Polygraph

VOLUME 30 2001 NUMBER 3

Contents

Published Quarterly

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph

Subscription Information: *Polygraph* is published quarterly by the American Polygraph Association. Advertising and Editorial Address is P.O. Box 10342, Ft. Jackson, SC 29207 (USA). Subscription Rates: One year $80.00 (domestic), $100.00 (foreign). Change of address: APA National Office, P.O. Box 8037, Chattanooga, TN 37414-0037.
The Sheridan Press

Fiscal Year 2000 Report to Congress on the Department of Defense Polygraph Program

I
DoD Use of Polygraph Examinations

The Department of Defense has used the polygraph for almost half a century. It is used in criminal investigations, counterintelligence cases, foreign intelligence and counterintelligence operations, exculpation requests, and as a condition for access to certain positions or information. The polygraph is a tool that enhances the interview and interrogation process. Often it is the only investigative technique capable of providing essential information to resolve national security issues and criminal investigations. The use of the polygraph as a condition for access is limited by a statutory quota for CSP examinations.

The following table reflects Department of Defense Polygraph statistics for Fiscal Year 2000.

Criminal	2,096	18.8%
Exculpatory	499	4.5%
CI Scope (CSP)	7,890	70.9%
All Others*	646	5.8%
Total**	11,131	100%

* Includes examinations conducted in support of personnel security investigations, counterintelligence and intelligence operations, and polygraph assistance to non-DoD federal agencies.

** Does not include polygraph examinations conducted by the National Security Agency (NSA). A breakout of polygraph examinations conducted by NSA is contained in a classified table submitted with this report. Nor does it include polygraph examinations conducted by the National Reconnaissance Office, which are conducted under the authority of the Director of Central Intelligence (DCI).

II
Fiscal Year 2000 Counterintelligence-Scope (CSP) Polygraph Examinations

Section 1121 of the National Defense Authorization Act for Fiscal Years 1988 and 1989 (Public Law 100-180, December 4, Car1987; 101 Stat. at 1147) authorizes the Department of Defense to conduct CSP examinations as a condition for access to certain information.

The purpose of the CSP Program is to deter and detect espionage, sabotage, and terrorism. The following topics are covered during the CSP examination: (1) Involvement with a foreign intelligence/security service, involvement in espionage; (2) Involvement in terrorism; (3) Unauthorized foreign contacts; (4) Deliberate failure to protect classified information; and (5) Damaging/sabotaging government information systems, clandestine collection, or defense systems. These CSP topics meet the needs of both DoD and the Intelligence Community facilitating the transfer of security clearances.

The Department published a handbook for federal polygraph examiners standardizing techniques and procedures for conducting polygraph examinations. The handbook also outlines the Quality Assurance Program (QAP) wherein DODPI inspects federal polygraph

During CSP testing, a contractor employee surrendered TOP SECRET/CODE WORD material he had in his possession. The examinee agreed to a permissive search of his residence where additional classified material was found. This matter has been referred for further investigation.

* * * * *

During a CSP examination, the examinee admitted to the unauthorized removal of TOP SECRET/SCI material from a facility on one occasion and unauthorized disclosure of classified information while deployed to South West Asia. Further investigation is ongoing.

* * * * *

During CSP testing, examinee admitted that in 1991, during Operation Desert Storm, he was provided a SECRET topographical map containing turning points and refueling requirements to friendly bases. He also had 10 to 15 SECRET target maps. He retained the maps at his residence until 1997 when he shredded the target maps, but retained the SECRET topographical map in his household goods. This matter was referred for further investigation.

* * * * *

During CSP testing, examinee provided information regarding her mishandling of classified information. The examinee stated that she could think of a half-dozen instances where she had accidentally divulged classified material. She admitted on three occasions she divulged NATO satellite communications to persons not authorized to receive such information. She also had a friendship with a Russian émigré to Israel who specialized in cryptography.

* * * * *

During CSP testing, examinee admitted that the day prior to his polygraph test, he destroyed approximately 200 pages of classified material at his residence. The classified material was obtained the prior year while he was assigned to an overseas position. He discovered the material while he was unpacking his personal effects and it was destroyed due to his anxiety over the pending polygraph test. Following this admission, he successfully completed his polygraph test.

* * * * *

During a CSP examination, a military member, assigned to the National Security Agency, admitted, that on two occasions, he had deliberately disclosed SECRET information to unauthorized persons.

* * * * *

During a CSP examination, examinee admitted that approximately two and a half years ago, while assigned to an aircraft carrier, he used his personal laptop computer to type a SECRET document regarding a weapons system. He could not recall if he ever deleted the document from his computer's hard drive, which was at his private residence. He also admitted discussing classified information with a relative and a former college classmate, and discussing ship movements with family and friends.

* * * * *

During CSP testing, examinee admitted storing classified information at his quarters in Hawaii. He stated that he moved to Hawaii from Fort Meade, Maryland and had attempted to obtain a courier orders but was unable to get this support, so he elected to hand carry the documents with him to Hawaii. He stored the documents in a closet at his residence in Hawaii. He denied providing the material to unauthorized persons. The documents were returned to Government control.

III
Utility Of the Investigative Polygraph

During Fiscal Year 2000, DoD investigations obtained unique and significant information from interviews conducted with the aid of the polygraph. In all illustrated instances, the polygraph examination process produced significant security or criminal information, which would not otherwise have been secured for the specific investigation. The polygraph examination process was also

valuable in helping to establish the innocence of persons charged with serious infractions.

* * * * *

An investigation was initiated regarding the theft of a diamond ring, valued at $4,000.00, from a guesthouse room operated by the U.S. Government. One of the housekeeping staff was suspected of stealing property in the past. It was determined that she was responsible for cleaning the room from where the ring was stolen. She was interviewed and denied stealing the ring and agreed to undergo a polygraph examination. The results of the examination indicated deception, and she confessed to stealing the ring which was recovered.

* * * * *

An investigation was initiated regarding a suspicious fire aboard a U.S. ship. The suspect denied involvement and requested a polygraph examination to support his denial. The polygraph examination results indicated deception and the suspect admitted to lighting a mop on fire with his cigarette lighter and then putting the mop into the fuel tank.

* * * * *

An investigation was initiated regarding the possible arson of a motor vehicle. One of the suspects was interviewed and admitted assisting another individual in stealing the vehicle, but denied setting the car on fire and agreed to undergo a polygraph examination. During the polygraph interview, he admitted assisting the other individual in setting the fire.

* * * * *

A National Guard Unit discovered that a number of intrusions had been made to its computer system. Two individuals assigned to a Health Clinic were identified as suspects. Both were interviewed and admitted hacking into the system, but said they did so to identify weaknesses in the system. Both of them admitted to obtaining not only their supervisor's passwords but also the systems administrator's passwords, which gave them root directory access. Additional polygraph testing determined that they did not plant any viruses or back doors in the system.

* * * * *

An investigation was initiated regarding a bomb threat made to a high school on a military installation. A military dependent was identified as a possible suspect. The suspect denied placing the bomb threat and agreed to undergo a polygraph examination. The suspect was evaluated as deceptive during the polygraph examination and admitted to placing the bomb threat in order to avoid attending school.

* * * * *

An investigation was initiated regarding allegations that a woman was raped. The victim stated that she and the accused had been to a local bar and had a few drinks. The victim believes she may have been drugged and did not recall the sexual events of the evening. She stated that she woke up the next day with only her under-garments on and was later told by the accused that they had sexual intercourse. The victim stated that she would not have consented to sexual intercourse and believed that she had been raped. The accused requested to undergo a polygraph examination to support his denial that he raped the victim. The polygraph results indicated deception and the accused admitted that the victim had passed out when he began having sexual intercourse with her.

* * * * *

An investigation was initiated regarding a reported robbery of $6,646 from a community club. During the investigation, several inconsistencies surfaced during the interviews of the club employee who reported the alleged robbery. The polygraph results were evaluated as deceptive and the employee admitted that he had fabricated the robbery incident and had stolen the funds himself over a period of time.

* * * * *

An investigation was initiated regarding an allegation of physical child abuse. The suspect, a licensed daycare provider on a

military installation, was accused of striking a child with her hand. The suspect was interviewed and denied the allegation and agreed to undergo a polygraph examination. The polygraph examination results indicated deception, and the suspect confessed to striking the child.

* * * * *

An investigation was initiated when it was discovered that a number of wild horses had been slaughtered in the deserts of Nevada. The investigation developed two military suspects and a civilian suspect. One of the military suspects agreed to undergo a polygraph examination. During the polygraph examination, the suspect admitted killing one horse and shooting into the herd. He also implicated the other military suspect and the civilian suspect.

* * * * *

An investigation was initiated regarding a threat and damage to government property when someone painted swastikas and left a threatening note on the door of U. S. Army quarters. A U.S. military member had been observed leaving the area and was subsequently interviewed. The suspect denied any involvement in the incident and agreed to undergo a polygraph examination. The results of the polygraph examination indicated the suspect was truthful in his denial. The victim was reinterviewed and confessed that he had committed the offense to get attention.

* * * * *

An investigation was initiated regarding an allegation that a military member had sexually molested his four year old step-daughter. The military member denied any sexual contact with the victim, and agreed to undergo a polygraph examination. The polygraph examination was evaluated as deceptive. Subsequently, the military member admitted to sexually molesting his stepdaughter.

* * * * *

An investigation was initiated regarding the theft of a lap computer from a military installation. During the investigation, a source reported that he heard a military member admit to the theft of the computer. The suspect was interviewed and denied any involvement in the theft of the computer and agreed to undergo a polygraph examination. The polygraph examination was evaluated as deceptive. Subsequently, the suspect admitted to the theft of the computer.

* * * * *

An investigation was initiated regarding the death of an infant on a military installation. The cause of death was the "shaken baby syndrome". Both the military mother and civilian father were interviewed and agreed to undergo a polygraph examination. Both polygraph examinations indicated deception. The mother refused to answer any further questions. The father admitted that he and his wife had an argument while the wife was holding and shaking the baby. All of a sudden, the wife threw the baby on the couch. The baby started choking and he administered CPR and took the baby to the hospital.

* * * * *

An investigation was initiated when a military member reported that his wife was missing. After a thorough investigation, the wife could not be located. The husband was interviewed and agreed to undergo a polygraph examination. The polygraph examination results indicated deception. Subsequently, the husband admitted to killing his wife and disclosed the location of her body.

* * * * *

A contractor, being sponsored for access to NSA information, disclosed during a polygraph examination, that he had been approached by a military officer of a foreign government to provide classified information on a sensitive government project. The details of this incident were obtained and disseminated for further investigation.

* * * * *

During a polygraph examination, a U.S. military officer, with access to NSA classified information, admitted friendships

with Russian and PRC military officers, and also admitted downloading classified information onto his home computer.

IV
Training and Qualification Standards for Department of Defense Forensic Psychophysiologists (Polygraph Examiners)

The Department of Defense maintains very stringent standards for polygraph examiners. The Institute's basic polygraph program is the only program known to base its curriculum on forensic psychophysiology, and conceptual, abstract, and applied knowledge that meet the requirements of a master's degree-level of study. Candidates selected for DoD polygraph positions must meet the following minimum requirements:

1. Be a United States citizen.

2. Be at least 25 years of age.

3 Be a graduate of an accredited four-year college or have equivalent experience that demonstrates the ability to master graduate-level academic courses.

4. Have two years of experience as an investigator with a federal or other law enforcement agency. Two years of comparable experience may be substituted for this requirement.

5. Be of high moral character and sound emotional temperament, as confirmed by a background investigation.

6. Complete a DoD-approved course of polygraph instruction.

7. Be adjudged suitable for the position after being administered a polygraph examination designed to ensure that the candidate realizes, and is sensitive to, the personal impact of such examinations.

All federal polygraph examiners receive their basic polygraph training at DoDPI. In Fiscal Year 2000, the Institute trained 62 new polygraph examiners. After completing the basic polygraph training, DoD personnel must serve an internship consisting of a minimum of six months on-the-job-training and conduct at least 25 polygraph examinations under the supervision of a certified polygraph examiner before being certified as a DoD polygraph examiner. In addition, DoD Polygraph examiners are required to complete 80 hours of continuing education every two years. To help meet this requirement, the Institute offers various specialized courses in forensic psychophysiology and related disciplines. In Fiscal Year 2000, approximately 555 students attended the specialized courses.

Department of Defense Forensic Psychophysiologists (Polygraph Examiners)

Fiscal Year	*Average Number of Examiners*
1994	192
1995	176
1996	164
1997	153
1998	147
1999	144
2000	138

V
Polygraph (Forensic Psychophysiology) Research

Mandated by Congress, the research program at the Institute is focused on: (1)developing new psychophysiological detection of deception (PDD) techniques, instrumentation and analytic methods to improve PDD technology; (2)evaluating the validity of existing and future PDD techniques; and (3)the conduct of research on PDD countermeasures.

To facilitate research, a small grant program was established in Fiscal Year 1992. During Fiscal Year 1997, the Institute developed a Prioritized Research Plan at the request of the Security Policy Board staff. This plan outlines a prioritized series of projects to be completed in support of the PDD community. The plan was approved by the Personnel Security Research Subcommittee and provides the basis for the acceptance of research efforts conducted by or on behalf on the Institute. In January 1999, the Institute began an effort to broaden its presence in the scientific and academic communities. This initiative seeks to give the DoDPI a research workforce that is competitive with the best minds from academia and persons knowledgeable of emerging technologies. As a result of this effort, during Fiscal Year 2000, the Institute published a revised research plan to include the Government's interest in less-invasive techniques and to enhance DoDPI's ability to answer agency-specific customer questions.

Current Research Projects

An Examination of Response Parameters of Electrodermal Recording (EDR) to Standard Stimuli. The objective of this project is to investigate whether equivalent electrodermal responses are obtained to equivalent psychological stimuli presented at different electrodermal tonic levels. The outcome will determine if resistance or conductance is a more accurate measure during PDD examinations.

Effects of Augmented Physiological Feedback on the Detection of Deception. This project will determine if augmented feedback improves the accuracy of PDD examinations.

Scaled P300Scalp Profiles in Detection of Deception. Previous Electroencephalographic (EEG) studies of deception have been limited to the changes in the amplitude of responses to specific questions. An investigator at North Western University has been awarded funds to pursue a unique measure of deception, the distribution of EEG activity over the scalp. A preliminary report was favorable.

High Definition EEG/ERP Recordings and the Detection of Deception. The University of South Carolina will conduct research on brain activity as it relates to the detection of deception. The project will use high-density EEG/ERP recordings, and correlate these findings with current autonomic nervous system recordings during a PDD examination.

Remote Sensing of Emotion and Stress Using Laser Doppler Vibrometer. This project involves the use of emerging technologies to develop methods for deriving simultaneous information from the Laser Doppler signal regarding multiple physiological functions including body tremor, respiration, cardiac function, muscle contraction, and sweating. Laser Doppler Vibrometry recording methods do not require the invasive attachment of physical transducers and could be adapted to multiple-examination settings. Preliminary results are very favorable.

Non-invasive Detection of Deception Methods for Field Use. A search for an effective non-invasive method of lie detection has persisted for many decades. Previous research on voice-based detection of deception methods has met with limited success. A new voice-based detection of deception device, called the "Vericator", has been developed by a private company. Capable of multiple modes of operation, "Vericator" offers a versatile package that can be used in a wide array of environments and situations, ranging from the strict question and answer format demanded of the polygraph to a free-flowing telephone conversation. The current project will assess the validity of the "Vericator's" on-line mode of operation. Three different scenarios will be

conducted in a pilot study to determine its effectiveness and sensitivity across different environments. Data analysis will be conducted using the direct veracity decisions produced by the Vericator in addition to its peripheral measures of stress and deception. Following the successful development of a scenario, multiple sites will be used to facilitate the rapid collection of data. This will allow for quick assessment of Vericator as an effective tool for the detection of deception.

Development of an Expert System for Human Assisted and Automated R/I. This project is designed to develop a novel waveform application for scoring the Relevant/Irrelevant test format polygraph data. A number of advanced analytical methods have been applied to electroencephalographic (EEG) and magneto encephalographic (MEG) signals which may be applied to the analysis of the physiological polygraph data. The goal of this project is to use traditional polygraph data and combine this data into a novel waveform that will enhance the accuracy of the PDD examiner decision. This technique is highly accurate and reliable for single trial analysis and should be adaptable to similar datasets collected from polygraph examinations. If successful, future analysis may include additional human physiological data for enhanced accuracy.

Test for Espionage and Sabotage (TES) Validation Project. TES is the primary counterintelligence security screening polygraph examination used in the federal government. There have been two previous research studies on TES which indicated that TES had an accuracy rate in the range of 80 to 90 percent. However, the sample sizes were rather modest. A subsequent study with a larger sample did not produce accuracies as high. It is unknown if the disparity in accuracy between the studies was associated with the participant variables, experimental conditions, or the test format. Two pilot studies have since been conducted with new methodology that closely approximates real world espionage activities and produces subject samples that better represent the typical TES candidate in the government setting. The goal is to conduct studies that have the best generability to government's use of the TES.

The Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Probable Lie and Directed Lie Polygraph Tests. Investigators at the University of Utah are examining the usefulness of administering an acquaintance test during a PDD examination. An acquaintance test is used to familiarize examinees with the PDD procedure, and reassure the examinee that the PDD procedure is effective. Unfortunately there have been no systematic studies to determine the validity or efficacy of this procedure. This study will examine the usefulness of the acquaintance test and also compare the directed versus probable lie comparison questions where the comparison test is used and not used.

Evaluation of DoDPI Evaluation Techniques. Several polygraph laboratory studies published by the University of Utah have reported higher accuracy rates for PDD chart evaluations than DoDPI has found in its laboratory studies. One of the possible sources of these differences is the method of chart interpretation. In the present study, the University of Utah decision rules will be compared to those of DoDPI to determine how the differences in rules influence PDD decision accuracy. The study will attempt to determine the best combination of decision rules to maximize decision accuracy.

Contracts

Application of Thermal Image Analysis to Polygraph Testing. Infrared thermal imaging, a non-intrusive and non-invasive technology may be used to determine if peripheral changes in skin surface temperature (SST) are related to psychological stress. Preliminary results are favorable and additional studies need to be conducted to determine if the technology is useful for PDD. The purpose of this study is to explore an alternative method of measuring the same physiological activity to increase sensitivity. This method would also have utility for screening interviews.

R/I Expert System. Each year the federal government performs thousands of multiple issue polygraph screening examinations on employees and prospective employees. One of the common testing techniques in this application of the polygraph is called the Relevant/Irrelevant (R/I) test. The R/I format

interprets relevant and irrelevant questions, with repetition, in a series that is unpredictable to the examinee. Currently R/I charts are interpreted globally by the testing examiner. The R/I technique is one of the last remaining techniques for which an acceptable, objective form of analysis has not been developed. An expert analytical system would enhance the consistency of chart interpretation.

