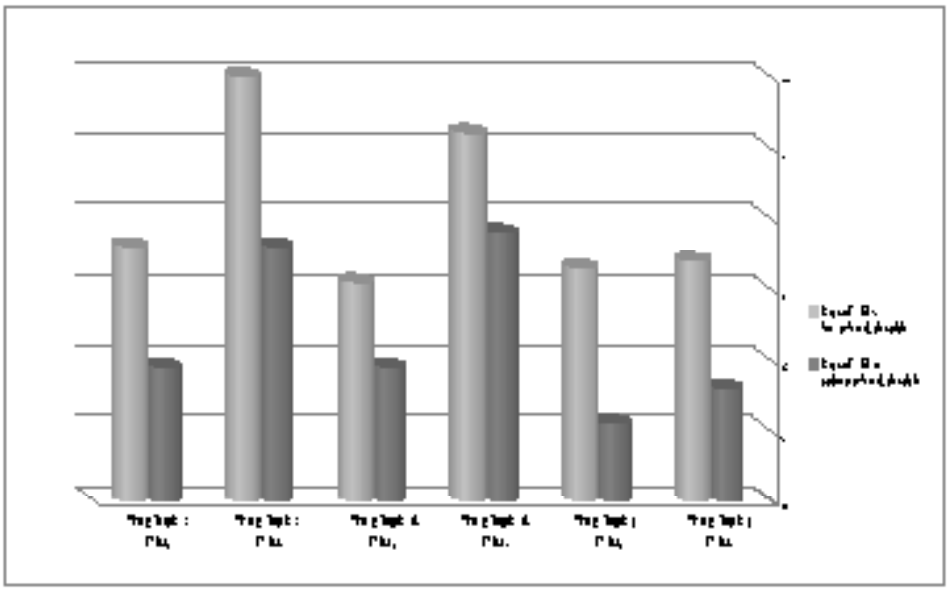
****p<.001*

No interaction of knowledge X country was detected ($F_{(4,156)} = 0.62$, $p>.05$), i.e. the decrease in scores with vs. without prior knowledge was found to be similar among the polygraphist from the different countries (see Figurer 1)

Furthermore the significant decrease in chart's score following prior knowledge produced 9 (11%) false positives as compared to one false positive (1.2%) with no prior knowledge. Generally, the rate of inaccurate decisions (i.e. false Positives or inconclusive, to a confirmed no deception indicated charts) was doubled as a result of prior knowledge (65.84% as compared to 32.9% respec-

tively). McNemar test for comparing frequencies in repeated measures with dichotomic variables (accurate/inaccurate decision) revealed that this difference in inaccuracy rate is significant ($c^2 = 25.04$, $p<.001$).

Figure 1. Average Score Change with vs. without prior knowledge



## Conclusion

The results of this study indicate that knowledge of case facts does have an influence on the polygraphist's evaluation and scoring of the physiological data recorded on polygraph charts. Therefore a Quality Control Review should be conducted in three stages:

The authors wish to express their gratitude to Dr. Frank Horvath for his contribution of research information and statistical data analysis and his invaluable advice on the conduct of this study.
The authors also wish to dedicate this study to the memory of Dr. Ronald M. Reuss and William J. Scheve, Jr. disciplines including forensic psychophysiology. However, another study by Eitan Eladd, et al (1998), that used blind scoring of polygraph records, reflected a different result. Prior expectations affected the polygraphists' judgments when the polygraph charts reflected inconclusive results which offered no indications of guilt or innocence.
However when the objective physiological evidence reflected strong indications of guilt or innocence which clearly contradicted the polygraphist's expectations, judgments were not affected by these expectations. Hence these results indicate that the effect of prior expectations on the interpretation and scoring of polygraph charts is limited to inconclusive polygraph records.

1. Blind scoring should be conducted without any knowledge of the case facts, examiner's impression, and results of the polygraph examination.
2. After blind scoring has been accomplished, the polygraphist should then review the case facts for adequacy of test question formulation.
3. Review the audio/video recording for assurance that no procedural violations were committed by the original polygraphist.

## References

Blackwell N.J. (1999), *PolyScore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examinations from actual criminal investigations*, Polygraph, 28 (2), 149–175.

Brisentine R.A. (1995, Nov. 13), Telephone conversation with J.A. Matte.

Brisentine R.A. (1974, March), *Polygraph Research in the U. A. Army*, Polygraph, 3 (1), 66–80.

Capps M.H., Ansley N. (1992), *Anomalies. The contributions of the cardio, pneumo, and electrodermal measures towards a valid conclusion*, Polygraph, 21 (4), 321–340.

Dror I., Rosenthal R. (Jul 2008), *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, Journal of Forensic Science, 53 (4), 900–903.

Elaad E., Ginton A., Ben-Shakhar B. (1998), *The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiners' Decisions*, Polygraph, 27 (4), 288–301.

Krapohl D.J. (1998), *A Comparison of 3- and 7-position Scoring Scales with Laboratory Data*, Polygraph, 27 (3), 210–218.

Matte J.A. (2002), *2002 Supplement to Forensic Psychophysiology Using The Polygraph*, J.A.M. Publications: Williamsville, New York.

Matte J.A., Backster C. (2000), *A Critical Analysis of Amsel's Comparative Study of the Exclusive v. Nonexclusive Comparison Question*, Polygraph, 29 (3), 261–266.

Matte J.A., Backster C. (2007), *Psychological Structure and Theoretical Concept of the Backster Zone Comparison Technique and the Quadri-Track Zone Comparison Technique*, Polygraph, 36 (2), 84–90.

Van Herk M. (1991), *Numerical evaluation: Seven point scale +/–6 and possible alternatives; A discussion*, Polygraph, 20 (2), 70–79.

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Svetoslav Zanev*
[uczelnia]

# The History of Polygraph Use in Bulgaria

The polygraph and polygraph methodology came relatively late to Bulgaria. It has a history of about 40 years, since the first polygraph instrument – a six-channel Stoelting – arrived in 1968. This instrument was for the needs of Bulgarian intelligence to develop a system for training in deceiving the polygraph. It was necessary for intelligence operations for the infiltration of agents in the NATO bloc and the former Yugoslavia (at this time considered, although not an enemy, not a friend either). Later, in 1972, another Stoelting Ultrascribe was received. After the creation of a laboratory (and later institute) of psychology at the Ministry of the Interior, all polygraph experiments were conducted there. Initially, different information was collected by written sources (books, articles, documents) and by stories of people who were tested by polygraph behind the Iron Curtain. This leads to the conclusion that the main focus of experiments for deceiving the polygraph had to be use of pharmacological substances. All experiments in the early 1970s were carried out in an attempt to search for the best combination of drugs that might be easy to access and used

* e-mail

to decrease physiological reactions in polygraph examination. Based on later information from participants in this training, the results had variable success. The key person in this period was Dr. G. Boyanov, a medical doctor and the head of the group which developed the programme.

In this period, the polygraph was used in just a few criminal cases. This use was generally as an experiment, and not for verification of truth or finding new information for investigation. In these experiments, Dr. Boyanov and colleagues tried to developed a test format that might be used in criminal investigations. Having access to theoretical sources for different testing formats, they developed a local version of the Relevant-Irrelevant test format. This format was used in examinations with multiple issues. In addition, another format used is POT. Without any obvious reason, zone techniques were considered solely for the needs of intelligence and were practically not used in criminal investigations.

During this period, a VSA was also imported from the USA. Many experiments were conducted with the same purpose – modifying behaviour by pharmacological substances. The experiments were cancelled in the early 1980s, for various reasons.

Up to 1989 the mainstream of development of the polygraph was creation of techniques for countermeasures. Training in the use of polygraph examinations was based on books and local experiments, and a small number of people was able to work with the polygraph. This was critical for the first years of the 1990s, because these people retired, leaving no successors to continue their work. In practice, the polygraph was used in this period in a few criminal cases and for selection of officers for a new service for fighting organized crime. Old instruments ceased to work (they can now be seen in the Museum of the Ministry of the Interior) and the Director of the Institute of Psychology (at the time Dr. Boyko Ganchevski) decided to buy new ones.