TES Algorithm. The John Hopkins Applied Physics Laboratory will develop and implement a computerized algorithm for evaluating the Test for Espionage and Sabotage (TES) polygraph screening examination used extensively within the federal government. This effort will have an immediate impact on the quality of the polygraph screening process.

Polygraph Countermeasure Detection Expert Assistant. The primary purpose of this effort is to quickly field a countermeasure screening tool. Polygraph experts will provide the knowledge necessary for the programmers to implement rules into the computer algorithm. This will provide the field examiner with a countermeasure screening algorithm.

Walter Reed Army Medical Hospital. Voice stress analysis has been widely promoted to law enforcement agencies as a means to detecting deception. Although controlled laboratory studies do not indicate the voice stress analysis technology is any more effective than chance, in detecting deception, there is a possibility that voice stress analysis is responsive to psychological stress. DoDPI working with Walter Reed, will record voice stress samples, objective stress questionnaire data, salivary cortisol, heart rate, and blood pressure as well as plasma hormones responsive to stress. The data will be combined with on-line voice stress data from another DoDPI study to independently validate the potential utility for voice stress data analysis.

U.S. Library of Congress. As part of the DoDPI mission to monitor foreign polygraph activity, the Institute collects research publications from foreign sources and under contract with the U.S. Library of Congress provides translations of selected articles and publications.

Other DoDPI Research. DoDPI is also conducting research on Cardio Element Analysis, A History of Comparison Questions, Scoring and Decision Rules in an Objective Scoring System, Rank Order Scoring Systems, Normative Respiration Data from Field Polygraph Examinations, Exploratory Study of Traditional and Objective Scoring Systems, Horizontal Scoring Systems, Case Study of a Spy, and Costs and Benefits of Spot Scoring.

Other Activities

The Institute maintains contacts with PDD examiners in other countries to keep abreast of PDD development around the world. The Institute issues quarterly reports summarizing international PDD activity. During the last few years PDD activity has increased significantly worldwide.

DoDPI taught a total of 15 courses to more than 1,000 students within the polygraph community. Additionally, the DoDPI staff provided polygraph related instruction to more than 100 state and local law enforcement students.

False Rape Allegations

Eugene J. Kanin[1]

Abstract

With the cooperation of the police agency of a small metropolitan community, 45 consecutive, disposed, false rape allegations covering a 9-year period were studied. These false rape allegations constitute 41% of the total forcible rape cases (n = 109) reported during this period. These false allegations appear to serve three major functions for the complainants: providing an alibi, seeking revenge, and obtaining sympathy and attention. False rape allegations are not the consequence of a gender-linked aberration, as frequently claimed, but reflect impulsive and desperate efforts to cope with personal and social stress situations.

Key words: rape, rape mythology, sexual assault, unfounded rape

Of the many controversies surrounding the crime of rape, no more thorny issue arises than that dealing with false allegations. Generally, this issue is couched in terms of unfounded rape. However, we are not addressing that concept here since unfounded rape is not usually the equivalent of false allegation, in spite of widespread usage to that effect. There is ample evidence, frequently ignored (see MacDonald, 1971; Brownmiller, 1975), that in practice, unfounded rape can and does mean many things, with false allegation being only one of them, and sometimes the least of them. Other factors that are typically responsible for unfounded declarations are victim's late reporting to the police, lack of corroborating evidence, lack of cooperation by the victim and/or witnesses, reporting in the wrong jurisdiction, discrepancies in the victim's story, wrong address given by the victim, victim's drunkenness, victim's drug usage, victim's being thought a prostitute, victim's uncertainty of events, victim's belligerence (Clark and Lewis, 1977; Hursch, 1977; Katz and Mazur, 1979; Kanin, 1985; LaFree, 1989). In sum, the foregoing largely represent those conditions that could seriously frustrate efforts to arrest and/or convict the offender. This paper deals exclusively with false rape allegations: the intentional reporting of a forcible rape by an alleged victim when no rape had occurred.

False rape charges have probably been in existence as long as the concept of rape. However, in the 20th century, medical jurisprudence saw a new development that enabled false allegations to be viewed as a singular instance of gender-related lying, something quite different in nature from the false accusations of robbery or burglary that were made by men. In short, false rape accusations became a reflection of a unique condition of women, not unlike that of kleptomania (Abelson, 1989). This new development was the masochistic nature of woman doctrine, a perspective that assumed women had a subconscious desire for rape, as evidenced by their rape fantasies (Freud, 1933; Deutsch, 1944; Horney, 1933), and that neurotic individuals would convert their fantasies into actual beliefs and memory falsification (for an extensive and critical treatment of this perspective, see Edwards, 1981, 1983; Kanin, 1982; Bessmer, 1984). In addition, some influential medical figures adopted the position that false rape allegations were widespread (Menninger, 1933; Guttmacher and Weihofen, 1952). Many legal

[1] Department of Sociology and Anthropology, Purdue University, 1365 Winthrop E. Stone Hall, West Lafayette, Indiana 47907-1365.

Originally published in *Archives of Sexual Behavior*, Vol. 23, No. 1, 1994. Reprinted with permission.

scholars enthusiastically endorsed this medical position (Wigmore, 1940; Juliver, 1960; Comment, 1973; Hibey, 1973) and commonly recommended that rape complainants be routinely subjected to psychiatric examination in order to determine their truthfulness (Guttmacher and Weihofen, 1952; Sherwin, 1973; Comment, 1973). An American Bar Association committee offered a similar recommendation to this effect as early as 1937-1938 (Weihofen, 1959).

In the legal literature, *pseudologia phantastica* became the authoritative scientific label for the condition responsible for false rape reporting (Grablewski, 1958; Juliver, 1960). Pseudologia phantastica was described as a "Delusional state in which the complainant truly believes that she had been raped although no rape, and perhaps no sexual contact of any kind had taken place. Since she firmly believes this non-fact, her story is un-shakable" (Bessmer, 1984). Less pretentious legal scholars made the same point by merely making references to delusional and hysterical states (Smith, 1953-1954; Comment, 1970). In recent years, however, possibly as a response to the women's movement, members of the mental health and legal community have become markedly less likely to express such a position on false rape allegations. In England, judges still rather freely comment on the mendacious nature of women (Lowe, 1984).

Currently, the two main identifiable adversaries involved in the false rape allegations controversy are the feminists and the police. The feminists are by far the most expressive and prominent on this issue. Some feminists take the position that the declaration of rape as false or unfounded largely means that the police do not believe the complainant; that is, the rape charges are real reflections of criminal assault, but the agents of the criminal justice system do not believe them (Brownmiller, 1975; Russell, 1984). Some feminists virtually deny the existence of false rape accusations and believe the concept itself constitutes discriminatory harassment toward women (see Grano, 1990). On the other hand, police are prone to say the reason for not believing some rape complainants resides in the fact that the rapes never occurred (Payton, 1967; Wilson, 1978; Jay, 1991). Medical Examiners lend support to this police position by emphasizing the ever-present possibility that rape complainants may be lying (Shiff, 1969, 1971).

The purpose of this paper is to report our findings on the incidence and dynamics of false rape allegations from a long-term study of one city's policy agency.

Method

This investigation is essentially a case study of one police agency in a small metropolitan area (population = 70,000) in the Midwestern United States. This city was targeted for study because it offered an almost model laboratory for studying false rape allegations. First, its police agency is not inundated with serious felony cases and, therefore, has the freedom and the motivation to record and thoroughly pursue all rape complaints. In fact, agency policy forbids police officers to use their discretion in deciding whether to officially acknowledge a rape complaint, regardless how suspect that complaint may be. Second, the declaration of a false allegation follows a highly institutionalized procedure. The investigation of all rape complaints always involves a serious offer to polygraph the complainants and the suspects. Additionally, for a declaration of false charge to be made, the complainant must admit that no rape had occurred. She is the sole agent who can say that the rape charge is false. The police department will not declare a rape charge as false when the complainant, for whatever reason, fails to pursue the charge or cooperate on the case, regardless how much doubt the police may have regarding the validity of the charge. In short, these cases are declared false only because the complainant admitted they are false. Furthermore, only one person is then empowered to enter into the records a formal declaration that the charge is false, the officer in charge of records. Last, it should be noted that this department does not confuse reported rape attempts with completed rapes. Thus, the rape complainants referred to in this paper are for completed forcible rapes only. The foregoing leaves us with a certain confidence that cases declared false by this

police agency are indeed a reasonable--if not a minimal reflection of false rape allegations made to this agency, especially when one considers that a finding of false allegation is totally dependent upon the recantation of the rape charge.

We followed and investigated all false rape allegations from 1978 to 1987. A ranking police official notified us whenever a rape charge was declared false and provided us with the records of the case. In addition, the investigating officers provided any requested supplementary information so that we could be confident of the validity of the false rape allegation declarations.

Findings

Incidence of False Allegations

Widely divergent viewpoints are held regarding the incidence of false rape reporting (Katz and Mazur, 1979). For example, reports set the figure from lows of 0.25% (O'Reilly, 1984) and 1% (Krasner et al., 1976) to highs of 80-90% (Bronson, 1918; Comment, 1968) and even 100% (see Kanin, 1985). All of these figures represent releases from some criminal justice agency or are estimates from clinical practitioners. The extraordinary range of these estimates makes a researcher suspect that inordinate biases are at work.

Regarding this study, 41% (n = 45) of the total disposed rape cases (n = 109) were officially declared false during this 9-year period, that is, by the complainant's admission that no rape had occurred and the charge, therefore, was false. The incidence figure was variable from year to year and ranged from a low of 27% (3 out of 11 cases) to a high of 70% (7 out of 10 cases). The 9-year period suggests no trends, and no explanation has been made for the year-to-year fluctuation.

Although very little information exists regarding the characteristics of the complainant, some data can be offered. These false complainants are all white, largely of lower socioeconomic background, and the majority were modestly educated. Only three complainants had any education beyond high school. The mean age of these women was 22. On the basis of the limited information available, these women could not be distinguished from those whose complaints were recorded as valid.

The study of these 45 cases of false rape allegations inexorably led to the conclusion that these false charges were able to serve three major functions for the complainants: providing an alibi, a means of gaining revenge, and a platform for seeking attention/ sympathy. This tripartite model resulted from the complainants' own verbalizations during recantation and does not constitute conjecture. Of course, we are not asserting that these functions are mutually exclusive or exhaustive; rather, these rape recantations focused on a single factor explanation. A possible objection to these recantations concerns their validity. Rape recantations could be the result of the complainants' desire to avoid a "second assault" at the hands of the police. Rather than proceed with the real charge of rape, the argument goes, these women withdrew their accusations to avoid the trauma of police investigation.

Several responses are possible to this type of criticism. First, with very few exceptions, these complainants were suspect at the time of the complaint or within a day or two after charging. These recantations did not follow prolonged periods of investigation and interrogation that would constitute anything approximating a second assault. Second, not one of the detectives believed that an incident of false recantation had occurred. They argued, rather convincingly, that in those cases where a suspect was identified and interrogated, the facts of the recantation dovetailed with the suspect's own defense. Last, the policy of this police agency is to apply a statute regarding the false reporting of a felony. After the recant, the complainant is informed that she will be charged with filing a false complaint, punishable by a substantial fine and a jail sentence. In no case, has an effort been made on the part of the complainant to retract the recantation. Although we certainly do not deny the possibility of false recantations, no evidence supports such an interpretation for these cases.

Alibi Function

Of the 45 cases of false charges, over one-half (56%, n = 27) served the

complainants' need to provide a plausible explanation for some suddenly foreseen, unfortunate consequence of a consensual encounter, usually sexual, with a male acquaintance. An assailant is identified in approximately one half of these cases. Representative cases include the following:

> An unmarried 16-year-old female had sex with her boyfriend and later became concerned that she might be pregnant. She said she had been raped by an unknown assailant in the hopes that the hospital would give her something to abort the possible pregnancy.
>
> A married 30-year-old female reported that she had been raped in her apartment complex. During the polygraph examination, she admitted that she was a willing partner. She reported that she had been raped because her partner did not stop before ejaculation, as he had agreed, and she was afraid she was pregnant. Her husband is overseas.

The above cases are prototypical cases where the fear of pregnancy is paramount in motivating the rape charge. This theme is constant, only the scenario changes in that the lover is black, the husband is out of state on a job, the husband had a vasectomy, the condom broke. Only three cases deviated from this tradition:

> A divorced female, 25 years of age, whose parents have custody of her 4-year-old child. She lost custody at the time of her divorce when she was declared an unfit mother. She was out with a male friend and got into a fight. He blackened her eye and cut her lip. She claimed she was raped and beaten by him so that she could explain her injuries. She did not want to admit she was in a drunken brawl, as this admission would have jeopardized her upcoming custody hearing.
>
> A 16-year-old complainant, her girlfriend, and two male companions were having a drinking party at her home. She openly invited one of the males, a casual friend, to have sex with her. Later in the evening, two other male acquaintances dropped in and, in the presence of all, her sex partner "bragged" that he had just had sex with her. She quickly ran out to another girlfriend's house and told her she had been raped. Soon, her mother was called and the police were notified. Two days later, when confronted with the contradictory stories of her companions, she admitted that she had not been raped. Her charge of rape was primarily motivated by an urgent desire to defuse what surely would be public information among her friends at school the next day, her promiscuity.
>
> A 37-year-old woman reported having been raped "by some nigger." She gave conflicting reports of the incident on two occasions and, when confronted with these, she admitted that the entire story was a fabrication. She feared her boyfriend had given her "some sexual disease," and she wanted to be sent to the hospital to "get checked out." She wanted a respectable reason, i.e., as an innocent victim of rape, to explain the acquisition of her infection.

Revenge

Essentially, this category involved a false rape report as a means of retaliating against a rejecting male. Twenty-seven percent (n = 12) of the cases clearly seemed to serve this function. These rejections, however, ranged from the very evident cases of women who were sexually and emotionally involved with a reciprocating male to those women who saw themselves spurned from what was in reality the females' unilateral involvement. Regardless, these women responded with a false rape charge to perceived rejections. Because the suspect is always identified, the false allegations potentially pose the greatest danger for a miscarriage of justice. Examples of these types of cases are as follows:

> An 18-year-old woman was having sex with a boarder in her mother's house for a period of 3 months. When the mother learned of her behavior from other boarders, the mother ordered the

man to leave. The complainant learned that her lover was packing and she went to his room and told him she would be ready to leave with him in an hour. He responded with "who the hell wants you." She briefly argued with him and then proceeded to the police station to report that he had raped her. She admitted the false charge during the polygraph examination.

A 17-year-old female came to headquarters and said that she had been raped by a house parent in the group home in which she lived. A female house parent accompanied her to the station and told the police she did not believe that a rape had occurred. The complainant failed the polygraph examination and then admitted that she liked the house parent, and when he refused her advances, she reported the rape to "get even with him."

A 16-year-old reported she was raped, and her boyfriend was charged. She later admitted that she was "mad at him" because he was seeing another girl, and she "wanted to get him into trouble."

Attention/Sympathy-Getting Device

Although this device seems to be the most extravagant, use for which a false rape charge is made, it is also the most socially harmless in that no one was identified as the rapist. Approximately 18% (n = 8) of the false charges clearly served this function. The entire verbalization of the charge is, by and large, a fabrication without base. The following are typical examples:

An unmarried female, age 17, abruptly left her girlfriends in the park one afternoon allegedly to go riding with a young man, a stranger she met earlier that morning who wanted her to smoke marijuana with him. Later that day, she told her friends she was raped by this man. Her friends reported the incident to the police, and the alleged victim went along with the rape charge because "I didn't want them to know that I lied to them." She explained that she manufactured this story because she wanted the attention.

An unmarried female, age 17, had been having violent quarrels with her mother who was critical of her laziness and style of life. She reported that she was raped so that her mother would "get off my back and give me a little sympathy."

An unmarried female, age 41, was in post-divorce counseling, and she wanted more attention and sympathy from her counselor because she "liked him." She fabricated a rape episode, and he took her to the police station and assisted her in making the charge. She could not back out since she would have to admit lying to him. She admitted the false allegation when she was offered to be polygraphed.

Related Findings

In addition to the foregoing, certain other findings and observations relevant to false allegations warrant comment. First, false allegations failed to include accusations of forced sexual acts other than penile-vaginal intercourse. Not one complainant mentions forced oral or anal sex. In contrast, these acts were included in approximately 25% of the founded forcible rape complaints. Perhaps it was simply psychologically and socially more prudent for these women to minimize the humiliation of sexual victimization by not embroidering the event any more than necessary. This phenomenon has been observed previously (McDowell and Hibler, 1987).

Second, although the literature liberally refers to various extortion scams as responsible for false rape charging (Comment, 1968; MacDonald, 1973), no such cases were encountered or could even be recalled by members of the police agency. This type of case may very well be a period piece, or perhaps it was even then the exceptional case. Extraordinary attention would readily have been forthcoming since this theory nicely meshed with the position of prevailing authorities who stressed the omni-present threat of female cunning and stealth. One authority, (MacDonald, 1973), for example,

cited a 1918 article (Bronson) to illustrate a blackmail case since he never encountered one himself.

In a similar vein, no apparent case of pseudologia phantastica surfaced. The earlier view of a deluded complainant, tenaciously affirming her victimization, just does not appear here. These women were not inclined to put up a steadfast defense of their victimization, let alone pursue it into the courtroom. Recantation overwhelmingly came early and relatively easily. Certainly, false rape allegations can arise from a deluded condition but we failed to find indicators for what was once offered as the most common explanation for false rape allegation.

One of the most haunting and serious implications of false rape allegations concerns the possibility of miscarried justice. We know that false convictions occur, but this study only tells us that these false accusers were weeded out during the very early stages of investigation. However encouraging this result may be, we cannot claim that false charging does not incur suffering for the accused. Merely to be a rape suspect, even for a day or two, translates into psychological and social trauma.

Conclusions

We feel that these false accusations can be viewed as the impulsive and desperate gestures of women simply attempting to alleviate understandable conditions of personal and social distress and that, as an aggregate, labels connoting pathology, e.g., delusional states, are uncalled for. One can be tempted to pigeonhole this type of conduct since we view it as extreme, as deviant, as criminally reckless. At first glance, false rape allegation seems to be a rather extreme gesture to satisfy alibi, revenge, or attention needs. Practitioners in the mental health and legal professions, however, will readily recognize that these false rape reports are not really exceptional exaggerations in light of what people rather commonly do in order to satisfy these same needs in other contexts. Consider the extravagant and perjurious accusations that routinely pepper divorce and child custody proceedings, and the inordinate departures from the truth that have accompanied credentialed and respected political and corporate figures in their quest for recognition and office. And think of the petty and commonplace transgressions that people frequently verbalize as reasons for having committed homicide.