In May 2007, Paul Redden, then senior polygraph examiner at the San Diego police department and representative of Lafayette Instruments, arrived at the institute. Mr. Redden taught a short course in the Backster zone technique and became the first person to give "first-hand" information on the contemporary use of the polygraph in investigations in the USA. After this visit, the author of this article had the opportunity to complete a full course for polygraph examiners at the Backster School of Lie Detection in San Diego. This was the first formal training of a Bulgarian polygraph examiner at an APA-accredited

school. After that, two other polygraph examiners attended the same school. All this catalyzed the process of finding an appropriate place for polygraph examinations. After 1997, use of the polygraph increased considerably. In the following years polygraph examination became decisive in the resolution of many criminal cases – murder, serial assaults, robbery and burglary. From 1999, the first results from polygraph examinations were presented before court. They were presented as "psychological expertise for the investigation of truthfulness". This is the only legal way to introduce polygraph examination in the court system. This expertise might be accepted or rejected by the court. The key issue is coherence with other evidence as described in the Bulgarian criminal code. In order to be presented to court, examinations must be made by a criminal investigator, prosecutor or the court itself.

In 1998 the Bulgarian Polygraph Association (BPA) was established. It was declared to be a professional body of people who work in the field of polygraph examinations. In order to be admitted to the association a candidate must have training and practice and accept the code. We decided to accept the ethical code and rules from the APA in order to guarantee good practice. The BPA now has 14 members working for government and in private practice (about 50–50 represented in the association). The Association has an Executive Board and Supervisory Board. The key task of the latter is to consult and resolve difficult cases in evaluation and administration of polygraph examination if an examiner or a customer needs. In 2004, the BPA organized its first international conference. This conference, which had a major influence on public opinion, took place under the patronage of Boyko Borisov (then General Secretary of the Ministry of the Interior, now Prime Minister of Bulgaria) and hosted guests from the USA, Russia and Israel. The conference generated fruitful discussion on the regulation of polygraph use in the court system leading to greater acceptance of polygraph examination results. The Association maintains very good contacts with polygraph examiners from different points of the world. On a regular basis, we organize seminars with Tuvia Shurany, a private examiner from Israel.

As mentioned above, polygraph examinations are now conducted for the government (at the Institute of Psychology in the Ministry of the Interior) and in private companies (most of which were established by former officers from the Institute).

Since 1996, in the Institute of Psychology, polygraph examinations have been situated in the department for "Psychological expertise and support of crimi-

nal investigation". Along with polygraph examinations, this department also works in criminal profiling, prevention of suicides and hostage negotiations. The key assignment of the department is investigation of murder, robbery and burglary cases. Polygraph examinations from 1997 until today have played a role in resolving a number of murder cases, which has led to great popularity and acceptance of the results of the polygraph. From 1998 to 2008 the head of this department was Dr. Nedelcho Stoychev. Today, Dr. Stoychev is Director of the Institute of Psychology at the Ministry of the Interior.

In 2008, following a restructuring programme of all activity in fighting corruption and organized crime, all group members of special units had to pass a polygraph examination. Members of these units are prosecutors, investigators and police officers.

Private polygraph examiners work mainly in cases of burglary and periodic loyalty testing. Both the government and the private sector conduct more than 1000 examinations per year.

In both sectors the main methodology used is the zone technique of Cleve Backster – You phase, Exploratory, SKY. In addition, different formats of POT are used. Some modifications have been made in comparison questions in order to make them easy to understand for the examinee and to fit the cultural and national specifics of Bulgaria. In addition, the Matte quadri-track ZST format is exploited, along with some modifications of RI in screening procedures.

Meanwhile, books have been published aimed at popularizing the use of the polygraph: by Svetoslav Zanev – *Polygraph or "Truth Machine"*; Boyko Ganchevski – *Use of the Polygraph in Criminal Investigation*; Radostin Belenski – *Lie Detector*.

Polygraph examiners use different instruments, but all are computerized versions of Lafayette Instrument and Stoelting products and polygraphs made in Russia.

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Vitas Saldžiūnas*
VIP Security Department
Under the Ministry of the Interior
Vilnius, LITHUANIA
Aleksandras Kovalenko**
Police Department
Under the Ministry of the Interior
Vilnius, LITHUANIA
Kalin Gaidarov***
BULGARIA

# The Problems of Truth Perception during Psychophysiological Examination

The topic of lies is discussed in the works of Paul Ekman (1992) and Aldert Vrij (2000). The specific case of the Russian nation is described by V. Znakov. Unfortunately, the problem of truth during psychophysiological examination has been researched by very few scholars. The majority of authors refer to validity and accuracy. No discussion has taken place about how such truth (the facts) is established during psychophysiological examination.

* vitas.saldziunas@vad.lt
** aleksandr.kovalenko@policija.lt
*** kdarov@abv.bg

First of all, we should recall that all the tests used for psychophysiological examinations can be divided into two groups. The first group, i.e. the group of comparative questions, is the most popular one among polygraphists. Comparative questions are such questions which are not related to the problem researched by plot, i.e. which are irrelevant. Through such psychophysiological examination those people who took part in the crime under investigation are distinguished from those who did not (Varlamov et al. 2008). According to the classical technique of psychophysiological testing such an examination should be followed by a post-test interview (Abrams, 1989; Matte, 1997), which aims at making the examined persons (those who were noticed as having stronger psychophysiological reactions to the relevant questions in comparison to the comparative ones during psychophysiological examination with the application of the polygraph) confess to having committed the crime. So far we do not see any direct problem of truth.

The second test group may include all the tests which help to verify and establish the circumstances of a given crime. This group includes POT, GKT, CIT and EKT. So the question is what is established when applying these tests. We will present several examples from the specific psychophysiological examination with the application of a polygraph.

Several years ago the Russian polygraphist Irina Nikolajeva applied a polygraph test when examining a person who was suspected of having murdered his wife. Nikolajeva used the modified GKT. The recorded psychophysiological reactions of the person allowed the assumption to be made that the person was completely familiar with the circumstances of the crime. However, there were no strong (marked) psychophysiological reactions to the question whether he had killed his wife.

At first Nikolajeva did not understand the reason for this discrepancy. Later it emerged that according to the examinee's understanding and belief he did not kill, but "punished" his wife.

The following example is different. A large amount of copper tubes was stolen in a company. The investigators (detectives) established four potential or possible suspects. All suspects agreed to take a psychophysiological examination with the application of a polygraph. During the psychophysiological examination using the polygraph EKT was used. Having conducted the psychophysiological examination with the application of the polygraph it was established that two of the suspects had not participated in the crime. Taking into consid-

eration the results of the examination the assumption could have been made that the other two suspects were well-informed or knew how the copper tubes had been stolen. One question was aimed at finding out what sum of money the criminal had received when selling the tubes. Having conducted the psychophysiological examination of the other two suspects with the application of the polygraph the results gained revealed inadequacies in the amount of sums, as the psychophysiological reaction of the examinees to different sums of money differed. The question might have arisen whether this was a mistake; however, this was not the case. One of the suspects, i.e. the organizer, told his accomplice the sum which they had received for the copper; this sum was, however, smaller than the real one.

The Codes of Criminal Procedure of the Republic of Bulgaria and of the Republic of Lithuania similarly define that the court, the prosecutor's office and subjects of the pre-trial investigation have to establish the objective truth using all possible legal and legislative measures. This implies that the investigators and prosecutors who investigate crimes and polygraphists who participate in criminal proceedings have to approach real events, to define the accident, to restore the course of the accident, and to define the body of the crime using the maximum defined evidence and other data. This means that, when defining the objective truth, two principles clash and disagree. On the one hand there are principles of "objective" truth, and on the other there are principles of proven "subjective" truth (Gaidarov, 2007).

What is "objective" and "subjective" truth? The problem is that the circumstances of the crime are established via the subject, who one way or another participated in that crime. How accurately can the subject recall past events? At this point we are not talking about a purposeful lie or reticence. There might be "facts", which are "proven", but in reality they did not occur in the past or they were different.

In classic criminology textbooks (Kertes, 1964) we may find examples of cases when wrong perception evokes mistakes. When asked how a prisoner had escaped, the warder explained that during a walk the prisoner had attacked him with a knife. During the investigation it was established that the criminal was holding not a knife, but a herring.
It is known that one and the same thing may be perceived differently by different people. Even the same person under different conditions may evaluate the same thing differently. This fact should be taken into consideration when interpreting the evidence of witnesses. Two witnesses' testimony regarding the accident often vary due to a different attitude towards the accident.