No evidence exists to suggest that something unique or defective is in the female condition that prompts such behavior. Rather, something biological, legal, and cultural would seem to make false rape allegations inevitable. If rape were a commonplace victimization experience of men, if men could experience the anxiety of possible pregnancy from illicit affairs, if men had a cultural base that would support their confidence in using rape accusations punitively, and if men could feel secure that victimization could elicit attention and sympathy, then men also would be making false rape accusations.

Most problematic is the question of the generalizability of these findings from a single police agency handling a relatively small number of cases. Certainly, our intent is not to suggest that the 41% incidence found here be extrapolated to other populations, particularly in light of our ignorance regarding the structural variables that might be influencing such behavior and which could be responsible for wide variations among cities. But a far greater obstacle to obtaining "true" incidence figures, especially for larger cities, would be the extraordinary variations in police agency policies (see Comment, 1968; *Newsweek*, 1983; Pepinsky and Jesilow, 1984); variations so diverse, in fact, that some police agencies cannot find a single rape complaint with merit, while others cannot find a single rape complaint without merit. Similarly, some police agencies report all of their unfounded rape cases to be due to false allegation, while other agencies report none of their unfounded declarations to be based on false allegation (Kanin, 1985). Some of these policies are really nothing more than statistical and procedural legerdemain. On the other hand, a degree of confidence exists that the findings reported here are not exaggerations produced by some sort of atypical population, that is, nothing peculiar exists about this city's population composition to suggest that an unusual incidence or patterning of false rape allegations would occur. This city is not a resort/reveling area or a center attracting a transient population of any kind, attributes

that have been associated with false rape reporting (Wilson, 1978). The major culprit in this city may well be a police agency that seriously records and pursues to closure all rape complaints, regardless of their merits. We may well be faced with the fact that the most efficient police departments report the higher incidence of false rape allegations. In view of these factors, perhaps the most prudent summary statement that is appropriate from these data is that false rape accusations are not uncommon. Since this effort is the first at a systematic, long-term, on-site investigation of false rape allegations from a single city, future studies in other cities, with comparable policies, must assess the representativeness of these findings.

Addenda

In 1988, we gained access to the police records of two large Midwestern state universities. With the assistance of the chief investigating officers for rape offenses, all forcible rape complaints during the past three years were examined. Since the two schools produced a roughly comparable number of rape complaints and false rape allegations, the false allegation cases were combined, n = 32. This represents exactly 50% of all forcible rape complaints reported on both campuses. Quite unexpectedly then, we find that these university women, when filing a rape complaint, were as likely to file a false as a valid charge. Other reports from university police agencies support these findings (Jay, 1991).

In both police agencies, the taking of the complaint and the follow-up investigation was the exclusive responsibility of a ranking female officer. Neither agency employed the polygraph and neither declared the complaint false without a recantation of the charge. Most striking is the patterning of the reasons for the false allegations given by the complainants, a patterning similar to that found for the non-student city complainants. Approximately one half (53%) of the false charges were verbalized as serving an alibi function. In every case, consensual sexual involvement led to problems whose solution seemed to be found in the filing of a rape charge. The complaints motivated by revenge, about 44%, were of the same seemingly trivial and spiteful nature as those encountered by the city police agency. Only one complainant fell into the attention/sympathy category. These unanticipated but supportive parallel findings on university populations suggest that the complications and conflicts of heterosexual involvements are independent of educational level. In fact, we found nothing substantially different here from those cases encountered by our city police agency.

References

Abelson, E. S. (1989). The invention of kleptomania. *Signs* 15: 123-143.

Bessmer, S. (1984). *The Laws of Rape*, Praeger, New York.

Bronson, F. R. (1918). False accusations of rape. *Am. J. Urol. Sex.* 14: 509-510.

Brownmiller, S. (1975). *Against Our Will*, Simon and Schuster, New York.

Clark, L., and Lewis, D. (1977). *The Price of Coercive Sexuality*, Woman's Press, Toronto.

Comment. (1968). Police discretion and the judgment that a crime has been committed. *Univ. Penn. Law Rev.* 117: 227-322.

Comment. (1970). The corroboration rule and crimes accompanying a rape. *Univ. Penn. Law Rev.* 118: 458-472.

Comment. (1973). Complainant credibility in sexual offense cases: A survey of character testimony and psychiatric experts. *J. Crim. Law and Crim.* 64: 67-75.

Deutsch, H. (1944). *The Psychology of Women*, Vol. 1, Grune and Stratton, New York.

Edwards, S. (1981). *Female Sexuality and the Law*, Martin Robertson, London.

Edwards, S. (1983). Sexuality, sexual offenses, and conceptions of victims in the criminal justice process. *Victimology* 8: 113-130.

Freud, S. (1933). *New Introductory Lectures on Psychoanalysis*, Norton, New York.

Grablewski, J. T. (1958). Necessity and corroboration of prosecutrix's testimony in prosecution for rape. *Villanova Law Rev.* 3: 220-222.

Grano, J. D. (1990, Spring). Free speech v. The University of Michigan. *Acad Quest.*, pp. 7-22.

Guttmacher, M., and Weihofen, H. (1952). *Psychiatry and the Law*, Norton, New York.

Hibey, R. A. (1973). The trial of a rape case: An advocate's analysis of corroboration, consent, and character. *Am. Crim. Law Rev.* 11: 309-334.

Horney, K. (1933). *Feminine Psychology*, Norton, New York.

Hursch, C. (1977). *The Trouble with Rape*, Nelson-Hall, Chicago.

Jay, D. R. (1991). Victimization on the college campus: A look at three high-profile cases. *Campus Law Enforce*, 35-37.

Juliver, M. (1960). Psychiatric opinions as to credibility of witnesses: A suggested approach. *Cal. Law Rev.* 48: 648-683.

Kanin, E. J. (1982). Female rape fantasies: A victimization study. *Victimology* 7: 114-121.

Kanin, E. J. (1985). Unfounded rape. Paper presented at the Academy of Criminal Justice Sciences annual meeting, Las Vegas, NV.

Katz, S., and Mazur, M. (1979). *Understanding the Rape Victim*, Wiley, New York.

Krasner, W., Meyer, L., and Carroll, N. (1976). *Victims of Rape*, U. S. Government Printing Office, Washington, DC.

LaFree, G. (1989). *Rape and Criminal Justice*, Wadsworth, Belmont, CA.

Lowe, M. (1984). The role of the judiciary in the failure of the sexual offenses (amendment) act to improve the treatment of the rape victim. In Hopkins, J. (ed.), *Perspectives on Rape and Sexual Assault*, Harper and Row, London, pp. 67-88.

MacDonald, J. (1971). *Rape Offenders and their Victims*, Charles C Thomas, Springfield, IL.

MacDonald, J. (1973). False accusations of rape. *Med Aspects Hum. Sex.* 7: 170-193.

Menninger, K. A. (1933). Cited in Wigmore, J. (1940). *Evidence in Trials at Common Law*, Little Brown, Boston.

McDowell, C. P., and Hibler, N. S. (1987). False allegations. In Hazelwood, R. R. and Burgess, A. W. (eds.), *Practical Aspects of Rape Investigation*, Elsevier. New York.

Newsweek. (May 16, 1983). Burying crime in Chicago.

O'Reilly, H. J. (1984). Crisis intervention with victims of forcible rape: A police perspective. In Hopkins, J. (ed.), *Perspectives on Rape and Sexual Assault*, Harper & Row, London, pp. 89-103.

Payton, G. T. (1967). *Patrol Procedures*, Legal Book, Los Angeles.

Pepinsky, H. E., and P. Jesilow (1984). *Myths that Cause Crime*, Seven Locks Press, Cabin John, MD.

Russell, D. E. H. (1984). *Sexual Exploitation*, Sage, Beverly Hills.

Schiff, A. (1969). Statistical features of rape. *J. Forensic Sci*; 14; 102-110.

Schiff, A. (1971, October). Rape needs a special examination. *Emergency Med.*, pp. 28-29.

Sherwin, R.V. (1973). Commentary. *Med Aspects Hum. Sex.* 7: 193.

Smith, T. L (1953-1954). Competency of female child as prosecutrix in case of sexual offenses. *Ala. Law Rev.* 6: 96-99.

Weihofen, H. (1959). Compensation for victims of criminal violence. *J. Public Law* 8: 209-218.

Wigmore, J. (1940). *Evidence in Trials at Common Law*, Little Brown, Boston.

Wilson, P. R. (1978). *The Other Side of Rape*, University of Queensland Press, St. Lucia, Queensland, Australia.

The Rank Order Scoring System: Replication and Extension with Field Data

Donald J. Krapohl, Donnie W. Dutton, and Andrew H. Ryan

Abstract

In this journal Honts and Driscoll (1987) reported the development of a scoring system for the single-issue Zone Comparison Technique format. It was called the Rank Order Scoring System (ROSS), and was based on the ranking of individual physiological responses, and determining algebraically from those ranks which category of question, relevant or comparison, elicited more physiological arousal, thereby permitting inferences of truthfulness or deceptiveness. Honts and Driscoll tested the method with laboratory cases, and finding it very effective, urged more research on the ROSS. The present study is a replication and extension of that work, using field cases instead of mock crime cases. Results here were also encouraging. Accuracy in the first cross validation, using 150 nondeceptive and 150 deceptive cases, was 80.3% with inconclusives, and 91.6% without them. A subset of that sample, consisting of 100 cases, was used in a second cross validation where the ROSS was directly compared to traditional 7-position manual scoring. Traditional scoring was 69.7% accurate including inconclusives, and 90.5% correct without them. ROSS produced an average of 75.0% correct decisions with inconclusives, and 88.2% without them. There were no significant differences in accuracy or rates of inconclusive results. These results suggest that the ROSS may be a valuable alternative to traditional 7-position scoring, especially in contested evidentiary cases. Since ROSS uses measured features, it holds the potential for exceptional inter-scorer reliability.

Key words: 7-position scoring, Horizontal Scoring System, rank order analysis, Rank Order Scoring System, reliability, ROSS, validity

The discipline of polygraphy has expended no small effort over its history in the exploration and development of analytical methods for scoring data, with the aim of maximizing decision accuracy. It has not been an easy task. One of the enduring challenges for polygraphy has been the sheer complexity of the physiological data, a complexity that has invited an unwelcome element of subjectivity to interpretation of the data. Striving to impose order on the data, several scoring methodologies have been spawned by polygraph practitioners. All have worked better than chance, and some better than others. The quantification rules for those systems have spanned the spectrum from the elegant to the baffling, and only a few have started with empirical findings and developed scientifically defensible scoring and decision rules (i.e., Bell, Raskin, Honts, & Kircher, 1999; Honts & Driscoll, 1987, Krapohl & McManus, 1999).

Acknowledgments

The authors are grateful for the assistance of the polygraph examiners from the Hillsborough County (FL) Police Department for volunteering to be scorers in this project: Eugene Godfrey, Det. Larry V. Lingo, and Det. Herbert Metzgar. We also appreciate the help of Kay Williams for managing the data. We'd like to extend a special thanks to Dr. Charles Honts, for his insightful review of the early manuscript. Dr. Stuart Senter was also extremely helpful with the data analysis.

The authors are with the US Department of Defense Polygraph Institute, Ft. Jackson, South Carolina. Donald Krapohl is assigned to the Research Division, Donnie Dutton is an Instructor, and Dr. Andrew Ryan is the Chief of Research Division.

The opinions expressed in this article are not necessarily those of the Department of Defense or the US Government. Reprint requests should be sent to: Donald Krapohl, PO Box 10411, Ft. Jackson, SC 29207, or via e-mail to dkrapohl@aol.com.

Some writers have advocated the use of rank order scoring systems as one method of minimizing subjectivity. In general, rank ordering entails the assignment of ordinal values to responses based on magnitude or intensity of the responses. If the same stimuli are repeated, responses that continue to be greater than others will accumulate the larger ranks. The sums of ranks across presentations of the same stimuli can provide an index of how much reactivity is associated with particular stimuli.

Rank ordering has much to commend it, including simplicity of rules and greater objectivity. The use of ranks instead of other types of response comparisons should decrease inter-scorer differences, since the only judgment is that of relative size of measurable features. Various forms of rank order scoring first appeared in the detection of deception literature for the electrodermal channel (Lykken, 1959; Suzuki, Watanabe, Ohnishi, Matsuno, & Arasuna, 1973; Suzuki, Ohnishi, Matsuno, Arasuna, 1974), followed by the respiration channel (Timm, 1982). Soon afterward, polygraph practitioners began to look at rank order systems for all of the traditional polygraph channels; respiration, blood volume, and the electrodermal response (Gordon & Conchetti, 1987; Honts & Driscoll, 1987). The polygraph literature now records four data ranking systems applicable to the field: Lykken Scoring (Lykken, 1959; 1998), the Horizontal Scoring System (Gordon, 1999; Gordon & Cochetti, 1987), the Rank Order Scoring System (Honts & Driscoll, 1987; Honts & Driscoll, 1988), and Rank Order Analysis (Miritello, 1999).

To date, only two ranking methods have been tested empirically with traditional polygraph recordings on single-issue examinations utilizing the Comparison Question Technique (CQT); the Rank Order Scoring System (ROSS), and the Horizontal Scoring System. Accuracy for the ROSS (Honts & Driscoll, 1987) was 90.2% without inconclusives, and it had an inconclusive rate of 31.7%. The 7-position scores from the same cases produced an accuracy of 86.6% without inconclusives, and an inconclusive rate of 25.0%. Honts and Driscoll suggested that the ROSS, with its use of simpler and more objective scoring rules, should make this system easier to teach polygraph students, as well as convey to a lay public. It would also be more familiar and acceptable to behavioral scientists, and should improve inter-scorer agreement.

Honts and Driscoll (1987) issued three caveats regarding their scoring system. First, their findings were restricted to laboratory data only, leaving questions as to whether ROSS would generalize to the field. Second, Honts and Driscoll regarded their cutting scores for ROSS as tentative, and that further research would be needed to determine whether they were optimal. Third, Honts and Driscoll did not know the efficacy of the ROSS in a mixed-issue examination, and urged caution using the ROSS with these types of cases.

The Horizontal Scoring System (Gordon & Cochetti, 1987) was more recently tested on field data in Egypt (Gordon, Fleisher, Morsie, Habib & Salah, 2000). Gordon et al collected 576 cases conducted by the Egyptian government between 1998 and 1999, and confirmed ground truth in 309 of the cases. Confirmation of ground truth was established by confession of the examinee, or judicial conviction. Gordon reported only one error among the 309 cases, and only 25 inconclusives, for an accuracy without inconclusives of 99.6%.

The ROSS differs from the Horizontal Scoring System primarily in the features it uses in the ranking. The ROSS assigns ranks to responses according to the features found most diagnostic in the detection of deception by Kircher and Raskin (1988). For respiration, it is the length of the waveform tracing for 10 seconds beginning at question onset, and this feature is called respiration line length (RLL) (Timm, 1982). RLL captures both decreases in respiration amplitude and changes in the inhalation/exhalation ratio (I/E) at once, because they both cause the RLL to become shorter. Shorter RLLs are associated with deception. For both blood volume (BV) and electrodermal activity (EDA), the diagnostic feature is the peak amplitude after stimulus onset. This ensemble of diagnostic features, sometimes called the "Kircher features," may perform at least as well as traditional scoring features (for a review, see Swinford, 1999) with

the 7-position scoring system (Harris, 2001). The Kircher features entail straightforward measurements that eliminate all subjectivity, and therefore it was chosen for the present study over the Horizontal Scoring System. The scoring rules for the Horizontal Scoring System are less precise, and have not yet been independently validated.

Like all rank order systems, ROSS ranking is done within charts, and within channels. EDA and BV phasic responses are assigned numbers according to magnitude, with larger responses receiving larger ranks. The smallest response receives a "1", and the largest response is assigned a number that corresponds with the number of questions being ranked. For example, in a three-question Zone Comparison Technique format, there are three relevant and three comparison questions, for a total of six questions. In this example the largest EDA or BV response would receive a rank of 6. In the case of ties, the larger rank is given to the response with the greater duration, and in the case of the EDA, the greater complexity. For distorted or artifacted tracing segments, the lowest rank (1) is assigned. In respiration, the greatest reaction, and hence the shortest RLL, receives the largest rank. Since respiration is recorded twice with conventional instruments, the ranks are averaged between those two channels. All ranks are summed by question type, with one sum for the relevant questions and another for the comparison questions. As the final step, the sum of the relevant question ranks is subtracted from the sum of the comparison question ranks, rendering a single value. Like the final value in traditional 7-position scoring, this number is compared to the polygraph decision thresholds. Honts and Driscoll (1987) tested all cutting scores symmetrical around 0, from +/-1 to +/-19, and found the best cutting scores at +/-13.

Though rank ordering of responses has some appeal among those who appreciate the value of limiting subjectivity in manual scoring, there are reasons to suspect that, even with optimal cutting scores, a rank order scoring system should not enjoy equal accuracy with traditional 7-position scoring. This is due to the forcing of equal-distant ranks on responses that are not equally distant from each other. For example, suppose one had three electrodermal responses (EDRs) on a particular chart, with sizes of 2 units, 3 units, and 99 units. These responses would receive ranks of 1, 2 and 3 with ROSS, respectively, thereby rescaling the data so that the true differences between the first two EDRs and the last two EDRs are obscured. Because the ranks are equally spaced, they significantly conceal the actual distance between the original response magnitudes in this example. The work of Ansley and Krapohl (2000) showed how this characteristic of rank ordering might influence accuracy. They investigated the incidence of 22 tracing criteria taught at DoDPI for manual numerical scoring, and uncovered an unexpected trend. It was found that each of the criteria was occurring with equal frequencies for all question types regardless of whether the cases came from deceptive or nondeceptive examinees. For example, if there were 100 incidences of respiration suppression on relevant questions for deceptive examinees, there were also about 100 respiration suppressions on the comparison questions for those same deceptive examinees, and 100 respiration suppressions taking place on both relevant and comparison questions of nondeceptive examinees. This finding had important implications. If the incidence of responses was equal for liars and truthtellers for both types of questions, then the information that allows examiners to correctly classify deceptive and nondeceptive examinees with only chart tracings must be found in the intensity of the responses, not just the presence or absence of the responses. Therefore, if intensity information is important, even essential, and rank ordering obscures or dilutes it with equal size ranks, one might reasonably expect a reduction of accuracy with ranking versus 7-position scoring, which better considers intensity information. Honts and Driscoll (1987) did not find ROSS to have lower accuracy than 7-position scoring with their lab data, but this is a testable question, and one that is addressed by the design of the present project.

We devised three goals for this effort: to determine the optimal ROSS cutting scores for single-issue field polygraph data; to use those cutting scores to test the decision accuracy of ROSS with new field data, and; to compare the accuracy of ROSS with traditional 7-position

scoring when the same cases are scored with the two techniques. The working hypothesis was that ROSS, despite the advantage it might afford in improving inter-scorer agreement, would not deliver accuracy equivalent to traditional 7-position scoring.

Method

PDD cases

Three separate samples of PDD cases were used here. All examinations were conducted with the Axciton digitized polygraph (Axciton Systems, Houston, TX) by local, state or federal polygraph examiners in the field. The samples are described below.

Training Set. As part of an algorithm validation project, Blackwell (1999) randomly selected 35 nondeceptive and 65 deceptive cases taken from the DoDPI confirmed case database. All cases were Zone Comparison Technique (ZCT) field examinations conducted according to rules published by the US Department of Defense Polygraph Institute (DoDPI, 1999). These cases were used in the present project for the development of cutting scores.

Cross Validation Set 1. Krapohl and McManus (1999) used 300 cases from the DoDPI confirmed case database to develop the scoring rules for the Objective Scoring System (OSS). All were three-chart three-question ZCT examinations conducted according to DoDPI testing protocol (DoDPI, 1999). Half of the cases were confirmed deceptive, and the other confirmed nondeceptive. Confirmation of ground truth was in the form of confessions from the examinee, confession from someone else who exculpated the examinee, urinalysis, or other reliable forensic tests. Cases that depended on forms of confirmation that are not generally considered as reliable (eye witness accounts, prosecution decisions, or judicial outcomes) were excluded. These 300 cases were used for the first cross validation of ROSS.

Cross Validation Set 2. This set was a subset of Cross Validation Set 1, using 50 confirmed deceptive and 50 confirmed nondeceptive cases drawn at random. These cases were used specifically to compare the ROSS with traditional 7-position scoring.

Software

Two software packages were used. The first was the Extract program (Applied Physics Lab, Johns Hopkins University, Version 3.0), a software package developed for DoDPI. This program measures and outputs the Kircher features from digitized polygraph data into a text file. The second software package was Excel 97 (Microsoft Corporation), an electronic spreadsheet. Excel was used to rank the data generated from the Extract program, perform the ROSS, and apply the statistical treatments.

Human Scorers

Three experienced polygraph examiners with a law enforcement agency agreed to serve as independent scorers of the cases in the Cross Validation Set 2. All were knowledgeable with the 7-position scoring procedures. They were kept blind to ground truth for the cases, in addition to base rates, case facts, and the decisions of the original examiner or each other.

Development of Cutting Scores

The Training Cases were subjected to the ROSS, as developed by Honts and Driscoll (1987). An abbreviated explanation of the ROSS is provided here. Each of the four channels of data was ranked separately, with greater reactions receiving larger ranks. Ranks of the two pneumograph tracings were averaged to a single rank for each question so that respiration would receive weighting equal with the electrodermal and cardiograph channels. The ranks of all channels of data for all charts for the relevant questions were summed, and those of the comparison questions were separately summed. The summed ranks of the relevant questions were subtracted from the summed ranks from the comparison question, and this difference became the score for that case.

The simple measurements of the Kircher features lack any metric for duration or complexity, factors Honts and Driscoll (1987) used in the manual ROSS for breaking of ties. In cases of ties with the Training Set and Cross Validation cases, the ranks were averaged among the tied responses. For example, if the two largest responses were tied, they would each receive the average of the highest two ranks available. If there were

six questions ranked, the highest two ranks would be 6 and 5, and they average to 5.5. Similarly, when the bottom three responses were of equal magnitude, they would each receive the average of ranks 3, 2 and 1, or the rank of 2. We did not know whether this departure from the Honts and Driscoll scoring protocol would affect accuracy. If there were an effect, we believed it would be more likely to reduce accuracy than improve it because there is diagnostic value in response duration (Kircher & Raskin, 1988) that is not captured by our method of breaking ties.

The purpose of the Training Set was to establish cutting scores for testing against the two cross validation sets. Honts and Driscoll (1987) had systematically tested 19 pairs of cutting scores that were symmetrical around 0. However, previous research suggests that physiological responding in field cases is shifted in the deceptive direction for both the deceptive and nondeceptive subjects (Franz, 1988; Kircher & Raskin, 1988; Krapohl & McManus, 1999; Krapohl, 2000). In the field, deceptive examinees respond stronger to the relevant questions, and nondeceptive respond more weakly to comparison questions than is predicted from laboratory data. Therefore, using 0 as the center point around which to test cutting scores with field data would probably not produce equal accuracies for deceptive and nondeceptive examinees. Rather, we would expect that the true and false positive outcomes would be high at a 0 center point.

Based on what is known regarding the response patterns of deceptive and nondeceptive examinees in the field, there was no assumption that balanced cutting scores would be symmetrical around 0. To find the center point for these data, descriptive statistics were calculated for the scores for the 35 nondeceptive cases and 65 deceptive cases in the Training Set. The deceptive cases produced a mean score of -23.01, and standard deviation of 20.02. Nondeceptive cases had a mean score of +8.17, and standard deviation of 17.61. Using these descriptive statistics, it was determined that -7 was the nearest whole number intersection point, a single cutting score where decision accuracy should be equal for deceptive and nondeceptive cases. With the center point established, all symmetrical cutting scores around -7, from -6/-8 to +1/-15, were tested. At the chosen cutting scores, -1/-13, the weighted average for the inconclusive rate of the deceptive and nondeceptive cases was 16.2%, and the accuracy without inconclusives was 88.6%.

Here it is important to document the rationale for selecting this particular pair of cutting scores. The validity standard of the American Society for Tests and Materials for evidentiary polygraph examinations (1998) was used as the reference: inconclusives must not exceed 20%, and accuracy without inconclusives must be 90% or greater. Moving the cutting scores outward from -1/-13 with the Training Set data increased the inconclusive rate nearly 4 percentage points over the ASTM standard of 20%. Since the averaged accuracy rate with the -1/-13 cutting scores was just below the 90% standard (88.6%) by only one case, we decided to accept the -1/-13 thresholds. Therefore, in this study, all final scores from -13 and lower were called Deception Indicated (DI), and -1 or greater were called No Deception Indicated (NDI). All others were called inconclusive. Spot scores (Light, 1998), that is, the use of decision rules for individual question scores, were not considered.

Cross Validations

The Extract software was used to measure the Kircher features with all of the cross validation cases, and the ROSS protocol as described in the previous section was followed. The rejection region for all statistics was .05.

Results

ROSS

Table 1 shows the accuracies of the Training Set, Cross Validation Set 1, and Cross Validation Set 2. Cross Validation Set 1 produced 241 correct decision, 22 errors, and 37 inconclusive results out of 300 cases. Overall accuracy was significantly greater than chance ($z = 10.49$, $p<0.000$. The proportions of accurate decisions for the deceptive and nondeceptive cases were not significantly different ($z = 0.15$, $p>0.05$). The average inconclusive rate for the 300 cases was 12.3%, and the overall accuracy without inconclusives was 91.6%.

Cross Validation Set 2 yielded 75 correct decisions, 10 errors, and 15 inconclusives out of 100 cases. Overall accuracy was significantly greater than chance ($z = 5.00$, $p<0.000$). The inconclusive rate averaged 15.0%, and accuracy without inconclusives was 88.2%. The proportions of accurate decisions for the deceptive and nondeceptive cases were not significantly different ($z = 0.69$, $p>0.05$).

Table 1. Accuracy for the Training Set, and Cross Validation Sets 1 and 2 for the ROSS with cutting scores of -1/-13.

	Deceptive				Nondeceptive				Total	
	Correct	Error	Inc	N	Correct	Error	Inc	N	Average Percent Correct	Percent Correct w/o Incs
Training Set	52	5	8	65	24	4	7	35	74.3%	88.6%
Cross Validation Set 1	121	10	19	150	120	12	18	150	80.3%	91.6%
Cross Validation Set 2	39	6	5	50	36	4	10	50	75.0%	88.2%

Inc=Inconclusive

Table 2. Number of correct, incorrect, and inconclusive decisions for three scorers using the 7-position scoring system, and the comparable accuracy for the ROSS for the Cross Validation Set 2.

	Deceptive (n=50)			Nondeceptive (n=50)			Percent Correct		Inc Rate
	Correct	Error	Inc	Correct	Error	Inc	With Incs	Without Incs	
Scorer 1	30	5	15	34	1	15	64.0%	91.4%	30.0%
Scorer 2	33	7	10	42	2	6	75.0%	89.3%	16.0%
Scorer 3	34	6	10	36	1	13	70.0%	90.9%	23.0%
Average	**32.3**	**6.0**	**11.7**	**37.3**	**1.3**	**11.3**	**69.7%**	**90.5%**	**23.0%**
ROSS	39	6	5	36	4	10	75.0%	88.2%	15.0%

Inc = Inconclusive

Human Scorers

Table 2 lists the rates of correct, incorrect, and inconclusive decisions for the three scorers and the ROSS with the same cases. For the human scorers, cutting scores of +/-6 were used, and the ROSS data from the Cross Validation Set 2 are carried over from Table 1. The average accuracy for the three human scorers was 69.7% including inconclusive outcomes, and 90.5% when inconclusives are excluded. Comparable results for the ROSS were 75.0% and 88.2%, respectively. Proportions of correct decisions for the average human scorers and the ROSS were not significantly different including inconclusives ($z=0.84$, $p>.05$), nor when inconclusive were excluded ($z=0.47$, $p>.05$). The human scorers averaged 23.0% inconclusives, while the ROSS produced 15.0%. These differences also failed to achieve statistical significance ($z=1.44$, $p>.05$).

Means and standard deviations for the scores produced by the three scorers using the 7-position scoring system, and the equivalent statistics for the ROSS, are found in Table 3. A within-subjects ANOVA for standardized scores from the ROSS and the 7-position scorers found no significant differences overall [$F(3,297)=0.0$, $p>0.05$], nor any differences among the scores for deceptive cases [$F(3,147)=0.344$, $p>0.05$] or nondeceptive cases [$F(3,147)=0.291$, $p>0.05$].

Table 3. Mean scores and standard deviations for deceptive and nondeceptive cases for three scorers using the 7-position scoring system, and for the ROSS for the Cross Validation Set 2.

	Deceptive		Nondeceptive	
	Mean Score	sd	Mean Score	Sd
Scorer 1	-6.54	8.94	7.62	6.68
Scorer 2	-8.88	10.16	10.12	7.16
Scorer 3	-11.68	15.38	12.16	11.11
ROSS	-24.65	18.97	7.40	17.06

sd=standard deviation

Inter-scorer agreement of decisions among the three traditional 7-position scorers and the ROSS are found in Table 4. For the 7-position scorers, scores of greater than +5 were called NDI (No Deception Indicated, scores lower than -5 were DI (Deception Indicated), and all others were Inconclusive. For the ROSS, scores greater than -2 were NDI, lower than -12 were DI, and all other Inconclusive.

Table 4. Proportion of agreement for decisions of the ROSS and the three scorers using the 7-position scoring system for the Cross Validation Set 2.

	Scorer 1	Scorer 2	Scorer 3	Ground Truth
ROSS	0.61	0.67	0.67	0.75
Scorer 1		0.71	0.70	0.64
Scorer 2			0.71	0.75
Scorer 3				0.70

Discussion

There were two findings worthy of special comment in the present data. First, decision accuracy using the ROSS was not different from that achieved by traditional 7-position scoring, an outcome that we had not predicted. It may be that intensity information is not as important as we had anticipated, or perhaps the ROSS capitalizes on a more rigid structure that reduces variability to produce better accuracy. Honts and Driscoll (1987) had also found similar accuracy for the ROSS and 7-position scoring with laboratory data, and the present findings give credence to both their conclusions and ours. If the ROSS delivers significantly improved inter-scorer agreement over 7-position scoring, an anticipated but as-yet unproven result, it could be the method of choice for at least evidentiary polygraph examinations. Given the simple rules and objective measurements of ROSS, it could be easy to integrate into computer polygraph operating software, something far less challenging than the programming of the semi-objective 7-position scoring system.

The second unexpected result was that manual scoring and the ROSS produced different balances in scores with Cross Validation Set 2. Traditional scoring of the present cases showed a nearly perfect balance of scores for deceptive and nondeceptive examinees, with a center-point for all cases falling just slightly above 0 (+0.94). Decision accuracy was slightly better for the nondeceptive cases than with deceptive cases for all manual scorers. This pattern of accuracy runs contrary to most field research, where the center-point of scores is almost always a strongly negative value, and detection of deception is better than the detection of truthfulness. In contrast, the ROSS data center-point with the same data was well into the negative range (-8.63), the expected direction.

Because the ROSS depends on absolute measurements of the tracing features, and the 7-position scoring system used here entails some subjectivity, it suggests that the absence of a shift of scores in the negative direction for the manual scorers is a function of their scoring rules, rather than a reflection of the actual pattern in the physiological data. Examiners did not assign more negative scores, despite the evidence that the underlying physiological phenomenon is clearly shifted in the negative direction. Because all of the scorers worked in the same polygraph program, we suspected that there may have been a common, but unknown factor influencing the scoring of the cases. At the conclusion of this study we contacted the senior polygraph examiner who participated in the project, and solicited a comment regarding their better performance with nondeceptive cases than with deceptive cases. He reported that all examiners in this study tended to give the examinee "the benefit of the doubt" when assigning scores. In other words, when scorers were less than certain, scores in a more positive direction were chosen. This was a practice in their polygraph unit, with the goal of being fair to examinees.

This scoring behavior may have something in common with examiner expectations, a source of scoring bias investigated by Elaad, Ginton and Ben-Shakhar (1994), in which examiners permitted personal beliefs about cases to influence the assignment of scores to reactions. Examiners' scores were shifted in the Elaad et al study according to examiner expectations of the examinee's guilt or innocence, though there was no effect found on examiner decision error. In the present study, both scores and decisions of the manual scorers were moved away from the deceptive direction as compared to the underlying pattern in the physiological tracings. These findings raise questions about the reliability of traditional manual scoring systems, and the related issue of using semi-objective scoring systems when more objective scoring systems are available, such as ROSS or automated algorithms.

Because the reliability (inter-scorer agreement) of a scoring system establishes the ceiling for its validity (accuracy), the more objective and rigid the scoring system is, the higher the potential validity can be. Complex or vague scoring rules can reduce inter-scorer reliability, as can inconsistent adherence to those rules by the scorers. Some polygraph educators blithely dismiss the wide difference in scores sometimes seen among scorers as simply being "a matter of degree" of the

polygraph scorings, and do not recognize it as the threat to the validity of the technique that it represents. Simply put, this type of variability among scorers erodes the possible validity of polygraphy. Given the simplicity of the ROSS, and the objective scoring rules, it is far less vulnerable to potentially lower levels of scorer reliability. It provides a promising method for increasing examiner reliability in the field.

Summary

The present study was a replication and extension of the Honts and Driscoll (1988) Rank Order Scoring System study. Using a training sample of 100 field cases we determined that cutting scores of -1/-13 would result in an accuracy rate and inconclusive rate that were statistically equivalent to the ASTM standards for evidentiary examinations. A cross validation with a sample of 300 cases also showed accuracies that met the ASTM standard for evidentiary examinations. A subsample of 100 cases was subjected to the ROSS and traditional manual scoring. The results indicated statistically equivalent performance for both systems. The existing research finds that the ROSS meets the ASTM scoring standard now under consideration, including the requirement for research replication, and the ROSS should now be considered for field use. Future research should compare the inter-scorer reliability of the ROSS and traditional 7-position scoring.

References

American Society for Tests and Materials (1998). *Standard Practice for Conduct of Research in Psychophysiological Detection of Deception (Polygraph).* Designation E 1954 - 98.

Ansley, N., & Krapohl, D.J. (2000). The frequency of appearance of evaluative criteria in field polygraph charts. *Polygraph*, 29(2), 169-176.

Bell, B.G., Raskin, D.C., Honts, C.R., & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.

Department of Defense Polygraph Institute. (1999). Zone comparison test. Ft. Jackson, SC.

Ellad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, 7, 279-292.

Franz, M. (1988). Relative contributions of physiological recordings to detect deception. Technical Report contract # MDA904-88-M-6612. Argenbright Polygraph, Inc., Atlanta, GA.

Gaines, K. (in progress). Cardio Element Analysis: A novel approach to the manual scoring of the cardiovascular signal.

Gordon, N.J. (1999). The Academy for Scientific Investigative Training's horizontal scoring system and examiner's algorithm for chart interpretation. *Polygraph*, 28(1), 56-64.

Gordon, N.J., & Cochetti, P.M. (1987). The horizontal scoring system. *Polygraph*, 16(2), 116-125.

Harris, J. (2001). Evaluation of the DoDPI Polygraph Chart Assessment Criteria. Report to the Scientific Review Committee of the DoD Polygraph Institute.

Honts, C.R., & Driscoll, L.N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16(4), 241-257.

Honts, C.R., & Driscoll, L.N. (1988). A field validity study of the Rank Order Scoring system (ROSS) in multiple-issue control question tests. *Polygraph*, 17(1), 1-16.

Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73(2), 291-302.

Krapohl, D.J., & McManus B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28(3), 209-222.

Krapohl, D.J. (2000). An assessment of the Total Chart Minutes Concept with field data. *Journal of the American Association of Police Polygraphists*, Oct., 31-37.

Light, G.D. (1999). Numerical evaluation of the Army Zone Comparison Test. *Polygraph*, 18(1), 37-45.

Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43(6), 385-388. Reprinted in *Polygraph*, 7(2), (1978), 123-128.

Miritello, K.M. (1999). Rank order analysis. *Polygraph*, 28(1), 74-76.

Suzuki, A., Watanabe, S., Ohnishi, K., Matsuno, K., & Arasuna, M. (1973). The objective analysis of GSR in the detection of deception: An analysis of GSR amplitudes in terms of rank scores. *Reports of the National Research Institute of Police Science*, 26(4), 237-245.

Suzuki, A., Ohnishi, K., Matsuno, K., & Arasuna, M. (1974). Amplitude rank score analysis of GSR in the detection of deception: Detection rates under various examination conditions. *Reports of the National Research Institute of Police Science*, 27, 142-148. Reprinted in *Polygraph*, 8(3), 242-252.

Timm, H.W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10(1), 47-51.