Research conducted by psychologists, psychiatrists and other physicians (Kavalieris et al., 2009) show that the situations described above, in which the eyewitnesses could not reconstruct the facts, happened due to three reasons:

Firstly, the things perceived by victims and by quite a few witnesses during the commitment of the crime may raise more or less negative emotions. For the sake of self-preservation the human psyche blocks information which raises negative emotions in the conscious memory. Therefore, many victims and eyewitnesses cannot recollect information which they have consciously perceived.

Secondly, witnesses (and sometimes victims) quite often take no notice of the things they have seen; therefore, the perceived facts do not stay in the memory as knowingly perceived and cannot be recollected and narrated.

Thirdly, sometimes the number of objects, features and actions related to the crime go beyond human perception. According to scientists, the number of objects that can be knowingly perceived is 5 +– 2 objects; the surplus is stored outside the knowingly perceived part of the memory.

According to the acclaimed specialist in the field of NLP (neuro-linguistic programming) R. Dilts (1998), it can be claimed that when acquiring information a person may be subject to three possible processes: generalization, deletion and distortion.

Generalization is a cogitative operation and product, the form of reflection of general features and attributes of actual phenomena.

Deletion (filtration) is a process by which a person "sorts" information, taking into consideration which information is important and necessary to him and which is not.

Distortion is a mechanism which changes the perception of sensory data.

Let us sum up what internal and external factors influence a person's ability to perceive the processes which take place around him/her, and to reproduce them from his/her memory.
1. Individualization of events. It is known that memories are constantly being "processed". The information which is stored in memory is influenced by subjective perception, evaluation and fantasies. Therefore, there is no doubt that

the available information will be distorted. The least that may happen is deformation of initial information. This can be illustrated by the following example. Information about different events may be deformed, subjectively outlived and as a result merged into one event.

2. Physical state and abilities of an individual. An individual may be in such a state when, due to some functional or organic changes, his/her capabilities of accepting and perceiving information decrease or become limited. The point is that an individual may be tired, intoxicated with alcohol or narcotic substances, may be ill or may feel unwell, may fail to fully perceive a prompt action due to the particularity of the organism. Also the process of accepting and perceiving information may be aggravated by the poor eyesight and hearing of an individual.

3. Emotional state of an individual. The capabilities of an individual to accept information are limited (Varlamov et al., 2000; Mijovic, 2002). If there is too much information, mechanisms of self-preservation automatically come into action. This may cause partial or full blocking of information acceptance. For example, this may happen when an individual is in a state of stress or euphoria etc.

4. Particularity of an individual due to age. Preteens taking part in an investigation of a crime can be taken as an example. Forensic scientists in Bulgaria have drawn attention to the fact that it was noticed that preteens' evidence was very unreliable. Children constantly change their evidence, supplement evidence with information or forget details, confuse reality with fantasies. Analogical problems may arise when interviewing people of a considerable age.

5. Sex, profession and education of an individual. The best example illustrating this factor would be the following. Women are better at noticing and remembering the details of female clothes and accessories, whereas men are better at remembering vehicles, work tools, mathematicians at figures, and other specialists at details of a machine, materials etc.

6. External conditions of an event. It is important to take into consideration lighting and weather conditions during the event and the observation position of an individual during the action.

When a participant in a crime takes a psychophysiological test with the application of polygraph it is examined whether that participant is hiding some information. It often happens that during psychophysiological examination using a polygraph due to lack of expertise in this sphere clients/initiators hope that the examination will reveal the whole objective truth of the event. Unfortunately, during this type of examination it is only possible to verify the subjective information which an individual has acquired, retained and restored.

Due to the aforementioned objective and subjective reasons the results (or information) may not exactly coincide with the results of other specialists, examinations and other proceedings. The results of several individuals who have taken a psychophysiological examination with the application of a polygraph may also partly differ. Therefore, the investigators of crimes should analyze discrepancies and look for the reasons for those discrepancies.

Our experience working with EKT (Saldțiűnas and Kovalenko, 2008, 2009 ) shows that with the benefit of properly assessed remarks which were put forward in this work it is possible to receive reliable and predictable results. We would like to remind experienced polygraphists of well-known recommendations:
– to conduct polygraph test on persons who are at least 16 years old,
– if an individual has taken a psychophysiological examination with the application of a polygraph while under the strong influence of alcohol or other intoxicants, there is a small chance of receiving comprehensive results,
– to conduct the examination with the application of a polygraph as soon as possible after the crime is committed,
– in some cases to induce hypnosis in order to help to sharpen an individual's memory (Kavalieris et al., 2009).

## References

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

Dilts R.B. (1998), *Modeling with NLP*, Capitola: Meta Publications.

Ekman P. (1992), *Telling Lies*, New York: Norton and Co.

Gaidarov K. (2007), *Судебно-психологическое установление правды*, Российский полиграф, 3 [text in Russian].

Kavalieris A. and Makans L. (2009), *Usage of hypnosis in interviewing – criminalistics recommendations for law enforcement structures of Latvia, Criminalistics and examination: science, studies, practice*, Vilnius, 182–187.

Kertes I. (1964), *Тактика и психологические основы допроса*, Москва: Юридическая литература [text in Russian].

Matte J.A. (1997), *Forensic Psychophysiology Using The Polygraph. Scientific Truth Verification – Lie Detection*, Williamsville, New York: Matte Polygraph Service.

Mijovic D. (2002), *Concept of temporary unsuitability of examinees in the use of polygraph methods*, Polygraph, 31 (4), 274–292.

Saldžiūnas V. and Kovalenko A. (2008), *The event knowledge test*, European Polygraph, 1 (3), 21–29.

Saldžiūnas V. and Kovalenko A. (2008), *The event knowledge test (EKT) in Polygraph Examination (in case murder)*, European Polygraph, 2 (4), 137–142.

Saldžiűnas V. and Kovalenko A. (2008), *The event knowledge test (EKT) in Polygraph Examination (common notice of tactics)*, European Polygraph, 3–4 (5–6), 209–220.

Saldțiűnas V. and Kovalenko A. (2009), *Problems of questions in event knowledge tests*, European Polygraph, 2 (8), 69–75.

Varlamov V.A. and Varlamov G.V. (2000), *Психофизиология полиграфных проверок* (Psychophysiology of PV Examination), Краснодар: Просвещение-Юг [text in Russian].

Varlamov V.A., Varlamov G.V. & Trufanov A.A. (2008), *Приемы получения правдивой информации*, Москва: Лига-Принт [text in Russian].

Vrij A. (2000), *Detecting lies and deceit*, John Wiley & Sons.

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Edward Lewandowski*
[uczelnia]
Łukasz Lewandowski**
[uczelnia]

# Forensic examination of memory traces Part 1

What is a memory trace? An ideal trace, or just the opposite, a trace on which it is easiest to find some flaws? Does it exist in a material manner, and can it be examined in an objective way, or is it nonmaterial and ephemeral, with both its observation and assessment as subjective as human perception? We are not trying to answer these questions here; we are interested in the way in which the fact that everyone in our country has for many years agreed that the object of the polygraphic examination are memory traces practically influences or should influence the manner in which such examinations are conducted. The approach that we present diverges significantly from the majority of publications in this area of forensic studies, which is why we consider it important to provide it with its own name: forensic examination of memory traces.

* e-mail
** e-mail

As has been mentioned before, we on purpose ignore here the question of memory traces themselves, moving immediately to the problem of disclosing them.

How is a memory trace disclosed in criminal acts? In most cases, this happens during interrogation. Testifying depends on the will of the suspect; he must neither be forced nor encouraged to do so. The suspect is not obliged to tell the truth. Yet if he feels unconnected with the subject of the investigation, it may be in his interest to disclose all the information that he possesses and that may be useful for investigation of the case. The obligations of a witness include testifying in accordance with their best knowledge and concealing no significant detail, even though we can of course not assume that that is what always happens. It may therefore be supposed that the person who has testified or given explanations in the matter in question has delivered a statement that he possesses or does not possess specific memory traces connected to the case (relevant for the case).

Article 192a of the Criminal Procedures Code envisages that – with the consent of the examinee – polygraph examinations may be conducted to define the value of the evidence of the disclosed traces (of a crime, event – authors' note). This also the case, therefore, with memory traces.