Spies and Lies

John F. Sullivan

The arrest of FBI agent Robert Philip Hanssen and the FBI's subsequent decision to expand its use of polygraph examinations have once again surfaced questions about the use of polygraph in the world of espionage. Why wasn't Hanssen tested has been one of the more frequently asked questions, and my answer is that it is not the FBI's policy to test its employees after they have been hired. At the time Hanssen entered on duty with the Bureau, applicants were not polygraphed, and he and many of his contemporaries have never been polygraphed.

More often than not, a discussion of the Hanssen case evolves into a discussion of Rick Ames, the CIA's infamous turncoat, and more questions about the use of polygraph in the world of espionage. Ames' arrest in 1994 focused a lot of attention on the CIA's polygraph program and raised questions about the CIA's use of polygraph, some of which I hope my comments will answer.

The CIA is, and has been since 1948, the intelligence community's primary user of the polygraph. In 1948, Director of Central Intelligence (DCI) Roscoe Henry Hillenkoetter authorized the use of polygraph on an experimental and voluntary basis. In 1949, the suggestion was made that recruited assets of the Directorate of Plans (DP), the CIA's clandestine service, be afforded polygraph tests

Thirty-one years as a CIA polygraph examiner have provided me with experiences that make it possible to discuss, and comment on from first-hand knowledge, how polygraph is used in the CIA's overt and operational arenas.

In the overt arena, CIA polygraph examiners test applicants for employment, industrial contractors working on CIA projects, staff employees as part of their periodic reinvestigations, and military detailees assigned to the CIA. Also, staff employees, against whom specific allegations have been made, on occasion, are afforded what are known as Specific Issue Polygraph (SIP) examinations.

Between 1949 and 1954, more than 100 staff employees had their CIA security clearances revoked, and several hundred applicants for employment with CIA were denied clearances based on polygraph-derived information. As a result, polygraph became an integral part of the CIA's clearance process.

Currently, there are several suits pending against the CIA by people who claim to have been falsely accused of wrongdoing during their polygraph tests. Those tests were conducted in the overt arena, and I feel it would be inappropriate to comment on overt testing at this time. Therefore, this article will focus on the operational arena or "ops testing," as covert testing is known.

Ops tests by CIA examiners are those tests conducted on agents and assets recruited by CIA case officers, and comparing polygraph testing in the overt arena with ops testing is comparing apples and oranges. The same instrumentation is used and the criteria for interpreting test results are the same, but any other similarities are coincidental.

Over the years, one aspect of ops testing that has caused a great deal of consternation and frustration among CIA

John Sullivan is a retired federal examiner. His book *Of Spies and Lies* is due for release in April 2002.

Requests for reprints can be sent to 1774 Wainwright Dr., Reston, Va. 20190.

examiners is the adversarial relationship that exists between case officers and polygraph examiners. It is the nature of the beast. Polygraph examiners' successes can be, and often are, perceived as case officers' failures. Recruited agents are the measurement of success for a case officer. A polygraph examiner is often put in the position of having to tell a case officer that a potential, or actually recruited, agent is not who he or she claims to be. On occasion, a polygraph examiner has to tell a case officer that one of his assets is a double agent (working for the other side).

In a perfect world, case officers would thank an examiner for determining that an agent is a fabricator, or worse. In the real world, when that happens, the case officer's stable of agents is depleted, his or her handling skills may be brought into question, previously reported information may have to be retracted, and the case officer is seldom pleased.

This is true even when an examiner uncovers a double agent. It is usually a case of "kill the messenger" and one of the real downsides of working in the operational world.

Another fact of life in ops testing is the lack of verified biographic information on the person being tested. In the overseas arena, running a background investigation is not an option, and often much, if not most, of what we know about the person we are testing is what he or she has told us. In many cases, the most basic biographic information (i.e., name and place and date of birth) cannot be verified.

Such a lack of biographic data is not only an impediment to formulating good test questions but also inhibits an examiner's ability to get in the proper "mind set" to do the test. A good background investigation can give an examiner a profile of the person to be tested that enables the examiner to prepare an appropriate approach to the test.

The more an examiner knows about the person he or she is testing, the better position the examiner is in to conduct an interrogation, should one become necessary.

In the ops world, the dearth of biographic information can put the examiner at a disadvantage. Establishing any kind of common ground with the person being tested and questioned is the essence of a good interrogation. Without accurate biographic information, establishing a common ground is difficult, if not impossible, and an examiner/interrogator working without that data is not only shooting in the dark but also more prone to error.

Conducting examinations in a foreign language is a common occurrence in the ops world and can be a problem. One of the initial criteria for CIA polygraph examiners was that the examiner be able to speak a foreign language. Too few applicants for polygraph examiner positions could meet that requirement, and interpreters became an integral part of ops testing.

Working with an interpreter is a skill that has to be learned, and even working with the best of interpreters is not as effective as communicating with the person being tested, one-on-one. Often, the interpreter's fluency is not as good as it should be, and meanings are lost in translation. It is essential that the individual being tested have no reservation about the meaning of a question, and, when he or she does, the validity of the test is diminished.

On occasion, I have caught interpreters taking license with the questions being asked, and, if an examiner is not careful, an interpreter can "take over" the examination session.

The "setting" or environment in which ops tests are conducted varies. The American Polygraph Association recommends that examinations be conducted in a sterile, quiet environment that affords the examiner and person being tested an opportunity to focus on the examination without distraction.

Sites in which I have conducted tests ran the gamut from five-star hotels to a peasant hut in which candles were the source of light and odors of urine, unwashed bodies, and nuoc mam (Vietnamese fish sauce) wafted throughout the hut. Some of the distractions I have encountered have been a B-52 strike that

caused the room to shake and plaster to fall from the ceiling and two earthquakes that scared the daylights out of me. Testing in a storage shack at the end of a runway as F-4 Phantoms took off and landed left something to be desired, as did testing in a dilapidated building redolent with the smell of decomposing bodies. On one occasion, a case officer brought me to a test site where there was not a stick of furniture - not a table, chair, or sofa. I declined his suggestion that I conduct the test on the floor.

More memorable than the sights, sounds, and distractions of ops testing is the clientele with whom I dealt. Criminals and clerics, heroes and cowards, intellectuals and dullards, idealists and cynics, teenagers and octogenarians, the good, the bad, the ugly, and the mentally as well as physically impaired are all part of the cast of characters in the operational testing environment.

Every individual who undergoes polygraph testing is different, but there is a degree of sameness among those afforded overt testing that does not exist in the ops world.

As many obstacles and frustrations as there are in the ops world, there are also compensations. First among them is the constant challenge. The mix of clients, varying circumstances under which tests are conducted, and the rare element of danger involved in these tests requires a certain amount of adaptability and initiative on the part of the examiner. Maintaining focus was never a problem in ops testing, and the occasional adrenaline rush extant in ops tests is a plus.

More than the challenge, however, is the job satisfaction that ops examiners experience. If one accepts that clandestine operations are a necessary and vital part of the United States' intelligence effort, so too is the need to maintain the integrity and security of those operations. Polygraph examiners are the gatekeepers of CIA's clandestine operations.

Ronald Reagan's "Evil Empire" may be moribund but as the Hanssen case demonstrates not dead. There has been no noticeable diminishment in the Evil Empire's intelligence activities directed against the United States, and there has been a dramatic increase in such activities by our friends and foes.

When I interviewed for my position with the CIA's Interrogation Research Division, as Polygraph Division was then known, my interviewer, Mr. William ("Bill") Osborne, had an inscribed picture on his wall of a former Director Central Intelligence, Allen Welch Dulles. The inscription read: "Dear Bill, polygraph is our first line of defense."

As true as that statement was then, it is more so today.

Electrodermal Activity: A Literature Review

Chauncey E. Farrell

Abstract

In the past few years, a wealth of research has been published on electrodermal activity. Researchers have demonstrated a correlation between brain activity, alertness, personality disorders and electrodermal activity. Persons with depression, attention deficient hyperactivity disorder, antisocial personality disorder, borderline, or psychopathy will usually have reduced electrodermal levels and responses. Males, according to most studies, will have a greater response than females and most show an increased response in the left hand. The response level and response activity may show differences in each hand. It is suggested that by using the hand with the stronger response, the examiner interpreting the charts may find them easier to read and possibly improve the accuracy in their evaluations. Suggestions are made on possible methods to increase arousal causing a more robust response.

Keywords: asymmetry, electrodermal activity, gender, literature review

Electrodermal activity (EDA), electrodermal response (EDR), and electrodermal level (EDL) are used as general terms. EDL refers to baseline levels, EDR refers to response away from baseline and EDA is the most general term referring to levels and/or responses. Other terms used in the literature included: skin conductance activity (SCA), skin conductance response (SCR), and skin conductance level (SCL) and infers the conductance rather than the resistance scale. Biofeedback literature uses terms such as: skin resistance activity (SRA), skin resistance response (SRR), and skin resistance level (SRL); skin potential activity (SPA), skin potential response (SPR), and skin potential level (SPL). All of these terms refer to measuring the response to a current through the body using electrodes attached to the skin. The recordings taken are measurements of a current passing through the body or the resistance of the body to that current (Peek, 1987).

Electrodermal activity (EDA) is thought to originate from activity in the amygdala and hippocampus in the brain (Ornstein 1997). These areas in the brain are part of the limbic system, the brain's emotional center. This paper will review studies that demonstrate the influence of the brain on the galvanic skin response. The studies show work that has been performed with a variety of patients and situations.

Several studies have shown that damage to the limbic system or to cortical areas of the brain, especially to the right hemisphere and to the frontal lobes, may dampen skin conductance responses to psychologically significant stimuli. Tranel and Damasio (1994) reported reduced electrodermal response in patients with brain lesions found in the ventromedial frontal area (right and left), the right inferior parietal region, and the anterior cingulate gyrus (right and left). These areas are considered to be a part of the limbic system. Zahn, Grafman and Tranel (1999) also showed diminished electrodermal activity in patients with damage to the lateral prefrontal and periventricular areas compared to controls. They found patients with right and bilateral lesions in the cingulate gyrus and/or frontal operculum demonstrated

Request for reprints should be sent to: Chauncey E. Farrell, MS, 1310 McClellan Drive, Klamath Falls, OR 97603, or to the e-mail address cfarrell@cdsnet.net.

diminished electrodermal activity. They concluded that certain frontal brain lesions affect the psychological response of the patients. These studies demonstrated changes in the electrodermal response occurring in patients with damage to the limbic system and other areas of the brain controlling emotions.

Other investigators compared the electrodermal response to activity within the brain. Critchlet et al. (2000) compared magnetic resonance imaging with electrodermal activity. They found increased EDA occurred with activity in the right orbitofrontal cortex, right anterior insula, left lingual gyrus, right fusiform gyrus, and left cerebellum. A lesser relationship, though significant, was found in bilateral medial prefrontal cortex and right inferior parietal lobule. They suggest that areas implicated in emotion and attention are differentially involved in generation and representation of peripheral electrodermal response.

Several studies have looked at the asymmetry of the electrodermal response found in patients with and without unilateral brain damage. A 1987 study (Martinez-Selva et al.) used 15 males and 15 females. They reported males displayed more asymmetry between hands with larger responses in the left hand. Females showed a less marked asymmetry but with larger skin responses in the right hand. Schulter and Papousek (1998) studied bilateral electrodermal activity. They observed both a stimulus free recording and stimulus evoked skin conductance responses (SCR). The stimulus consisted of tones in the dominant or nondominant ear. The stimulus-free recordings did not demonstrate asymmetry of SCRs. However, during auditory stimulation, they found higher amplitudes of SCRs contralateral to the preferred hemisphere in strong dextals while weak right-handers demonstrated a different pattern of bilateral asymmetries, partly in the opposite direction. In 1998, Naveteur et al. reported on a patient with right frontal damage who exhibited an unexpected stronger left than right asymmetry of SCLs. They suggested that the frontal lesion had induced a contralateral disinhibition control of tonic electrodermal level. They discounted peripheral factors such as sweating or temperature as being related to the asymmetry. In another study using bilateral electrodermal activity recording, Brand et al. (1999) used a specific odor (lavender) in a single nostril as the stimuli. They reported no difference between the two nostrils but all 30 subjects exhibited a consistent direction of electrodermal asymmetry. Two-thirds of the subject systematically demonstrated greater response amplitude for the right hand while one-third demonstrated a greater response for the left hand. In another study using olfactory stimuli, Conesa (1995) used a one-subject design experiment. Over a period of 26 days, 7 a.m. to midnight, a total of 760 measurements were taken. For the single subject, electrodermal asymmetry was confirmed during the right nostril condition showing higher right than left hand conductance. The asymmetry was less pronounced during the left nostril stimulation.

A paper from Russia, Dementienko et al. (1999), found that a decrease in wakefulness accompanied a drop of the electrodermal activity. They suggested that discrepancies in experimental studies of psychophysiological events using EDA might be explained by an insufficient control of the subject's alertness throughout the experiment. Other studies have also found alertness to be a factor in the galvanic skin response. Smoking, a recognized stimulant, has been shown to enhance electrodermal activity. Two studies, Furedy et al. (1999), and the 1993 study of Morris and Gale, both demonstrated a significant increased electrodermal activity in males after smoking compared to EDA prior to smoking.

Several investigators have looked at levels of arousal using electrodermal responses and brain waves simultaneously. Smith et al. (1995) studied introverts and extroverts using both electroencephalograph (EEG) and electrodermal activity. EEG frequencies are described as delta, deep sleep; theta, drowsiness; alpha, relaxed wakefulness; beta, ranging from attentiveness to strong excited emotion. They reported a high correlation of arousal seen on both the EEG and EDA. They found a higher level of arousal demonstrated by the introverts compared to the extroverts. In 1996, Lim et al. investigated cerebral cortical activity and sympathetic autonomic activity using both EEG and skin conductance

level. Using 10 normal adult subjects, they found a significant correlation of increased skin conductance with alpha and beta frequencies. These studies indicate a strong correlation between psychological arousal and electrodermal activity.

Several investigators have looked at galvanic skin response, emotions, and different psychological conditions. Collet et al. (1997) reported different emotions demonstrated different responses in electrodermal activity. In 1999, Andersson, Krogstad, and Finset reported on electrodermal activity correlated to apathy found in traumatic brain injured patients. They reported no significant correlation between apathy and electrodermal activity. In a study comparing 18 neurological patients with severely disturbed vigilance to 18 healthy subjects, Schuri and von Cramon (1982) reported significant differences. They found the patients had reduced electrodermal responses on initial trials but had increased responses on later trials to acoustic stimuli. They interpreted the response pattern as reflecting changes in activation. Herpertz et al. (1999) investigated electrodermal responses in 24 female subjects with borderline personality disorders compared to normal females. The stimuli consisted of photographic slides with pleasant, neutral or unpleasant emotional valence. They reported the borderline patients showed lower electrodermal arousal as compared to normal subjects. They concluded that borderline personality disorder patients suffer from autonomic under-arousal rather than affective hyperresponsiveness suggested by current theories.

Twenty-nine female schizophrenic patients were compared to female controls in a study reported in 1994 by Wieselgren et al. In a two-year longitudinal study, they reported higher skin conductance levels and increased spontaneous skin conductance fluctuations in the group with poor social functioning outcomes. This was the reverse of an earlier finding of schizophrenic men that showed poor social functioning associated with reduced electrodermal activity. Raine, Venables and Williams (1995) reported a 14-year prospective study testing the hypothesis that antisocial adolescences who desist from crime by the age of 29 have greater psychological arousal than antisocial adolescence who become adult criminals. They found that 15-year-old male schoolchildren who showed significantly reduced electrodermal and cardiovascular arousal developed criminal behavior by the age of 29. That study was followed a year later by another study with Raine as a co-author. Raine, with seven other co-authors (Brennan, 1997), wrote in the American Journal of Psychiatry, about their study with matched controls of subjects with fathers who were criminals. They again reported that subjects at high risk for criminal behavior who showed increased autonomic nervous system responsiveness compared to controls appeared to have a lower likelihood of criminal activity outcome. Raine continued his research. In 1998, Raine et al. reported a study comparing 15 predatory murderers and nine affective murderers. Affective murderers were described as those who lost their temper and then committed the crime. They reported the affective violence offenders, compared to predatory murderers, showed lower left and right prefrontal functioning, higher right hemisphere subcortical functioning, and lower right hemisphere prefrontal/subcortical ratios. They hypothesized that excessive subcortical activity predisposed to aggressive behavior and lack of prefrontal control over emotional regulation. In 1996, Raine reported autonomic under-arousal found in infants and young children with a disinhibited temperament that is thought to be a predisposition to juvenile delinquency and adult aggressive behavior. He hypothesized that aggressive children may be stimulation seekers who are relatively fearless. Scarpa and Raine (1997) report antisocial populations to be in a state of psychophysiological under-arousal with reduced skin conductance levels and increased slow wave EEG.

Raine then looked at skin conductance and schizotypal criminals. Raine, Bihrle, Venables, Mednick, and Pollock (1999) reported a prospective and longitudinal study of alcoholism in 134 male schizotypal subjects assessed during adolescence between the ages of 18 and 20 and later at the ages of 30 to 33. They reported decreased prefrontal functioning combined with reduced skin conductance responses and increased alcoholism compared to the criminal controls. In the year 2000,

Raine directed his attention to the antisocial personality disorder. Raine et al. (2000) reported the antisocial personality disorder group compared to a healthy subject group, a substance dependency group, and a psychiatric patient group demonstrated and 11 percent reduction in prefrontal gray matter volume. They concluded that these findings provide the first evidence for a structural brain deficit in the antisocial personality disorder. This prefrontal structural deficit may underlie the low arousal, poor fear conditioning, lack of conscience and decision-making deficits found to characterize antisocial, psychopathic behavior.

Blair, Jones, Clark, and Smith in 1997 investigated psychopathic individuals. They recorded electrodermal responses of the subjects watching slides indicating distress cues, threats and neutral stimuli. When compared to 18 incarcerated control individuals, the psychopathic individuals showed reduced electrodermal response to the distress cues but did not differ in their response to the threatening stimuli and to the neutral stimuli. In Antonio Damasio's book (1994) is a description of a study demonstrating that the patients with frontal lobe damage had very little anticipatory electrodermal response concerning behaviors when gambling. His notes referred to a study by Dr. Hare (1971) showing similarities in patients diagnosed as psychopathic and with criminal records behave similar to patients with frontal lobe damage.

Lazzaro et al. (1999) found adolescents with attention deficit hyperactivity disorder also had reduced brainwave beta activity, increased theta and alpha activity and reduced skin conductance levels when compared to control subjects. Depressed patients demonstrated reduced electric skin conductance levels compared to normals as reported by Ward and Doerr (1986). Storrie, Doerr and Johnson (1981) also reported low responses to skin conductance found in depressed individuals that did not change after successful therapeutic intervention.