## How can we examine a memory trace and register it?

Functional magnetic resonance imaging (fMRI) is a technique of imaging the brain introduced in the 1980s, with publications concerning it being available since the mid-1990s. Examinations make use of the fact that while the examined person performs thinking tasks, for instance, the activity of certain areas of the brain increases. Detection of these areas of the brain is possible thanks to the fact that the neuronal activity (e.g. thinking) causes topical changes of blood flow, blood volume, and oxygen consumption. During the examination, a magnetic field is run through the patient's head. Passing radio waves through the head of the examinee results in their reflection from hydrogen particles, which resonance is picked up by the scanner. It can be said that this examination is based on finding areas of increased uniformity of the magnetic field.

During over a decade of conducting examinations (using fMRI and PET, i.e. positron emission tomography), it was found which regions of the brain are responsible for storing information, thinking, emotions, etc. An activation map of the brain was developed, although research has not yet been completed.

The basic examination procedure is based on scanning the area in question (in layers) repeatedly/several times in the rest state, and later after the examined

has been subjected to a specific stimulus. It is important that the geometric centres of the selected layers lie in the axis of the magnet.

Neuropsychological examinations disclose whether the given person "tells the truth or lies" on the basis of the areas of the brain showing activity. When the area of the brain responsible for problem-solving is active, it may be concluded that the examinee is lying, and is making up the answer. If the active areas are those responsible for storing information (and emotions), this testifies that the examinee recollects the facts that were actually recorded in the brain earlier.

Experiments conducted during the research (with simultaneous observation of changes taking place in the brain and changes occurring on the surface of the body of the examined person) disclosed that activation of the area of the brain responsible for storing emotions is manifested by stimulation of the autonomous nervous system. In an indirect manner, using a polygraph to achieve this goal, we may observe when the area of the brain responsible for storing information and emotion becomes activated. It was on this that we based our concept of disclosing memory and emotional traces during examination.

During further research that we conducted, we used a stimulus on the patient (as in the case of fMRI), which results in the activation of specific structures in the brain (usually responsible for storing information and emotions, like in fMRI), and observed and registered the effect of the activity of these brain structures which manifests itself in physiological processes that can be measured in a non-invasive manner on the surface of the body of the examinee. To be more precise, if we are asked a question concerning an event during an examination, the content of the question will be compared to the content of our memory. The content of the question will reach the thalamus through the organ of hearing, with the initial assessment of the features of the stimulus performed from the thalamus and the content of the question passed to the higher parts of the brain, where precise analysis of the question is performed and its content is compared with the resources stored. From the many levels of processing the information, the result of the query is passed to the amygdale. If specific records are found in our memory, emotions accompanying this record are also found, which is manifested in stimulation of the autonomous nervous system.

Peak of tension tests are used to define the evidential value of memory traces. It is significant that – unlike the functioning practice and theory of conducting polygraph examinations, and, moreover, unlike the approach that assumes the sole use of "guilty knowledge tests" – "our" concept assumes the necessity of using three types of peak of tension tests, depending on the needs and capacity of their implementation, without excluding the use of control question tests. It is obviously not by coincidence that peak of tension tests play such an impor-

tant role in this concept. It is only thanks to their application that we can claim unambiguously that the goal of the examination is disclosure and registration of memory traces of the event in question.

Presenting forensic concepts of examining memory traces, it seems important to emphasise the differences that led us to discuss here a new approach to conducting polygraph examinations. First, whether the person examined conceals the knowledge of the event is a matter of secondary, if not tertiary importance. Yet we do not pass entirely over this matter, as we realise that, in most people, lying is also a source of emotions. Secondly, in the approaches to these examinations, the primary role has so far been attributed to control question tests, and peak of tension tests were given an ancillary role. It is otherwise in forensic examination of memory traces: peak of tension tests play a primary role, while control question tests have a supportive character, as they are used mainly for selecting people and the directions of the examination using peak of tension tests. Last but not least, this concept requires greater care and precision in preparing and conducting the examination, and at the stage of interpreting the registered emotional changes.

As we mentioned before, there are three types of peak of tension tests, each of them having a different reason for using it as well as diagnostic value. Below, we will present both a brief description of each of these tests and examples of their use. We shall put special emphasis on the discussion and illustration of confirmation tests.

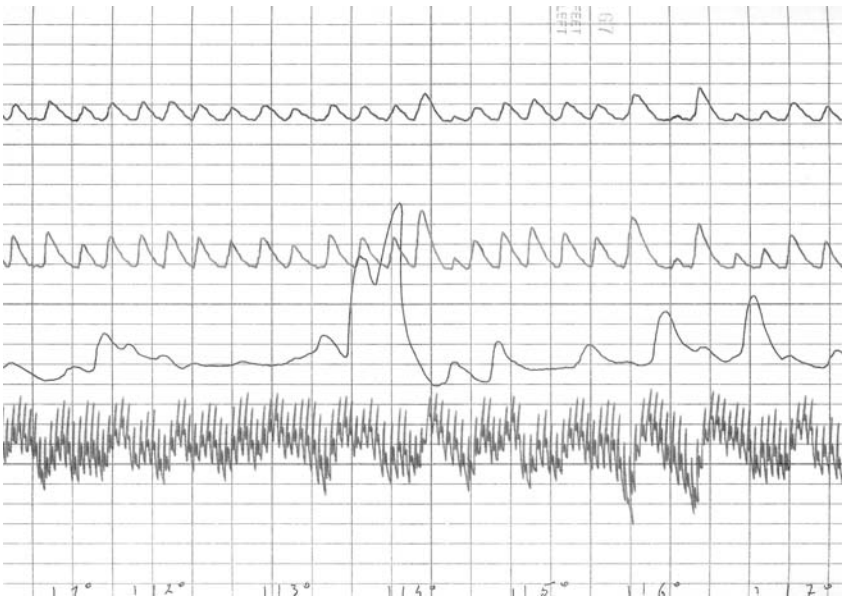
The guilty knowledge test is constructed in such a manner that a single critical question concerning a significant detail of the crime is hidden among other, similar questions. It is assumed that only a participant in the event in question knows which of the questions is the critical one. For the person who did not participate in the event, all of them should be equally critical. The objective of such a test is to check whether the examinee knows the details of the event, which are highly improbable to be known to anyone else but the perpetrator.

The test with a solution that is unknown and sought makes it possible to define significant details of the event that were not previously known to the investigators. These details should, on the other hand, be known to the person who committed the action that is the object of the investigation or to the person who acquired such knowledge from other sources, e.g. from the perpetrator or witness. To give an example, using this test, one may try to identify the perpetrator or accomplice in a crime, the place where the objects coming from the crime or used to commit it are kept, the place where the body was deposited, or the place where a kidnapped person is being held.

The confirmation test makes it possible to establish significant details of an event that raised doubts that may result among others from the mutually ex-

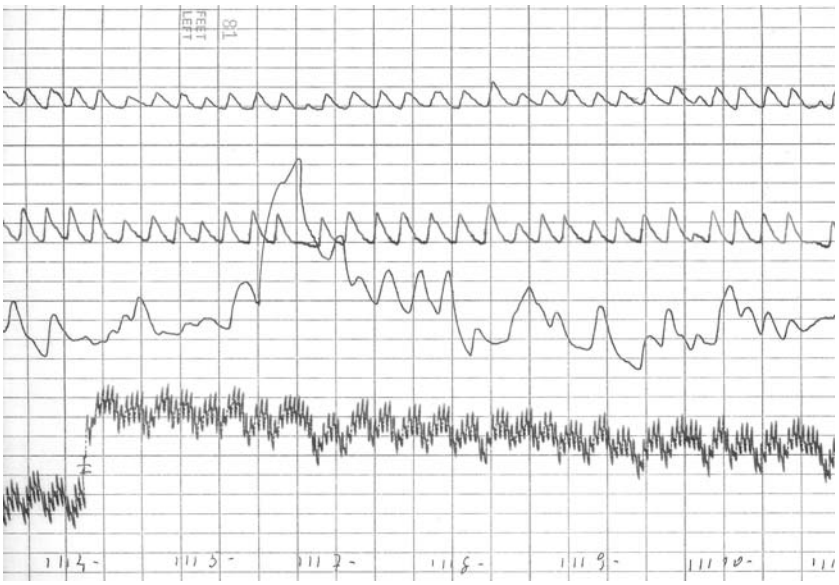
clusive testimonies of persons connected to the investigation. In such a test, the examined person reacts emotionally to the question concerning a fact that actually took place or its date, according to the person's knowledge, independent of the fact or date stated during the interrogation. In this type of test, both the version that assumes the participation of the examinee in the given event and the one that excludes it are taken into account. For example, this may be the fact of having been at the place of the event, as opposed to the place stated by the examinee during the interrogation, with the assumption that neither of these options has strong support from other evidence. Without resorting to the theory of probability, we assume that the likelihood is minimal that a given person 1a) reacts weakly or does not react emotionally to the fact connected to the question being investigated that at the same time places that person at the scene of the event, and 1b) also reacts in a strong and legible manner to the question considering the place that the person stated, with the 2) fact of that person's brain having registered a memory trace of being in the place where the crime was committed on the given day and at the given time.
Significant is the fact that the registered trace is lasting, and it is possible (where necessary) to repeat its "examination".