Discussion

Recording electrodermal activity is dependent on a functioning autonomic nervous system and sufficient arousal in the normal brain produced by a stimulus. The autonomic nervous system, arising from the limbic system, reacts to perceived threats and produces emotional arousal responses. The brain must be able to receive stimuli, perceive a threat, and react accordingly.

Some interesting work has been done comparing electrodermal response to activity in the brain. The fields of psychology and psychiatry have a wealth of information with their research using EDA. Their studies show reduced electrodermal activity in patients with brain damage or brain dysfunction of specific areas. These studies demonstrate clearly that the limbic system is closely related to psychological arousal as monitored by the electrodermal response. Other studies have shown that asymmetrical activity of the brain is related to asymmetry with skin conductance. More recent studies have demonstrated reduced brain activity in areas including the limbic system in persons that are ADHD, antisocial, borderline or psychopathic. Because these people are under-aroused, they stimulate themselves by engaging in activities that are dangerous, on the edge or may cause them or others physical harm. They have a difficult time with sustained focus on any activity. They are easily bored and are constantly seeking behaviors that might get them into trouble. Many of these personalities are likely to become criminals and be tested by polygraph. Hare (1993) estimated that on average, about 20 percent of male and female prison inmates are psychopathic. He reported that the psychopaths are responsible for more than 50 percent of the serious crimes committed.

The polygraph examiner will certainly see his share of individuals that are psychopathic, antisocial, depressed and borderline. These individuals will most likely demonstrate reduced skin conductance response. This reduced response is central in origin, i.e., from the brain. No amount of skin conditioning will change a reduced response that is central in origin. These individuals must be psychologically or physiologically aroused at the time of testing. The question is how to arouse the psychopathic or antisocial individual sufficiently to show more to normal amplitudes of response to the questions posed during the examination.

Research suggests several remedies, but incorporating them into a polygraph exam may be problematic. Since smoking has been found to increase the amplitude of the EEG and skin conductance, the examiner might allow the examinee to take a smoke break between the pretest interview and the actual testing. Or, the examinee could be permitted to stand up, stretch or move about between each chart recording. Scents may also be stimulating. Sullivan et al.(1998) found a scent of peppermint increased performance on tests of vigilance. Van Toller et al.(1983) recorded increased skin measurements in subjects who perceived an odor as unpleasant. Two studies (Peretti & Zweifel, 1983; Myskja & Lindbaek, 2000) found significant differences in skin responses to individuals reacting to musical preference. An act as simple as chewing a commercially flavored gum demonstrated aroused psychosomatic responses (Masumoto et al.,1998). Future studies may look at some of these areas and see if methods exist to increase arousal in subjects that are demonstrating reduced EDA in a testing situation.

References

Andersson, S., Krogstad, J.M., & Finset, A. (1999). Apathy and depressed mood in acquired brain damage: Relationship to lesion localization and psychophysiological reactivity. *Psychol Med*, 29(2), 447-456.

Blair, R.J., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34(2), 192-198.

Brand, G., Millot, J.L., & Henquell, D. (1999). Olfaction and hemispheric asymmetry: Unilateral stimulation and bilateral electrodermal recordings. *Neuropsychobiology*, 39(3), 160-164.

Brennan, P.A., Raine, A., Schulsinger, F., Kirkegaard-Sorensen, L., Knop, J., Hutching, B., Rosenberg, R., & Mednick, S.A. (1997). Psychophysiological protective factors for male subjects at high risk for criminal behavior. *Am J Psychiatry*, 154(6), 853-855.

Collet, C., Vernet Maury, E., Delhomme, G., & Dittmar, A. (1997). Autonomic nervous system response patterns specificity to basic emotions. *J Auton Nerv Syst*, 62(1-2), 45-47.

Conesa, J. (1995). Electrodermal palmer asymmetry and nostril dominance. *Precept Mot Skills*, 80(1), 211-216.

Critchley, H.D., Elliott, R., Mathias, C.J., & Dolan, R.J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *J Neurosci*, 20 (8), 3033-3040.

Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York, NY: Avon Books.

Dementienko, V.V., Dorokhov, V.B., Koreneva, L.G., Markov, A.G., & Shakhnarovich,V.M. (1999). The characteristics of the electrodermal activity during changes in the level of human wakefulness. *Zh Vyssh Nerv Deiat Im I P Pavlova*, 49(6), 926-935.

Furedy, J.J., Algan, O., Vincent, A., Demirgoren, S.,& Pogun, S. (1999). Sexually dimorphic effect of an acute smoking manipulation on skin resistance but not on heart rate during a cognitive verbal task. *Integr Physiol Behav Sci*, 34(4), 219-226.

Hare, R.D., & Quinn, M.J.(1971). Psychopathy and autonomic conditioning. *J Abnorm Psy*, 77, 223-235.

Hare, R.D. (1993). *Without conscience: The disturbing world of the psychopaths among us.* New York, NY: The Guilford Press, p.87.

Herpertz, S. C., Kunert, H.J., Schwenger, U.B., & Sass, H. (1999). Affective responsiveness in borderline personality disorder: A psychophysiological approach. *Am J Psychiatry*, 156(10), 1550-1556.

Lazzaro, I., Gordon, E., Li, W., Lim, C.L., Plahn, M., Whitmont, S., Clarke, S., Barry, R.J., Dosen, A., & Meares, R. (1999). Simultaneous EEG and EDA measures in adolescent attention deficit hyperactivity disorder. *Int J Psychophysiol*, 34(2), 123-134.

Lim, C.L., Barry, R.J., Gordon, E., Sawant, A., Rennie, C., & Yiannikas, C. (1996). The relationship between quantified EEG and skin conductance level. *Int J Psychophysiol*, 21(2-3), 151-162.

Martinez-Selva, J.M., Roman, F.,Garcia-Sanchez, F.A., & Gomez-Amor, J. (1987). Sex differences and the asymmetry of specific and nonspecific electrodermal responses. *Int J Psychophysiol*, 5(3), 155-160.

Masumoto,Y., Morinushi, T., Kawasaki, H., & Takigawa, M. (1998). Spectral analysis of changes in electroencephalographic activity after the chewing of gum. *Psychiatry Clin Neurosci*, 52(6), 587-592.

Morris, P.H., & Gale, A. (1993). Effects of situational demands on the direction of electrodermal activation during smoking. *Addict Behav*, 18(1), 35-40.

Myskja, A., & Lindbaek. M. (2000). How does music affect the human body? *Tidsskr Nor Laegeforen*, 120(10), 1182-1185.

Naveteur, J., Godefroy, O., & Sequeira, H. (1998). Electrodermal level asymmetry in a unilateral brain-damaged patient. *Int J Psychophysiol*, 29(3), 237-245.

Ornstein, R.,(1997). *The right mind: Making sense of the hemispheres.* New York, NY: Harcourt, Brace and Co.

Peak, C. J.(1987). A Primer of Biofeedback Instrumentation. In M. S. Schwartz, (Eds.), *Biofeedback, A Practitioners Guide*, 2nd ed. (pp. 45B 96). New York, NY: The Guilford Press.

Peretti, P.O., & Zweifel, J. (1983). Effect of musical preference on anxiety as determined by physiological skin responses. *Acta Psychiatr Belg*, 83(5), 437-442.

Raine, A., Venables, P. H., & Williams,M. (1995). High autonomic arousal and electrodermal orienting at age 15 years as protective factors against criminal behavior at age 29 years. *Am J Psychiatry*, 152(11), 1595-1600.

Raine, A. (1996). Autonomic nervous system factors underlying disinhibited, antisocial, and violent behavior. Biosocial perspectives and treatment implications. *Ann N Y Acad Sci*, Sep 20(794), 46-59.

Raine, A., Meloy, J.R., Stoddard, J., LaCasse, L., & Buchsbaum, M.S. (1998). Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affective murderers. *Behav Sci Law*, 16((3), 319-332.

Raine, A., Bihrle, S., Venables, P.H., Mednick, S.A., & Pollock, V. (1999). Skin conductance orienting deficits and increased alcoholism in schizotypal criminals. *J Abnorm Psychol*, 108(2), 299-306.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch Gen Psychiatry*, 57(2),119-127.

Scarpa, A., & Raine, A. (1997). Psychophysiology of anger and violent behavior. *Psychiatr Clin North Am*, 20(2),375-394.

Schulter, G., & Papousek, I. (1998). Bilateral electrodermal activity: relationships to state and trait characteristics of hemisphere asymmetry. *Int J Psychophysiol*, 31(1), 1-12.

Schuri, U., & von Cramon, D. (1982). Electrodermal response patterns of neurological patients with disturbed vigilance. *Behav Brain Res*, 4(1), 95-102.

Smith, B.D., Kline, R., Lindgren, K., Ferro, M., Smith, D.A., & Nespor, A. (1995). The lateralized processing of affect in emotional labile extraverts and introverts: Central and autonomic effects. *Biol Psychol*, 39(2-3), 143-157.

Storrie, M.C., Doerr, H.O., & Johnson, M.H. (1981). Skin conductance characteristics of depressed patients before and after therapeutic intervention. *J Nerv Ment Dis*, 169(3), 176-179.

Sullivan, T.E., Warm, J.S., Schefft, B.K., Dember, W.N., O'Dell, M.W., & Peterson, S.J. (1998). Effects of olfactory stimulation on the vigilance performance of individuals with brain injury. J Clin Exp Neuropsychol, 20(2), 227-236.

Tranel, D., & Damasio, H. (1994). Neuroanatomical correlates of electrodermal skin conductance responses. *Psychophysiology*, 31(5), 427-438.

Van Toller,C., Kirk-Smith, M., Wood,N., Lombard, J., & Dodd, G.H. (1883). Skin conductance and subject of assessments associated with the odor of 5-alpha-androstan-3-one. *Biol Psychol*, 16(1-2), 85-107.

Ward, N.G., & Doerr, H.O. (1986). Skin conductance. A potential sensitive and specific marker for depression. *J Nerv Ment Dis*, 174(9), 553-559.

Wieselgren, I.M., Ohlund, L.S., Lindström, L.H., & Ohman, A. (1994). Electrodermal activity as a predictor of social functioning in female schizophrenics. *J Abnorm Psychol*, (3), 570-575.

Zahn, T.P., Grafman, J., & Tranel, D. (1999). Frontal lobe lesions and electrodermal activity: effects of significance. *Neuropsychologica*, 37(2), 1227-1241.

A Concise History of the Comparison Question

James F. Waller

Key words: comparison question, comparative response question, control question, directed lie, emotion standards, history, probable lie

Over the years many claims have been made as to who first developed and utilized a comparison question technique in a detection of deception exam. This article will trace the earliest development and use of the comparison question to its current status within the polygraph discipline.

The earliest published report of a comparison question technique while recording the responses of a subject was done by Cesar Lombroso. In the second edition of his book *L'Homme Criminal* (1895), Lombroso used a plethysmograph and sphygmograph to record the responses of criminal suspects as they were questioned concerning their knowledge and involvement in certain crimes. In a celebrated case Lombroso used the hydrosphygmograph to determine that a suspect in a train robbery did not steal 20,000 francs, but was involved in the theft of certain documents and passports from the train. Lombroso came to this conclusion when the suspect showed no response to the questions about the train robbery, but a significant blood pressure drop was recorded when the question concerning the theft of the documents was asked. Lombroso reported that, based on these results, the suspect was being truthful about not being involved in the theft of the money and that he was lying about his involvement in the theft of the documents. This was the first documented case of the use of a comparison-type question technique in a psychophysiological detection of deception examination. It differed from current comparison question methods in that the two issues used for comparison purposes by Lombroso were actually relevant issues, since the examinee was suspected of both crimes. Lombroso's conclusion concerning the suspect's guilt was reportedly substantiated later.

Another major development in the realm of the comparison question technique is attributed to Dr. William M. Marston in the early 1920s. Dr. Gordon H. Barland (2001), reported that between 1984-85 he was present when Norman Ansley interviewed Dr. Marston's former assistant, Olive Richard. Dr. Barland stated that she described the use of a "hot question" on some of Dr. Marston's exams. The description of this question corresponded to the non-exclusive probable-lie comparison question published by Reid (1947). It would seem that Dr. Marston was already using this a form of probable-lie question some 17 years before Reid was credited with his probable-lie comparison question. Whether it was used for direct comparative purposes was unstated. Olive Richard related that Dr. Marston did not publish anything dealing with this question, to keep its use out of the hands of people having to take the tests. It was clear that even at this early date there was a concern about countermeasures. Because of the absence of published reports, Marston is rarely credited for his early use of comparison questions. Marston also utilized another type of comparison question in some of his deception tests. They were unrelated to the relevant question, yet calculated to stimulate various emotions, and they were alternated with relevant questions (Larson, 1922).

CW3 James F. Waller is a Special Agent with U.S. Army Intelligence, U.S. Army Intelligence and Security Command, and currently an instructor at the Department of Defense Polygraph Institute. The views expressed in this paper do not necessarily reflect those of the Department of Defense or the US Government. Requests for reprints should be directed to wallerj@jackson-dpi.army.mil.

Meanwhile, in the late 1920s to early 1930s Leonard Keeler was trying several types of comparison questions with his Relevant/Irrelevant technique. One was the Personally Embarrassing Question (PEQ), which was designed to evoke a response in the innocent person. The reasoning was that the guilty person's attention would be on the relevant issue, and the innocent person would be more concerned with the embarrassing issue. The fatal shortcoming of this technique was that both guilty and innocent subjects tended to respond equally to the PEQ. This led Keeler to abandon the PEQ. The other comparison questions used by Keeler were called "cover questions." These questions were of the inclusive type, not unlike those used in the Reid Test. Cover questions, however, were used for more than their comparative value. The examiner was interested in whether the subject had committed a similar crime, and could use cover questions to reveal them. Consequently, this method had something in common with Lombroso's method, which compared the reactions from relevant questions about different crimes the same examinee may have committed. Keeler also used non-lie questions that linked the examinee to the event, and these questions could be used as reference points against which the relevant questions were compared (Abrams, 1989).

A major contributor to the comparison question technique was Reverend Walter G. Summers. In a paper entitled, "Science can get the confession" (1939), Summers outlined how he had been experimenting with a psychogalvanograph to get a more precise index of emotion. These studies influenced Summers in the creation of a questioning method for his own deception test. He developed three types of questions. The first were "significant questions," which correspond with what are called relevant questions today. Within one record he presented three different, but related significant questions, and there were three presentations of each significant question. Significant questions were asked among a larger number of non-significant questions such as, "Are you wearing a black coat?" or "Did you eat breakfast this morning?" Summers also used what he called "emotional standards," such as, "Were you ever arrested?", "Are you living with your wife?" and "Do you own a revolver?", questions designed to be emotional in nature. He wrote:

> The emotional standards are selected after careful analysis of the suspect's life history and after the examination of his psychogalvanic reactions to a preliminary series of questions. When chosen properly, the emotional standards tend to evoke within the individual rather intense psychogalvanic reactions due to surprise, anger, shame, or anxiety over situations which he would ordinarily prefer to conceal. In the examination of suspects an emotional standard precedes each significant question. For purposes of interpretation we contrast and compare the reactions to the significant questions with the reactions to the emotional standards. If the deflections to the critical (significant) questions are consistently greater than the deflections to the emotional standards, the individual is consciously trying to deceive the examiner. If, on the other hand, the deflections to the critical questions are not consistently greater than those to the emotional standards, the individual is truthfully expressing his state of mind. This is the essential criterion of interpretation (Summers, 1939).

From his description, it can be seen that Summers was directly comparing reactions to relevant question to those of non-relevant comparison questions, and proposing the beginnings of a decision rule for results of truthfulness or deception. Summers published this technique almost eight years before Reid's paper appeared. And, in what appears to be a foreshadowing of some future techniques, Summers preceded each relevant question with a comparison question.

Though John Reid may not have invented the comparison question technique, he was highly influential in how the technique is practiced in a modern polygraphy. Reid advocated what would later be known as the probable-lie comparison question, which differed from Summers approach. Summer's emotional standard questions were answered truthfully by the examinee, while Reid's

examinees were expected to lie to the comparison questions. Reid found, as did Summers, that innocent individuals respond physiologically greater to the comparison questions, while the guilty reacted greater to the relevant questions. Most believe that Reid's approach, using probable lies, enhanced the accuracy of the polygraph technique (Abrams, 1989).

After Reid 's contribution to the development of the comparison question, Cleve Backster was next to influence the technique. He was responsible for the exclusive comparison question. This version of comparison question is formulated to be in the same category of offense as the crime in question, but is worded to exclude the relevant crime. Backster explained the function of the comparison question within the framework of what he calls "psychological set." He reported that attention is involuntarily directed to whatever the person perceives as the greatest threat. Therefore, if the subject is lying to the comparison question, but truthful to the relevant question, the theory predicts that he or she will show greater physiological arousal to the comparison question because of concern the lie will be detected. Conversely, if the examinee is lying to both the relevant and comparison questions, his or her concern will be directed primarily toward the relevant item because the consequences for detection of that lie are greater than those for the comparison question topic. Backster's comparison questions are arguably the most commonly used in the field today.

The last noteworthy development in the comparison question technique was the advent of the directed-lie comparison (DLC) question. In the DLC family of techniques, the examiner and examinee agree before testing begins that the examinee will lie to the comparison questions, unlike the PLC procedure, which entails some subtle maneuvering by the examiner to have the examinee lie to the PLCs. The theory of psychological set also plays an important part in the DLC technique. For the innocent subject there is no threat from the relevant questions, and they tend to concentrate on the DLC questions. Guilty examinees know the importance of both questions, however, the greatest threat comes from detection of the lie to the relevant issue because of the greater adverse consequences.

Louis Fusse wrote in 1982 that the DLC had been developed some 16 years earlier. It came out of the intelligence arena where multiple-issue tests were being employed in examinations of intelligence sources. Fusse claimed that, as sources were subject to repeated testing over time, the effectiveness of PLC questions tend to diminish. It was found that DLCs remained effective, however, and could be used to test the same individual over many years.

Not all field examiners have embraced this latest polygraph technique. It has been pointed out that the theoretical underpinnings of the DLC are different from those of the PLC, which are better understood. Some believe the DLC may actually pose a greater threat to the guilty if he or she believes the physiological reactions to the DLC are used as a gauge for assessing reactions to relevant questions (Matte, 1996). Other research has found the DLC superior to the PLC (Horowitz, Kircher, Honts & Raskin, 1997). The final answer to the efficacy of the DLC is waiting for more research.