Case study No. 1.
The following are registered (top to bottom of the recording): changes in the thoracic breathing cycle, in the abdominal breathing cycle, galvanic skin re-

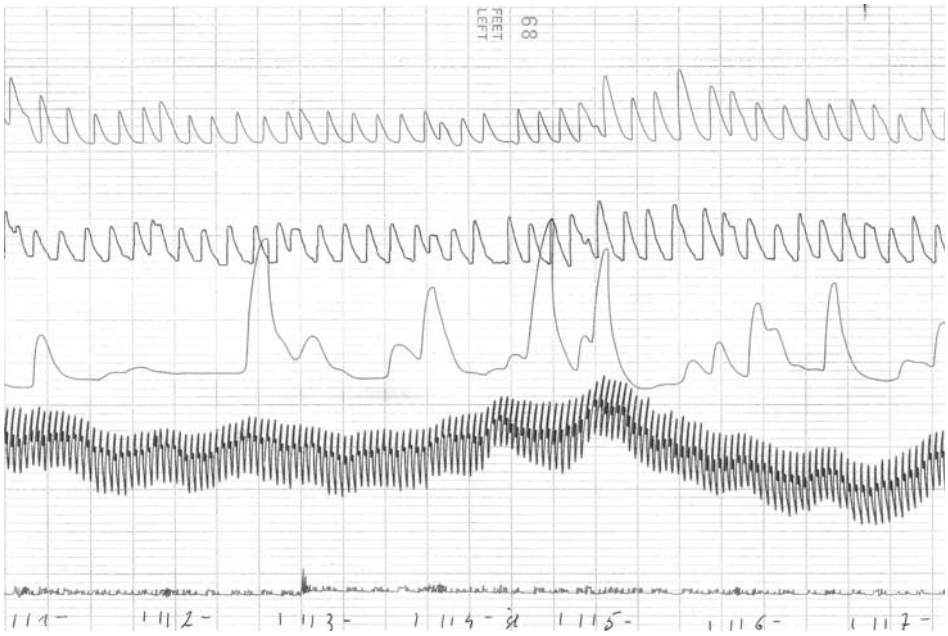


sponse (in the conventional, i.e. traditional, polygraph, the recording of changes in the galvanic skin response "outruns" the point at which a question is asked by a few seconds, which results from the fact that the galvanic skin response pen is longer than the others by 3/4"), changes in pulse and blood pressure. The distance between two vertical lines is 1/2", and the speed of paper transmission is 1" per 10 seconds.
The examination used the guilty knowledge test. The questions concerned the amount of money stolen from a day safe in a hotel. Question No. 4 concerns the amount that was actually stolen. Visible are marked emotional changes in the skin galvanic response and in breathing after question No. 4 was asked. Earlier, the examinee declared that he had no knowledge concerning the amount stolen. This gives grounds to assume that the examinee knows what amount of money was stolen from the safe.

Case study No. 2.
The examination makes use of the test with an unknown result. In this test, the names of four different people known to the examinee were mentioned, with an accompanying question whether he knew that any of the people mentioned performed the action that was the object of the investigation. Visible emotional changes in skin galvanic response and in breathing followed the asking of question No. 7. This gives grounds to assume that the examinee knows or suspects that the person mentioned in question No. 7 is connected to the theft being investigated.



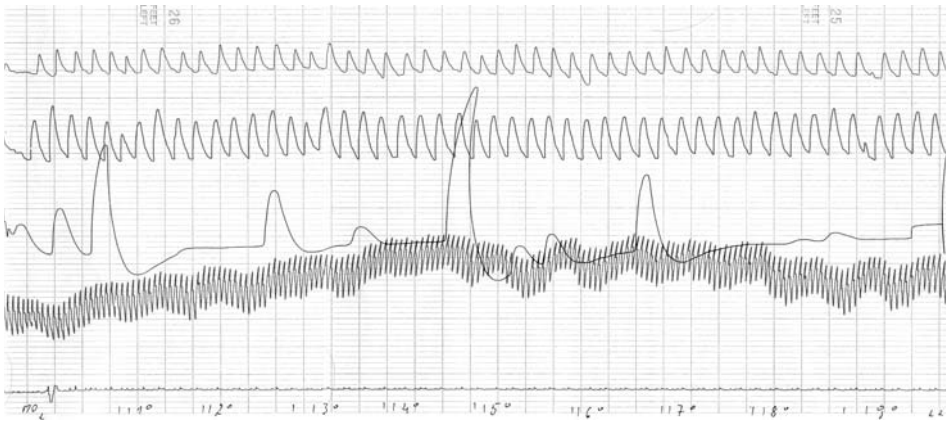
Case study No. 3.
The examination made use of the confirmation test. A few months after the crime, while browsing photographs in the police register, the injured party recognised the examinee as the perpetrator of a robbery. He recognised him, among other reasons, thanks to his characteristic moustache. In the courtroom, the accused stated that he had shaved off the moustache two years before the robbery he was charged with was committed. He remembered this very well, as the reason for removing the facial hair was the issue of an arrest warrant in which he was identified as having the moustache. He shaved the moustache off to make recognition more difficult. Question No. 5 concerns the period when the accused (according to his statement) did not have a moustache. Question No. 7 concerns the period when the robbery which the examined was accused of was committed.
The largest emotional changes in the arterial blood pressure and the skin galvanic response followed the asking of question No. 5, which is to be interpreted as meaning that the examinee – according to his knowledge – did not have a moustache during the period listed in question No. 5.
This examination establishes what memory and emotional trace concerning the time when the examinee did not have a moustache is registered in his nervous system. This trace is lasting, and may be disclosed in successive examinations, should the need to repeat them arise.



Case study No. 4.
Adverse circumstances, such as a lengthy passage of time and repeated participation in legal proceedings, do not rule out the possibility of conducting forensic examination of memory traces. It should nevertheless be remembered that repeated interrogations and questionings concerning the actions the examinee was charged with, presentation of the charges, bringing the case to court, participation in the trial, and sentencing at the court of first instance also result in the creation of memory and emotional traces connected to the event in question. For these reasons, it may not be ruled out that in examination of certain people, sudden emotional changes present after asking critical questions may be caused by the experiences listed above and not because of perpetrating the action they are charged with.
This example concerns the brutal killing of an elderly married couple. The examination was conducted more than 13 years after the crime had been committed. Polygraph examination was proposed to the people who had the keys to the house or who could have been let inside at any time. This was suggested by the way that the crime was committed. During the investigation, the time of the killing was defined fairly precisely. All those examined were interrogated as witnesses within several hours of the discovery of the crime. Among others, they described where they were and what they were doing at the time of the crime.

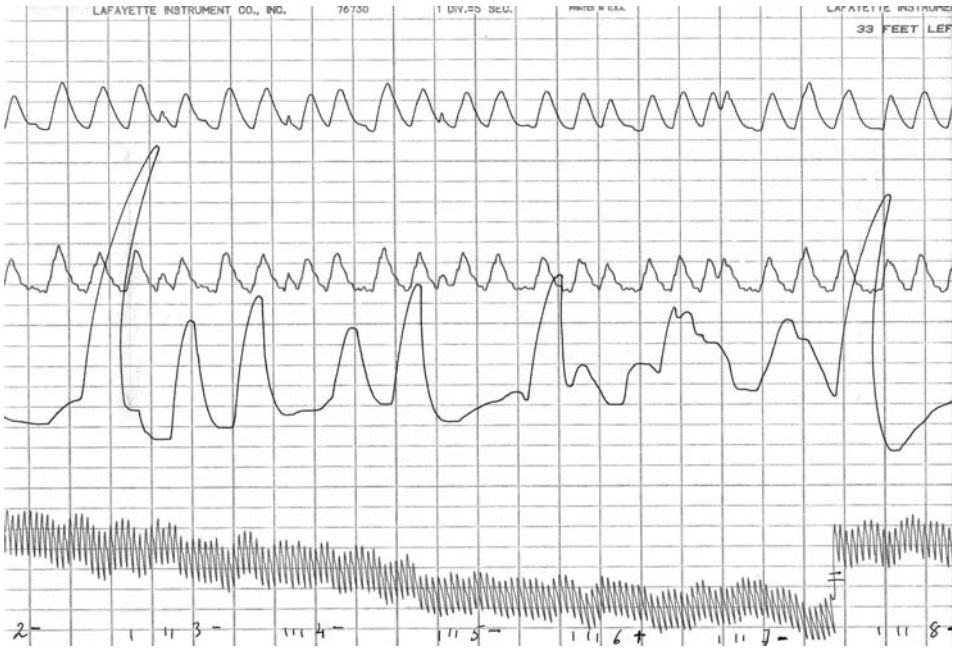
One of the tests used concerned the place where the people examined stated they were when the crime was committed. Question No. 4 concerned being in the house of the parents while they were killed. Question No. 5 concerned the place where the examinee was – according to the testimony – when the parents were killed.
No significant emotional changes were registered after the asking of question No. 4 concerning being in the parents' house during the murder. Decidedly the strongest emotional changes were recorded after question No. 5 was asked, that is when the place named by the examinee as the one where he was at the critical moment was mentioned. Changes are visible both in skin galvanic response and in arterial blood pressure.
It is hard to imagine that memory and emotional traces were not registered in the nervous system of the examinee if he were a witness to or perpetrator of the murder of his parents. The remaining questions concerned other probable places where of examinee might have been.
This examination established what memory and emotional trace connected to the location of the examinee when the murder was performed is registered in the nervous system of the examinee. This trace is lasting, and may be disclosed in successive examinations, should the need to repeat them arise.

Case study No. 5.
This case refers to a suspicion of insurance fraud. The examinee notified the police on 6th January 2008 that somebody had stolen his car from the parking spaces by the house where he lived. At about 2 p.m. the day before, he had left the car in the parking spaces, and he had seen it for the last time at about 9 p.m. on 5th January. He was convinced that the car was stolen from him, and did not know who did it.



The police officer conducting the preliminary proceedings issued a decision to terminate the investigation concerning the theft of the car, due to the lack of data sufficiently substantiating the actual crime. One of the basic reasons for undertaking such a decision was an official note which claimed that the police had "operational evidence" to prove that W.T. submitted a false claim to obtain damages under false pretences, and actually sold or abandoned the car.
Question No. 3 was "Did you abandon your car?", question No. 5 was "Was your car stolen from you?", and question No. 7 was "Did you sell your car?"
By far the strongest emotional changes were present after question No. 5 was asked. Visible are clear emotional changes in the skin galvanic response and arterial blood pressure.
Established through this examination is what memory and emotional trace concerning the loss of the car is registered in his nervous system. This gives reasons to assume that the examinee notified the police about the theft of the car with the best of his knowledge, and the version accepted by the police is false. This trace is lasting, and may be disclosed in successive examinations, should the need to repeat them arise.

Case study No. 6.
This example concerns a case of a false accusation. The examinee was accused of claiming and receiving a financial gain amounting to several tens of thou-

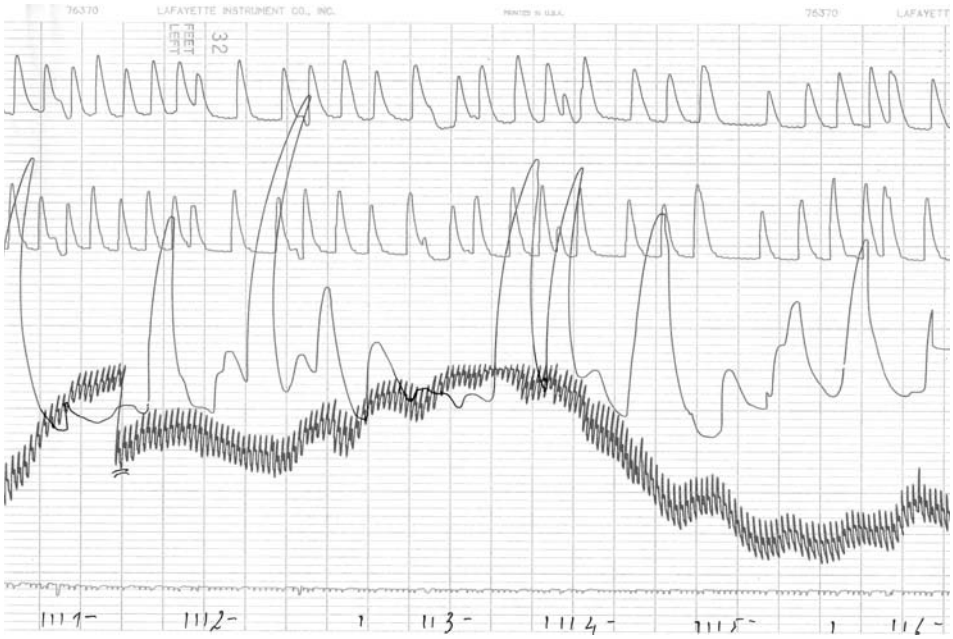


sand zloty. Due to the post he held, he could eliminate the corporation he worked for from competing for highly lucrative contracts. A criminal procedure against the examined was initiated. The examinee admitted that he had maintained professional contacts with the man falsely accusing him for a few months, and even received from him on two or three occasions small presents of a nominal value.

Question No. 5 pertained to receiving money from the slanderer, and questions Nos. 3 and 8 concerned receiving small presents in kind. It is visible that stronger emotional changes occurred after questions Nos. 3 and 8 were asked than after question No. 5. This provides grounds to assume that the examinee did not accept any money from the slanderer. Established through this examination is what memory and emotional trace concerning receiving a financial gain is registered in the nervous system of the examinee. This trace is lasting, and may be disclosed in successive examinations, should the need to repeat them arise.

Case study No. 7.

This case concerns a false accusation. The examined man was accused by an acquaintance of forcing her with violence and threats to have sexual intercourse. Question No. 4 was "Was the sexual intercourse the initiative of your acquaintance?", and question No. 5 was "Was the sexual intercourse your ini-



tiative?". Decisively stronger emotional changes were present after asking the question No. 4, which gives reasons to assume that the sexual intercourse of the examinee with the slandering woman occurred on her initiative. This excludes the element of threats and use of force to coerce her to sexual intercourse. This trace is lasting, and may be disclosed in successive examinations, should the need to repeat them arise.

As the cases above prove, polygraphic examination is vested with far greater capacity than just disclosing the connection of the person examined with the event in question. This is significant not only for the practice but also for the manner of perceiving this field of forensic studies by persons dealing with the broad field of evidence procedure (and law).

## References

Góraj B., Kordek P. (1999), *Czynnościowy Rezonans Magnetyczny w tworzeniu map aktywacyjnych mózgu człowieka*, Polski Przegląd Radiologiczny" No. 64 [in Polish].

Jaworski R. (1999), *Opinia z ekspertyzy poligraficznej, jako dowód odciążający*, Wydawnictwo Kolonia Limited, Wrocław.

Kołecki H. (1977), *Pojęcie i klasyfikacje śladów kryminalistycznych*, Zeszyty Naukowe ASW" No. 18, Warszawa [in Polish].

Kulicki M. (1998), *Wariografia kryminalistyczna*, ed. by Mariusz Kulicki, Wydawnictwo Wyższej Szkoły Policji, Szczytno [in Polish].

LeDoux J. (2000), *Mózg emocjonalny*, Media Rodzina, Poznań [in Polish].

Lewandowski E. (2006), *Badania poligraficzne. Fakty i mity*, Współczesne Problemy Kryminalistyki, Vol. X, Warszawa [in Polish].

Lewandowski E., Lewandowski Ł. (2008), *Alibi Testing Potential in Polygraphic Examination*, European Polygraph, No. 3, Kraków.