The goal of this paper was to trace the history of the comparison question, and identify those who advanced its development to its current stage. We credit Cesare Lombroso with the first comparison question technique, using the reactions to one of two relevant questions to infer deception to one and truthfulness to the other. The earliest pioneers of what we know today as the emotion-evoking comparison question were Dr. Marston and Reverend Summers, who recognized that reactions to non-relevant questions could be useful to diagnose deception. John Reid was responsible for refinement and popularization of this principle, devising the probable-lie comparison question, while Cleve Backster and Louis Fuse promoted new forms that found wide application. All have contributed toward making polygraphy what it is today.

References

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

Fuse, L.S. (1982). Directed lie control testing (DLCT) technique. Unpublished manuscript.

Horowitz, S.W, Kircher, J.C., Honts, C.R., & Raskin, D.C. (1997). The role of comparison questions in physiological detection of deception, *Psychophysiology,* 34(1), 108-115.

Larson, J.A. (1922). Modification of the Marston Deception Test. *Journal of the American Institute of Criminal Law and Criminology,* 12, 390-399.

Matte, J.A. (1996). *Forensic Psychophysiology Using the Polygraph, Scientific Truth Verification-Lie Detection.* J.A.M. Publication: Williamsville, New York.

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

Summers, W. G. (1939). Science can get the confession. *Fordham Law Review,* 8, 334-354.

Trovillo, P.V. (1939). A history of lie detection. *Journal of Criminal and Criminology,* 29 (6), 848-881 and 30 (1), 104-119.

Psychological Set: Its Origin, Theory And Application

James Allan Matte and Robert Nelson Grove

Abstract

In 1960, Cleve Backster used the term *psychological set* to explain anti-climax dampening, the basis for his Zone Comparison Technique. Since then, the term has been attributed to Backster in numerous articles pertaining to polygraph. Backster however has always attributed its source to a *Psychology and Life* textbook by Floyd L. Ruch (1948). Recently (Honts 2000; Krapohl 2001) the term *psychological set* has been questioned as a non-scientific or non-modern term, an embarrassment to the scientists trying to explain the 'real' basis of anti-climax dampening. This paper sets forth the historical origins of its application to forensic psychophysiology and shows how *psychological set* continues to be widely used and regarded in contemporary psychological, industrial/organizational and legal/political science communities.

Keywords: psychological set, selective attention, theory

Background

In 1960, Cleve Backster, in a school handout, described anti-climax dampening as follows:

> Anti-climax dampening effect involves the inter-relationship of two issues, questions, or topics, in close proximity to each other, where the more important, bothersome or stimulating issue suppresses or completely eliminates emotional response to the other issue, question or topic which the person might have responded to had the other stronger issue, question or topic not been present.

Anti-climax dampening was described as an effect based on natural psychological phenomena associated with attention and set. Backster defined attention as 'readiness to respond to stimuli.' Set was 'an adjustment of an organism in preparation for a certain kind of activity.' Next he gave the definition of *Psychological set* found in a widely-used textbook *Psychology and Life* by Floyd L. Ruch (1948).

Ruch clearly and elegantly defined psychological set in the following way:

> From all the energies about us, our sense organs select only certain ones. The others are tuned out just as effectively as we tune out the voice of one speaker on the radio so that we may hear that of another. But the selectivity of the human organism goes far beyond physiological selectivity—a heightening or lessening of responses to certain stimuli within range—or its lack of, in the sense organs. This selectivity is more than sensory. Although several stimuli compete, only those fitting the need of the moment are selected. For example, when you are deep in an interesting conversation the sounds of traffic noises outside are not heard.

Dr. Matte is a member of APA, the owner of Matte Polygraph Service, Inc., and author of three textbooks: *The Art and Science of the Polygraph Technique* (1980), *Forensic Psychophysiology Using The Polygraph: Scientific Truth Verification - Lie detection* (1996), and *Examination and Cross-Examination of Experts in Forensic Psychophysiology Using The Polygraph* (2000). Dr. Grove is a Medical Psychologist and Adjunct Faculty, Aliant University, formerly known as California School of Professional Psychology, San Diego, CA.

The authors wish to express their sincere appreciation to Dr. (Hon.) Cleve Backster, Director of the Backster School of Lie Detection, and Dr. Gordon H. Barland, DoDPI Research (retired), both APA members, for their contribution of information to this project.

Thus the term *psychological set* was introduced by Backster to the polygraph field 42 years ago. Since then the discipline of forensic psychophysiology (ASTM 1999) has emerged, infused with knowledge brought by scientists from behavioral and medical disciplines never dreamed of in those early years. Yet in spite of Backster's declaration that he acquired the term from a book on psychology (Ruch 1948), the polygraph community has generally attributed Backster as the source of that term, while others (Honts 2000; Krapohl 2001) have questioned the scientific basis, definition and applicability of that term, regardless of who inserted it into our discipline.

Honts (2000) was very clear on his reservations: "The notion of *psychological set* is a contrivance of the polygraph profession and has received little scientific validation. Moreover, *psychological set* is not a term that is currently much used in mainstream psychological science. While the hypothetical construct, *psychological set*, may have some heuristic value as a descriptive tool, it has no reality in science or the real world."

Krapohl (2001) acknowledges the use of the term in the past, but yearns for a more modern, less antiquated term than *psychological set*: "As for the expression *psychological set*, in the mainstream literature it relates to expectancies, not attention. The research goes back to the 30s, and there is a lot of it. Our profession cooped the term for our use. We at DoDPI are looking for a substitute term, one that corresponds with the rest of the psychological sciences, but it will take a while. In the meantime, it is only a borrowed term." Note that Krapohl (2001) assumes *psychological set* is biased toward expectancies, and away from attention, as if expectancies were not as relevant to polygraph examinations as was the concept of selective attention. Yet during the early 1960s, Backster himself authored several articles pertaining to the Backster Zone Comparison Technique wherein he used the term *psychological set* interchangeably with "attention set" (Backster 1962).

Others in the polygraph community have embraced *psychological set* as fundamental to understanding polygraph protocols. For example, as early as 1965, the United States Army Military Police School (USAMPS) Department of Resident Instruction's Summary Sheet[1] articulates "The Psychological Theory of The Polygraph Examination" and the basis of "The Backster Zone Comparison Technique." It further defines the *psychological set* as follows:

"2. Psychological Set.

a. A person's fears, anxieties, and apprehensions are channeled toward the situation which holds the greatest immediate threat to his self-preservation or general well being. He tunes in that which indicates trouble or danger by having his sense organs tuned for a particular stimulus, and he tunes out that which is of a lesser threat to his self-preservation or general well being. In other words, he establishes a *psychological set.*

b. *Psychological set* is selective in nature and depends upon the present frame of reference."

"4. The Anticlimax Dampening Concept.

a. The anticlimax dampening concept is based on the theory of *psychological set.*

b. In a series of questions containing a relevant and a control question, the guilty (lying) subject will tune in the relevant question and out the control question, and the innocent (truthful) subject will tune in the control question and out the relevant question.

c. If a series of relevant questions are asked during a test, the guilty (lying) subject will direct his attention to the most intense relevant question. He will basically perceive but may not be materially affected by the weaker relevant questions, i.e., he may tune them out.

[1] USAMPS Polygraph School's 1963 Summary Sheet Introduced the Backster Zone Comparison Technique and defined the term *Psychological set* and Anti-Climax Dampening.

d. Backster calls this tuning out of the weaker relevant questions the anticlimax dampening concept."

Apparently *psychological set* has become an important concept within much of the polygraph community because no one has provided an alternative concept that so closely links the importance of the preparation for the exam to specific procedures which if violated negate achieving a valid and reliable result.

Psychological Set in Contemporary Psychology

"Set" is widely used throughout modern psychology and even psychophysiology. Psychologists have prefixed the term "set" with such qualifiers as 'perceptual,' 'preparatory,' 'attention' as well as 'psychological.' These qualifiers determine their meaning and definition. For example the term "set" is defined in the *Penguin Dictionary of Psychology* (Reber 1995) in a number of ways, depending on the qualifiers:

"1. n. A classification, aggregate or series of things sharing some defining property or properties such that they can be regarded collectively. This general meaning encompasses a variety of uses from the purely mathematical characterization embodied in set theory, through the more common-sense denotations such as the set of respondents to a questionnaire, the set of stimulus items in an experiment, the country-club set in upper-class society, etc. Note that sets may be infinite in size (the set of integers), finite (the set of correct answers on a multiple-choice test), empty (the set of immortal persons) or poorly defined (the set of all young persons. See here *fuzzy set*)."

"2. n. Any condition, disposition or tendency on the part of an organism to respond in a particular manner. Note that the term 'respond' here may encompass a number of acts. Thus, one may have an attentional or perceptual set for particular kinds of stimuli (see here Einstellung), a task-oriented set for a problem (see here Aufgabe), a functional set which directs the manner of use of objects (see here functional fixedness), a muscular set in which a particular motor act is optimized (preparatory set), etc. To distinguish among these various uses many authors will use qualifiers, as in some of the following entries. It should also be recognized that the term is generally used with the connotation that the set under consideration is a temporary (although potentially recurring) one and, as such, its meaning is contrasted with terms like habit and trait, which refer to enduring dispositions or conditions, and distinguished from schema (l), which is used for more general orientations to situations. The longer term determining set is often used synonymously for 2, particularly for sets that exert some measure of control over how the organism is to respond. adj. set; vb. (for 2), set." (underlined text by the authors)

Note the last sentence in this definition of "set" by W. F. Hill (1970) as "Transfer may be from either recent or more distant experience. The effects of very recent experience are often spoken of as a set. When a person fails to solve a problem because he has recently used the essential tool in its familiar way, or because he has just solved a series of problems with a formula that will not work in the present situation, he is exemplifying set. Set may also be manipulated by telling him something that biases him either toward or away from the correct approach." (underlined text by the authors)

C. G. Morris (1973) explains "Apart from these emotional factors, two other things may affect your ability to solve a problem. These are set and functional fixedness. You usually approach a problem with some sort of direction or expectation which is the result of experience. This is a set - a kind of habit, the way you are used to perceiving certain situations. The value of previous experience in problem solving is that you have learned certain methods or ways of perception in the past, and you can apply them to the present situation. In the example of the conflicting appointments, one set you might have is that it is not polite to break appointments. Without that set, our solution might have been just to go off to the tennis court and forget the dentist entirely. A set can function as a 'hint' toward the solution to a problem." (underlined text by the authors)

Hence it can be seen that it is the prefix qualifier to "set" that determines the

ultimate definition of the whole term, i.e., functional set, muscular set, perceptional set, attention set, mental set, or psychological set.

It should be noted that in England the use of the term *mental set* in lieu of *psychological set* was the preferred term since the 1940s. It is alive and well today. For example, the recent use of *mental set* is described in "Processing in the Stroop task: Mental set as a determinant of performance." (Bauer & Besner 1997). It reads in part as follows:

"Subjects took part in a Stroop experiment in which they responded to the print color of an irrelevant word that spelled a congruent or incongruent colour word. In the CLASSIFY condition, subjects were instructed to map one colour to one response button and the other colour to another response button. In the DETECT condition, subjects were instructed to signal the presence of a target colour with one button, and its absence with a different response button. The CLASSIFY instructions produced the standard result: The incongruent condition was slower than the congruent condition. In contrast, there was no Stroop effect given DETECT instructions. These results are discussed in terms of *mental set* as an important determinant of processing, and contrasted with the received view that reading the irrelevant word is largely 'automatic' and virtually always results in a 'Stroop Effect.' Hence *mental set* is not expectation or attitude, but a disposition to respond to an immediate situation by filtering out irrelevant stimuli. This is similar to the definition of *selective attention*, "(t)he process involved in situations in which one is confronted with multiple stimulus inputs and must select but one aspect of them and attend to it." (Reber 1995).

Interestingly, the term "selective attention" as defined in the contemporary text, *Abnormal Psychology* (Bootzin, et al 1993) appears to be synonymous with "psychological set" as defined in *Psychology and Life* (Ruch 1948). An excerpt from Abnormal Psychology is quoted below:

"ATTENTION: When the mind takes in only some of the information it is exposed to, it is engaging in *selective attention*. *Selective attention* is an indispensable adaptive function. We cannot possibly attend to, let alone process, all the information that impinges on our faculties at any given moment. So we focus on what seems to us most important and filter out the rest."

A significant number of research studies have been published within the recent past that clearly validate the *selective attention* concept. (Eimer Jan 1996 and Aug 1996; Freedman et al 1987; Garcia-Larrea et al 1995; Haken 1998; Kappas et al 1997; Kropotov et al 1997; Lorist et al 1996; White et al 1997; Ward et al 1996; Trejo et al 1995; Karayannidis et al 1995; Kenemans et al 1995; LaBerge 1990; Treisman 1998; Van Der Molen et al 1996; and Woldorff et al 1998).

Of course, *psychological set* did not originate from Ruch (1948). It should be recognized that the term *psychological set* was used by Titchener to describe the gelling of attitudes and beliefs in 1914. In both America and England (*psychological set* was called *mental set*), research on *psychological set* blossomed in the 1950s in two areas: Cognitive Sciences and Industrial/Organizational/Political (Applied) Psychology.

In the Cognitive Sciences, *psychological set* was a cornerstone of the new discoveries about reification (the mind's natural tendency to simplify and categorize events). Difficulties in cognitive flexibility are attributed to functional fixedness, a sub-category of *psychological set.*

In Applied Psychology, *psychological set* is used to integrate diverse findings of human tendencies towards manipulation, prejudice, persuasion, and even brainwashing (see 1977 Senate Hearings). Advertisers and marketers study *psychological set* to sell products; political reformers study *psychological set* to explain cognitive dissonance and psychological profilers study *psychological sets* to explain deviant behaviors.

As noted in above (1977 Senate Hearings), *psychological set* was used by law enforcement to help explain narco-interrogations. Even today, *psychological set* is used to explain how people shift their focus based on their background, knowledge and

experience. The applications of *Psychological set* theory today are diverse; we know some sets are inflexible (functional fixedness) and others seem to be more malleable, based on immediate needs (consumer preference sets in marketing studies).

Forced Choice and Psychological Sets

It is quite apparent that Backster did not invent the concept of *psychological set.* His application of *psychological set* to forensic psychophysiology using the polygraph was unique, but a logical application of two other then-new discoveries: (1) The double-bind effect, and (2) cognitive dissonance.

Double-Bind Effect: Backster's concept of focusing the issues to clearly establish one of two mutually exclusive *psychological sets* is similar to another discovery in cognitive sciences, the double-bind effect. Backster's method clarifies immediate threats to being deceptive to one of two situations: Lying to the relevant issue or lying to the control issue. A guilty person can lie to both, but is more threatened by the crime-related lie (Anti-climax dampening). In double-bind studies, a situation is constructed where only two escapes are obvious (other options are logically possible, but not allowed). For example, you can be truthful or you can lie; this restricts your obvious choices to only two - a double-bind. Under clear double-bind conditions, persons make forced choices based on Fechner's Law of Least Resistance - taking the easy way out.

Cognitive Dissonance: Set choices are not just based on the options available; choices are also influenced by the stakes involved in each choice. Festinger found that in double-bind situations where the stakes were high, people often developed cognitive dissonance (c. 1953), holding more firmly to beliefs proven wrong later. This sounds like the opposite of Anti-Climax Dampening, but in fact it is Backster's complimentary concept of Dampening Outside Issue,[2] a way of dampening the fear of any other issue outside the double-bind.[3] Persons hold to an old belief in part because they share a wider belief that outsiders do not really understand (as perhaps in Ruby Ridge). Cognitive dissonance is based on the link between the belief about a specific event and the threat of abandoning that belief, creating a conflict resolved by outside issues, not the specific issue. In forensic psychophysiology failure to account for the threat associated with outside beliefs will negate the desired double-bind effect about the target issue, as Backster had predicted.

In summary, Backster's synthesis of the concept of *psychological set* is entirely consistent with the latest thinking and a long tradition in academic psychology and forensic sciences. Furthermore, as shown above, the definition of the term *psychological set* by Ruch (1948) is consistent with the definition of the contemporary term *selective attention* by Bootzin et al 1993, a current, validated psychological concept. Hence *psychological set* remains alive and well in its continued role within Backster's Anti-Climax Dampening Concept. The polygraph community should be proud to embrace it as fundamental to all scientifically-based protocols.

References

ASTM (1999). Standard Guide for Terminology Relating to Forensic Psychophysiology. Committee E-52, Standard E-2035-99. American Society for Testing and Materials, Conshohocken, PA.

[2] Backster developed symptomatic questions to address outside issues that may interfere with the examinee's psychological set, self-directed onto the relevant or comparison questions.

[3] Double-bind is also a technique of hypnosis, and of interrogation; however, research on double-bind is wider than both.

Backster, C. (1960). Anti-Climax Dampening Effect. School Handout. Backster School of Lie Detection, New York, N. Y.

Backster, C. (1962, May-June). Methods of strengthening our polygraph technique. Polygraph and Interrogation. Academy for Scientific Interrogation. New York, NY.

Backster, C. (1963, October). Anticlimax Dampening Concept. *Military Police Journal*, Augusta, GA.

Backster, C. (1964, February). Polygraph spot analysis versus non-localized analysis. Academy for Scientific Interrogation, *Law and Order*, 12(2). New York, NY.

Barland, G. H. (2001, February 21). Personal communication.

Bauer, B., Besner, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Experimental Psychology*, 51: 61-68.

Beck, T. R., Berry, R. L. (1963, January). Orientation on lie detection examination: Glossary of lie detection terminology. USAMP, Fort Gordon, GA.

Berry, R. L. (1965, May). Psychology of the Backster Zone Comparison Technique. USAMPS, Fort Gordon, GA.

Eimer, M. (1996). ERP modulations indicate the selective processing of visual stimuli as a result of transient and sustained spatial attention. *Psychophysiology*, 33(1): 13-21.

Eimer, M. (1996). Selection and inhibition in attention and motor activation. Abstract of paper presented at 36th annual meeting of the Society for Psychophysiological Research. *Psychophysiology*, 33(1S): S12.

Festinger, Leon (1957). Theory of Cognitive Dissonance. Stanford University, Stanford, Calif.

Fechner, Gustav. (1966). Translated by Helmut E. Adler. *Elements of Psychophysics.* Holt, Rinehart and Winston, Inc. New York.

Freedman, R., Adler, L. E., Gerhardt, G. A., Waldo, M. C., Baker, N., Rose, G. M., Drebing, C., Nagamoto, H., Bickford-Wimer, P., and Franks, R. (1987). Neurobiological studies of sensory gating in schizophrenia. *Schizophrenia Bulletin*, 13: 669-678.

Garcia-Larrea, L., Lukaszewica, A. C., and Mauguiere, F. (1995, November). Somatosensory responses during selective spatial attention: The N120-to-N140 transition. *Psychophysiology*, 32(6): 526-537.