Lewandowski E., Lewandowski Ł. (2008), *Teoria i praktyka testów potwierdzających wykorzystywanych w badaniach poligraficznych*, Współczesne Problemy Kryminalistyki, Vol. XII, Warszawa [in Polish].

Lewis M., Haviland-Jones J.M. (2005), *Psychologia emocji*, red. Lewis, M., Haviland-Jones, J.M., Gdańskie Wydawnictwo Psychologiczne, Gdańsk [in Polish].

Martin G.N. (2001), *Neuropsychologia*, Wydawnictwo Lekarskie PZWL, Warszawa [in Polish].

Sehn J. (1960), *Ślady kryminalistyczne*, Z zagadnień kryminalistyki, Vol. 1 Warszawa [in Polish].

Widacki J. (2002), *Kryminalistyka*, ed. by Jan Widacki, Wydawnictwo C.H. Beck, Warszawa [in Polish].

Wójcikiewicz J., Szumakowicz E. (1984), *Ślady pamięciowe*, Studia Kryminologiczne, Kryminalistyczne i Penitencjarne, Vol. 15, Warszawa [in Polish].

# Book reviews

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Daniel T. Wilcox (Ed.)
*The Use of the Polygraph in Assessing, Treating and Supervising Sex Offenders. A Practitioner's Guide,*
Wiley–Blackwell, Chichester, 2009, pp. 332

Dr. Daniel T. Wilcox (a chartered clinical and forensic psychologist with the British Psychological Society), one of the pioneers in claiming that polygraph testing can be very useful in monitoring offenders' post-release behaviour, has compiled a collection of articles written by leading researchers from Britain, Europe, Australia, New Zealand and the United States, summarising the recent findings in this field. As the title suggests, the reader will find an enormous amount of information on polygraph post-conviction sex offender testing (PCSOT). The editor has undertaken to present methods and utilities of polygraph testing used in assessing the risk of re-offending. By cooperating with authors covering a wide variety of subjects, he offers the reader the opportunity to become aware of the history as well as the current and future developments on this topic. Moreover, the book explains the matter of polygraph testing in general, especially dealing with PCSOT.

The combination of authors from different scientific fields shows that cooperation between clinical specialities, governments and professionals in polygraph testing is strongly needed, when the best results in treatment, correction, and risk assessment are in our interest.

Despite the fact that the beginning of the polygraph as a machine dates from the first decade of the previous century, for many people it is still known as

a "lie detector". Moreover, not only are its purpose and function generally misunderstood, but the way in which it is operated is too. This probably explains why a whole chapter is devoted to presenting its history and explaining the way in which it works. Daniel Wilcox and Lars Madsen (a clinical psychologist) have written a historical article containing not only the evolution of polygraph testing, but also the main differences between pre- and post-conviction testing. The value of post-conviction polygraph testing is accurately described by Wilcox and Madsen using empirical evidence.

Even though the first PCSOT took place in the early 1970s and was developed in the judicial community, a large group of people still treat polygraph testing as an unimportant tool with a low level of accuracy and validity. In the first chapter, David Whittingham (a clinical and forensic psychologist) describes the polygraph's clinical application in comparison with other uses in pre-conviction polygraph tests. He also demonstrates the benefits over cost and risk, as well as claiming that the monitoring of situational factors provides a greater understanding of human behaviour, which is very helpful in prevention.

There is no doubt that polygraph testing requires a well-prepared expert as an examiner. Chapter 5 is especially for professionals, as it provides them with the basics for understanding exactly how the polygraph works and how to use it in sex offender treatment. Four different types of denial during testing are also shown. In my opinion, the most interesting part of this chapter, written by the editor and Daniel E. Sosnowski (APA accredited polygrapher, president of the American Polygraph Association), are the common mistakes which are made during testing, which has an influence on its accuracy.

As mentioned before, for the best results in assessing a risk of relapse, professionals from different fields have to work together. Chapter 6 goes some way towards explaining the ways in which all experts collaborate, using six case studies. These show that polygraph testing is also very useful during the treatment process; for example, it helps to improve openness between therapist and offender. Risk theory is dealt with in chapter 8, and is necessary for a comprehensive assessment of offenders' risk. By connecting this theory and professional polygraph testing, we will achieve better results than a self-report tool.

Sex offenders, especially paedophiles, use computers to watch films or look at photographs featuring pornography. Computers are used not only for watching pornographic movies, but also for making contact with children in chat rooms, flirting with them or sending them pornography, which can be a tool for normalizing child/adult sexual behaviour. For cases of deviant sexual inter-

est, the Sexual History Disclosure Polygraph Examination (SHDE) was created, which is used to disclose previous sexual offences and/or previous deviant sexual interest. It also gives the opportunity to assess the risk of cyber-abuse becoming abuse in reality. In chapter 7, Jos Buschman (APA accredited polygrapher and forensic psychologist) and Stefan Bogaerts (a professor of forensic psychology) describe different cases of using the polygraph in cases of internet sexual offences.

The accuracy mentioned previously was the most important subject in the work conducted by Grubin, Madsen and others. Taking it as a base, Lars Madsen has written an article (chapter 9) in which he carefully explains the circumstances for low and very high accuracy.

In most chapters, the related examinations took place in the United Kingdom. However, Peggy Heil (Colorado Department of Corrections) and Kim English (Colorado Division of Criminal Justice) are the authors of a chapter (10) describing the history of the use, standards and many studies conducted in the United States of America. They also question the effectiveness of polygraph testing with psychopaths.

The polygraph examination can be provided not only in sexual abuse cases, but also in other crimes, such as domestic violence, stalking and general violence. Shay Addison (APA accredited polygrapher and psychologist) and Lou Crisilla (APA accredited polygrapher and vice president of the Florida Polygraph Association), in the chapter entitled "Other post-conviction applications of polygraphy", deal with the limitations of this type of testing, proposing using the polygraph not only as a stand-alone technique, but always with other forensic tools. The above-mentioned types of crimes were used as examples of the different uses of polygraph testing.

The book includes other modern techniques useful in the detection of deception, some of which can be used without the knowledge of the examinee, for example: voice, thermal imaging, and pupillometry. Brain waves and functional magnetic resonance imaging are also needed to improve the accuracy level. The chapter includes information about fusing all of these technologies together.

It is not only necessary in assessing the risk of sex offenders to know the sexual interests of the examinee. One chapter reviews the forensic assessment on this subject, using physiological and non-physiological methods.

Despite the fact that the title "The use of the polygraph in assessing, treating and supervising sex offenders" is focused on sex offenders, many outcomes presented in the book might be useful for those who work with other types of criminal. Certainly this is an outstanding source of knowledge, which helps

in understanding the most important items in post-conviction testing and shows how to involve them in prevention, policing and judicial processes. This book is highly recommended for professionals and students in various fields of criminalistics and social welfare.

Małgorzata Wrońska*

* e-mail

EUROPEAN
POLYGRAPH

Volume 3 • 2009 • Number 3–4 (9–10)

Daniel T. Wilcox (Ed.)
*The Use of the Polygraph in Assessing, Treating and Supervising Sex Offenders. A Practitioner's Guide,*
Wiley–Blackwell, Chichester, 2009, pp. 332

This newest book on the global market devoted to polygraph testing is edited by Dr. Daniel T. Wilcox, a famous British clinical and forensic psychologist. It is a joint effort, including contributions mostly by British and American authors, as well as those from Australia, New Zealand and the Netherlands. The volume consists of a foreword, followed by fourteen relatively lengthy chapters. Each of the chapters is highly competent, well documented and comprehensive in scope, and focuses on the subject matter of polygraph post-conviction sex offender testing (PCSOT).

The introduction, authored by the former president of the American Polygraph Association and current chair of the APA PCSOT committee, states that polygraph testing of sex offenders is the most rapidly developing sector of polygraph examinations in the world. In the USA such examinations are used in 46 states, and the remaining 4 states are in the analytical stage of the process of their introduction. Never before in the history of polygraph testing did such examinations have such support in this area as they do now. Apart from the natural support of polygraphists, these tests are also endorsed by the judiciary and by politicians.

It is likely that these circumstances, coupled with the fact that Great Britain is now preparing for the introduction of PCSOT on a large scale, are where the book stems from.

Overall, the volume presents the scope of the use of polygraph testing in rehabilitation of sex offenders. Irrespective of the complex and varied corrective and therapeutic concepts, a constant element of working with offenders who were convicted and then released on parole is their comprehensive monitoring, aimed at the reduction of the possibility of their committing another crime. Clearly this job is well suited for polygraph testing. The use of polygraph examinations for monitoring sex offenders on parole was initiated by Stanley Abrams in 1973. Despite certain shortcomings in terms of scientific grounding, PCSOT was developed in the USA in a number of wide-ranging programmes.

The polygraphist who carries out the PCSOT is one of the elements of the triad: therapist (rehabilitator) – probation officer (usually a police officer) – polygraphist; usually in the monitoring capacity and with the task of supplying the other two officials with information on the offender.

The programmes carried out nowadays usually consist of four types of examinations. The first group consists of examinations aimed at obtaining a confession with regard to the offence that was the basis of conviction. This naturally applies only to offenders who had not confessed earlier during the criminal trial, hence the name: specific issue denial testing. If the offender acknowledges his/her guilt status, further corrective methods are easier to apply.

The second group consists of sexual history disclosure examinations, which allow for a better understanding of the past interests and sexual behaviours of the subject, as well as a better selection of measures and more accurate estimate of risk and direction of possible re-offending.

The third group contains maintenance polygraph examinations, where the extent to which the offender on parole implements the requirements of the court is examined. Tests deliver up-to-date information on the behaviour of the offender and are conducted every 3–6 months.

The last group consists of monitoring polygraph examinations, which are directed at discovering possible new offences or breaches committed while on probation. The tests are conducted on the basis of information obtained by the authorities pertaining to the fact that the offender fails to observe the requirements imposed upon him/her, e.g. he/she was seen meeting minors with no supervision, visited websites with banned material or used public libraries for accessing such materials, or entered an area where he/she was not allowed to go.

Clearly, PCSOT is conducted not with the purpose of learning more about the crimes which are under investigation (with the exception of monitoring examinations), but rather in order to make the offender acknowledge his/her

guilty status and accept rehabilitation as a value. This is reflected both in the way the offenders are motivated to take the tests (e.g. by informing them of the consequences of being deemed "liars" at the current phase of probation), as well as the offenders' perception of the examinations. These perceptions are usually favourable, since the very fact of appropriate cooperation with the polygraphist creates the opportunity for improving the offender's image.

Polygraph examination also improves the cooperation of the person undergoing rehabilitation in their contact with the therapist. Moreover, it is a source of new useful information for both the therapist and the probation officers. Of course, the close cooperation of all parties involved is a key factor. This cooperation is carried out within the framework of a rehabilitation plan, drafted individually for each offender. There is a lot to be said for the utility of polygraph tests in predicting the risk of future undesirable behaviours of subjects. In this respect, it is vital to discover the past frequency of offending and of failing to observe probation requirements. The possible discovery of unknown facts pertaining to using violence within a sexual context is also crucial.

Questions arise as to the diagnostic value of polygraph examinations in this context. Scientific knowledge in this respect is largely imperfect, since most experiments and studies have so far focused on the effectiveness of polygraph testing pre-conviction. The modest scope of research material does provide scientific grounding for optimism, albeit with a clear recommendation to treat results which are unfavourable to the subjects as "red flags", i.e. warning information only, which should not become the sole basis for further actions.

The American Polygraph Association has been working on developing a set of standards for examinations of this type. The minimum has been set at completion of a 40-hour specialist course for polygraphists intending to carry out PCSOT. Many state jurisdictions in the USA have accepted this requirement either in its original form or with adaptations to their own specific circumstances.

There is no intention to hide the deficit of knowledge on validity, reliability and accuracy of polygraph examination in the case of convicted sex offenders. Moreover, there is no option to release the polygraphist community from the duty to diligently research the effectiveness of polygraph examination in PCSOT usages. The following questions remain in force: Would those people who are being monitored and rehabilitated be more honest if polygraph examinations were excluded from the inventory of measures used towards them? Are the tools of clinical diagnostics available to therapists and the control measures used by probation officers more efficient than polygraph examinations?

These are no easy questions. It seems, however, that negative answers to them are more justified. One must also agree with the following statement: "Errors with deceptive individuals can lead to new offences against children, whereas errors with truthful individuals can devastate people's lives" [T. Cross, L. Saxe (2001), Polygraph testing and sexual abuse: The lure of the magic lasso, *Child Maltreatment*, 6, p. 203]. While this caution is important to remember, the same errors can be made without the polygraph and result in similar consequences" (p. 212).

The last part of the volume to some extent broadens its primary scope. Namely, it appears that the same assumptions that underlie the use of polygraph examinations in rehabilitation of sex offenders are applied also in respect to other offenders, and in particular to perpetrators of acts such as stalking, domestic abuse and other violent behaviours, some with the sexual aspect included. Despite the first – and encouraging – attempts to use polygraph examinations in these types of cases, the considerations on this new field of research remain in the realm of speculation. However, they give rise to hopes on the expansion of post-conviction examinations. It is natural that there is a tendency to appreciate a potential new source of independent information on an offender.

The book ends with a few notes on the tactics of interviewing and interrogation, alternate new technologies of lie (deception) detection (a very interesting review of the newest research in this area) and a final review on forensic (but other than polygraphy) assessment of sexual interest.

As I mentioned earlier, all of the constitutive parts of the volume are very well written. The language is clear and, as far as it is possible to be so in scientific texts, simple. The editing of the volume is also excellent, with the possible exception of the (perhaps inevitable?) repetitions of historical references. It is worth stressing again that the documentation of each article/chapter is impressive.

Since it is very difficult to criticise what the book contains, let me say a few words on what it fails to contain and what it in my opinion definitely should contain, considering that the focus of the volume is, apart from the sex offender, on the polygraphist-practitioner. This polygraphist-practitioner should be able to find in this book two more chapters: one on the specific role and the details of the pre-test interview in PCSOT, and another on the optimal techniques of examination. These issues are mentioned in the book on rare occasions only, and are scattered throughout the volume. This is certainly insufficient. Possibly in future editions these matters will be dealt with in more depth. Maybe a monograph is in the works already, and the authors and the publisher are aware of that? That would be very good news.

The final conclusion is clear: we have gained a book that is very good and very important for lawyers, criminologists, forensic scientists, police officers, experts on polygraphy and those concerned with the penitentiary system, therapists, probation officers, activists and everyone interested in counteracting sex crimes.
May 27, 2009

Jerzy Konieczny*

* jerkonieczny@wp.pl

# The basic information for Authors

To publication will be accepts unpublished research papers as well as review article, case reports, book reviews and reports connected with polygraph examinations.

Submitted manuscripts must be written in English.

All papers are assessed by referees (usually from Editorial Board), and after a positive opinion are published.

Texts for publication should be submitted in the form of normalized printout (1800 characters per page) and in electronic form (diskette, CD), or sent by e-mail to Editorial Office.

The total length of research papers and review article should not exceed 12 pages, case reports – 6 pages, and other texts (book review, report) – 5 pages.

The first page of paper should contain: the title, the full name of the author (authors), the name of institution where the paper was written, the town and country.

Figures should be submitted both in printed form (laser print, the best) and electronic form.

Tables should be numbered in Roman numerals and figures in Arabic ones.

Figures, tables, titles of figures and titles of tables should be included on a separate page. The places in the text where they are to be included should be indicated.

The references should be arranged in the alphabetical order according to the surnames of the authors.

The references should be after the text.

Each reference should include: the surname (surnames) of the author (authors), the first letter of author's first name, the title of the book, year and place of the publication, the name of publisher, or the title of the paper, the full title of the journal, the year, the volume, the number and the first page of the paper.

For example (in references):

Reid J., Inbau F. (1966), *Truth and Deception: the Polygraph ("Lie-detector") Techniques*, Williams & Wilkins, Baltimore.

Abrams S. (1973), *Polygraph Validity and Reliability – a Review*, Journal of Forensic Sciences, 18, 4, 313.

and (Reid, Inbau, 1966), (Abrams, 1973) inside text.

Texts for publication in "European Polygraph" should be mail to:

"European Polygraph"
Andrzej Frycz Modrzewski Krakow University
ul. Gustawa Herlinga-Grudzińskiego 1
30-705 Kraków (Poland)

Or e-mail: margerita.krasnowolska@kte.pl