Haken, M. D. (Ed) (1998, February). Visual Attention: Understanding the mind's eye. Basic research in psychological science. *Observer*, American Psychological Society, Special Issue, AHI Report 6.

Hill, W. F. (1970). *Psychology: Principles and Problems.* Philadelphia, PA: J. B. Lippincott. Co.

Honts, C. R. (2000). A brief note on the misleading and the inaccurate: A rejoinder to Matte (2000) with critical comments on Matte and Reuss (1999). *Polygraph*, 29(4): 321-325.

Kappas, A., and Smith, C. A. (1997). The psychophysiology of emotion: The role of stimulus relevance, motivational congruence, and coping potential. Abstract paper presented at the 37th annual meeting of the Society for Psychophysiological Research. *Psychophysiology*, 34(1S): S4.

Karayannidis, F., Andrews, S., Ward, P. B., and Michie, P. T. (1995). ERP indices of auditory selective attention in aging and Parkinson's disease. *Psychophysiology*, 32(4): 335-350.

Kenemans, J. L., Smulders, F. T. Y., and Kok, A. (1995). Selective processing of two-dimensional visual stimuli in young and old subject: Electrophysiological analysis. *Psychophysiology*, 32(2): 108-120.

Krapohl, D. (2001, February 12). Personal communication.

Kropotov, J. D., Etlinger, S. C., and Ponomarev, V. A. (1997). Human multiunit activity related to attention and preparatory set. *Psychophysiology*, 34(5): 495-500.

LaBerge, L. D. (1990). Attention. *Psychological Science*, 1: 156-162.

Lorist, M. M., Snel, J., Kok, A., and Mulder, G. (1996). Acute effects of caffeine on selective attention and visual search processes. *Psychophysiology*, 33(4): 354-361.

Matte, J. A. (1996). *Forensic Psychophysiology Using The Polygraph.* Williamsville, NY: J. A. M. Publications.

Morris, C. G. (1973). *Psychology: An Introduction.* Englewood Cliffs, NJ: Prentice-Hall.

Ruch, F. L. (1948). *Psychology and Life.* Chicago, IL: Scott Foresman.

Reber, A. S. (1995). *The Penguin Dictionary of Psychology.* Penguin Books, Ltd. London, England.

Titchener, E. B. (1914). *Systematic psychology: prolegomena.* New York: Macmillan (last edition, 1929)

Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A: 202-237.

Trejo, L. J., Ryan-Jones, D. L. and Kramer, A. R. (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology*, 32(4): 319-328.

U. S. Senate Hearings (1977). Project Mkultra, The CIA's program of research in behavioral modification. http://www.parascope.com/ds/documentslibrary/documents/mkultrahearing/mkultraHearing04.htm.

Van Der Molen, M. W., Somsen, R. J. M., and Jennings, J. R. (1996). Does the heart know what the ears hear? A heart rate analysis of auditory selective attention. *Psychophysiology*, 33(5): 547-554.

Ward, R., and Goodrich, S. (1996). Differences between objects and nonobjects in visual extinction. A competition for attention. *Psychological Science*, 7(3): 177-180.

White, P. M., and Yee, C. M. (1997). Effects of attentional and stressor manipulations on the P50 gating response. *Psychophysiology*, 34(6): 703-711.

Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., and Bloom, F. E. (1998, May). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, 35(3): 283-292.

A Brief Rejoinder to Matte & Grove Regarding "Psychological Set"

Donald Krapohl

Keywords: psychological set, rejoinder, theory, selective attention

Drs Matte and Grove have provided the community with an excellent and expansive justification for retaining the expression *psychological set* in the polygraph lexicon. It was prompted in part from an informal e-mail I had sent Dr. Matte earlier this year. Because mine is the e-mail that launched ten thousand words (to torture the words of Homer) I feel obligated to reply briefly and courteously here to give the readers a better understanding of the logic behind my personal communication with Dr. Matte.

If I may be permitted to restate the Matte and Grove position, *psychological set* is the scientifically agreed upon expression for the selective attention that is the driver for the differential arousal patterns used for diagnostic purposes in polygraphy. The expression appeared in a 53-year-old introductory psychology textbook, and was recognized by Cleve Backster as the explanation for larger reactions on relevant questions by the guilty, and larger reactions on the probable-lie comparison (PLC) questions for the innocent. No scientific discoveries since 1948 have superseded this original theory, and *psychological set* as a concept remains as vital today as it did since it was first introduced to polygraphy.

Psychological set has served as an important expression in teaching and communicating polygraph principles within the profession for a couple of generations. I depart from Matte and Grove's position on a couple of points, which will be detailed here. First, I would maintain that the assertion that selective attention is the phenomenon that causes the differential arousals, though a longstanding premise in polygraphy, is not the established truth that many of us would believe. In fact, selective attention as understood by scientists may not even be the same as the selective attention in polygraphy as described by Matte and Grove. If one reviews the literature on selective attention, one finds that the paradigms are much different from those in polygraphy. In the standard selective attention research, more than one stimulus is presented to the subject simultaneously, so that researchers can monitor which stimulus is chosen. For example, several symbols are presented concurrently on a computer monitor, or different information is presented at the same time in each ear of the subject, and scientists infer effects or processes from the subject's recall or behavior. In the selective attention paradigms, examinees genuinely attend to one stimulus at the cost of attention to other stimuli in the same sensory field. In contrast, in polygraphy the stimuli (test questions) are presented serially, not simultaneously. There is no competition for attention in the cognitive sense, because the examinee can, and does attend to every question for some period after its presentation. If that were not true, the examinee would be unable to give a deliberate answer to each of the questions. Therefore, the selective attention model as the operating mechanism explaining examinee response patterns in the CQT is at least incomplete.

Here is another view. Most would agree that the polygraphic differential-arousal phenomenon probably involves more than one cognitive component, and possibly many. Attention, as the frontloading component, is necessary, certainly, but is probably not sufficient in itself. A more compelling case can be made that the arousals are largely due to a subjective assessment of the threat value of the question, rather than from the exclusiveness of attention. In other words, it is the unique salience of the test question, not how much attention is given that determines the intensity of the body's affective response.

Salience probably mediates the attention, but it is the salience alone that evokes the response. The centerpiece of the psychological set hypothesis is the notion that PLCs somehow compete for limited attentional resources, and larger responses signal more attention. If one were able to control for attention, or even direct a subject's attention, it is reasonable to speculate, within limits, that the differential responses would still take place, contrary to the predictions that naturally flow from the psychological set hypothesis. However, a theoretical shortcoming is not the most significant problem for the expression *psychological set.* There is a larger, more intractable issue: precedence.

In 1999 the American Polygraph Association Board of Directors voted to discontinue use of the expression *control question* because the term *control* already had an established meaning in the larger scientific community that was different from what we attributed to the term. DoDPI made the change from *control question* to *comparison question* a year earlier, for the same reason, and much of the mainstream scientific literature has made the transition. Many of us learned to use *control question* when we were but infant polygraphers, and updating our language in recent times has included some painful moments. A similar conflict of meaning exists for *psychological set.* This expression, as Drs. Matte and Grove cite from my personal e-mail, relates to expectancy, not attention, though they were comfortable in folding expectancy in with attention rather than addressing their important differences. What was not clearly articulated in either my e-mail or the Matte and Grove article is what the expression *psychological set* means specifically to the much larger psychological community. I hope to correct that now with an example. Below is a synopsized demonstration of psychological set taken from the article by Cory (1990) entitled "Psychological set and the solution of anagrams" to give the reader what I hope is a more straightforward appreciation of the scientifically accepted meaning of *psychological set.* This paradigm works reliably, and readers might try it themselves with students or colleagues.

One begins with two lists of scrambled words (anagrams). They are:

List 1. SIFH, CALEM, NUKKS, SEUMO, BAZER, EAP

List 2. NORC, NOONI, MATOOT, PREPE, TEBE, EAP

One list is given to a group of subjects, and the other list to a second group. The subjects are instructed to unscramble these anagrams. Each subject works independently, and completes each word before moving to the next.

The first five anagrams on each list can only be unscrambled one way. The unscrambled words on List 1 are all animals (FISH, CAMEL, SKUNK, etc.), while List 2 consists of vegetable words (CORN, ONION, TOMATO, etc). However, what the subjects don't know is that the sixth anagram on both lists can be rearranged as either APE or PEA. Researchers find about 80% to 90% of subjects with the animal list solve the sixth anagram as APE, and an equal percentage of subjects with the vegetable list interpret the same anagram as PEA. The experimenter, using the principle of psychological set, largely shapes a subject's choices by giving one list or the other to the subject. As one can readily see, mainstream scientific use of *psychological set* (choices affected by prior stimuli) has little in common with our discipline's use of *psychological set* (physiological reactions to one of two categories of questions), and as one might expect, the difference can be an issue in the communication between polygraphers and behavioral scientists. At some time it may become necessary to reconcile our language with them, and either we, or the American Psychological Association, will have to adjust. Those in our profession who believe that the other camp has to change will have much work to do.

Unfortunately, I cannot offer a solution to this dilemma, and find that I have reluctantly come to the same conclusion regarding the continuance of *psychological set* as did Drs. Matte and Grove, though by another route. Unlike the *control question* problem, there is no readily available term for *psychological set* that we in polygraphy can

turn to as a substitute. That new term, in whatever form it takes, will be the product of theory development. In the meantime, I believe we should retain *psychological set* as a placeholder, an interim expression, but also we should be aware that it is imperfect, and may give rise to confusion when used outside of our profession. I support Matte and Grove in continuing the use of *psychological set* within polygraphy, until we can do better.

Reference

Corey, J.R. (1990). Psychological set and the solution of anagrams. *Activities Handbook for the Teaching of Psychology*, Vol 3.

A Belated Look at Symptomatic Questions

Donald J. Krapohl and Andrew H. Ryan

Abstract

Backster (1962) suggested that unresponsiveness during polygraph testing was often caused by the examinee being apprehensive about issues outside the scope of the test questions. These concerns drew the examinee's attention away from the diagnostic questions, thereby diminishing differential responsivity, and giving rise to inconclusive polygraph outcomes. Backster named this phenomenon the "super-dampening concept", indicating that all responses during testing were dampened by the examinee's diversion of attention to outside concerns. Backster recommended the use of symptomatic questions, which he reported were useful to identify when the examinee was more worried about an outside issue than the test issues. In the present project, 100 field polygraph cases were subjected to 7-position scoring and an automated form of the Rank Order Scoring System (ROSS, Honts & Driscoll, 1987) to investigate the relationship between dampened responsiveness to diagnostic questions, as represented by polygraph scores, and responses to symptomatic questions. Consistent with previous research, reactions to symptomatic questions had no correlation with the strength of polygraph scores in either the manual 7-position scorings or the automated ROSS. The predicted super-dampening effect was not found. The future of symptomatic questions is discussed.

Key words: outside issue, Rank Order Scoring System, super-dampening concept, symptomatic question, validity

One of the abiding questions for the field of polygraphy is why examinees occasionally fail to react to polygraph questions to which they are lying. Even when the test conditions are ideal, the examination is competently conducted, the examinee appears to be suitable, and the test issues are clearly defined, there are instances where examinees do not react physiologically during polygraph testing. There are possibly as many theories for this phenomenon as there are writers on the topic. Proposed causes for this unresponsiveness have been adrenal exhaustion (Reid & Inbau, 1966), emotional lethargy (McInerney, 1961), and even a fatalistic attitude or a lack of a sense of guilt (Arther, 1977). Despite these suggestions from the field, there is no research known to the present writers that directly answers this question.

Cleve Backster (1962) proposed a theory based on attention. Called the "super-dampening" concept, Backster asserted that examinees were often distracted by issues outside of the polygraph session, which can limit their attention to the polygraph questions. He wrote:

Acknowledgements

We are grateful for the assistance of DoDPI Instructor William Gaschler for scoring the cases used in this project, and to Kay Williams for data management. We also appreciate the methodological guidance and critical manuscript review of Dr. Dean Pollina and Dr. Stuart Senter of DoDPI.

Donald J. Krapohl is a Researcher with DoDPI, and Andrew Ryan Ph.D. is the Chief of DoDPI's Research Division. The opinions expressed in this article do not necessarily represent those of the Department of Defense or the US government. Reprint requests should be directed to: Donald J. Krapohl, PO Box 10411, Ft. Jackson, SC 29207, USA, or e-mail to dkrapohl@aol.com.

> Based on all indications to date it seems apparent that the major portion of "inconclusive" polygraph examination results are not caused by a subject who is a non-reactor at that time, but are caused by the sensory perception "set" of the subject being diverted or tuned in on the wrong protective frequency. It now seems apparent that this is caused by the subject's anticipation of a question involving some outside factor. With such a person the outside issue, about which he is so apprehensive, is much more important to him - or more directly affects his well-being - than does the reason for the polygraph examination, thus causing a "super-dampening" of all responses that ordinarily would have occurred (Backster, 1962, p 65).

Based on his observations, Backster concluded that suppressed responsivity, and the resulting inconclusive outcome, is largely attributable to outside issues competing for an examinee's attention. As a remedy, Backster advocated the inclusion of additional questions, called "symptomatics," among the other polygraph questions. The more common phrasing of symptomatic questions are: "Are you completely convinced that I will not ask you a question during this chart that has not already been reviewed?" and "Is there something else you are afraid I will ask you a question about even though I told you I would not?" Backster warned that when responses occurred to these symptomatic questions, "the subject's outside issue is causing an inter-play on his charts" (Backster, 1962, p 67). In other words, it was an indication that the examinee was overly concerned that the examiner would inquire about other topics beyond those on the reviewed question list, and that the examinee might not react to the other diagnostic questions.

Symptomatic questions are now found in all of the major techniques that are offsprings of the Backster Zone Comparison Technique (ZCT): the Department of Defense Polygraph Institute (DoDPI) ZCT, Matte Quadri-Track, and the Utah ZCT. Though the logic of the super-dampening is credible, empirical data to support the use of symptomatic questions, or even verify the underlying super-dampening concept is surprisingly meager. Capps, Knill and Evans (1993) set out to test the hypothesis that the mere inclusion of the symptomatic question would reduce the incidence of inconclusive results. They directed three field polygraph examiners to alternately conduct their live criminal examinations with symptomatic questions, and the next examination without symptomatic questions. Capps et al (1993) collected 75 field cases with symptomatic questions, and 75 without symptomatic questions. Based on the original examiner scores, they found that among the symptomatic question cases, there were 4 inconclusive results from manual scoring, whereas the cases without symptomatic questions had 12 inconclusive results. Using an early version of the PolyScore algorithm, the numbers were 5 and 8, respectively. Capps (1993) subjected the manual scores to a chi-square statistical treatment, and concluded that there was a significant effect for the presence of symptomatic question. They reported that their data supported Backster's claim that symptomatic question would significantly reduce the number of inconclusive calls. These findings have been touted by proponents as evidence of the efficacy of the symptomatic question (Matte, 1996; 2000).

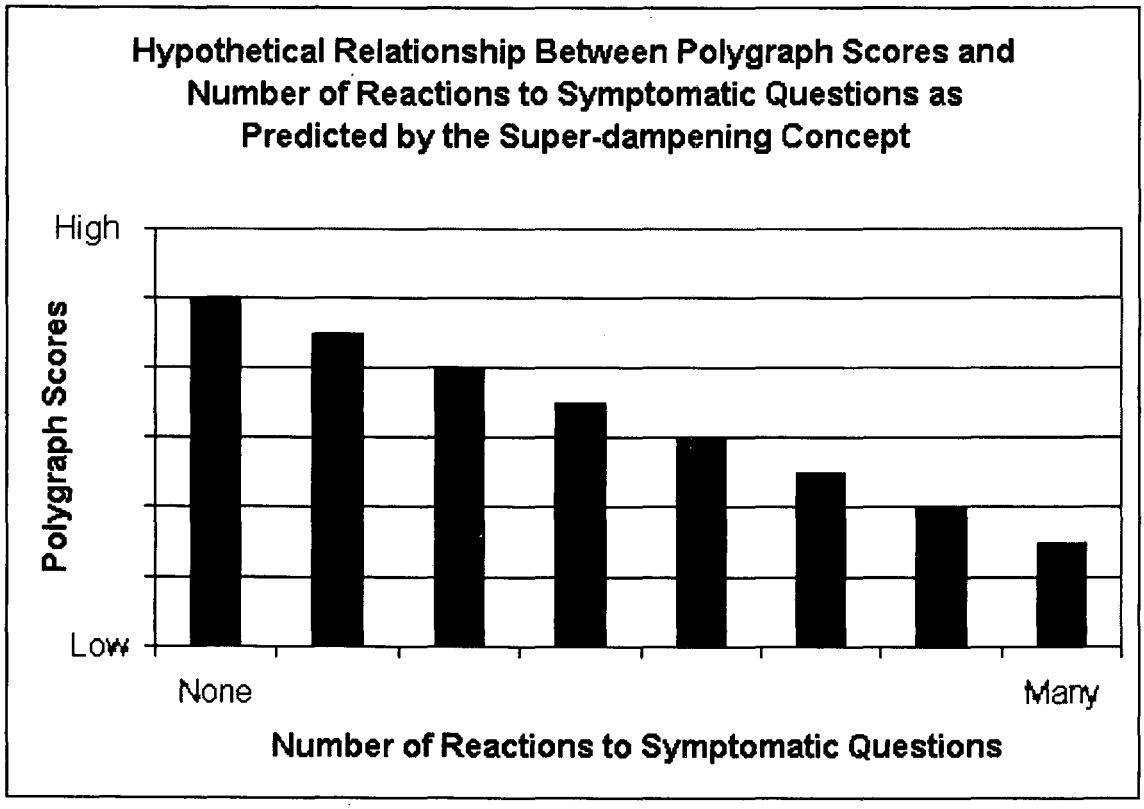
It should be noted that in the Capps et al (1993) study there were only a relative handful of inconclusive findings by the original examiners, merely 4 and 12, and as such, subject to more variability than if the sizes had been more substantial. This does not disprove their conclusions, of course, but very small sample sizes are notoriously unstable, and generalizing their findings to the hundreds of thousands of polygraph examinations conducted each year would be premature. Moreover, had the PolyScore decisions been used instead of the examiner calls, the outcome would have been very different. Using the Capps (1993) data for the PolyScore decisions, we conducted a chi-square for with-symptomatics and without-symptomatics for two types of decisions: conclusive, and inconclusive. Those results did not indicate a significant difference in decisions for the presence or absence of the symptomatic question (X^2 (1)=0.758, p>.05). In contrast to the manual scores, the PolyScore decisions were unaffected by the symptomatic questions.

Since this automated algorithm could not be influenced by anything other than the physiological data, we are more inclined to accept these results than those produced by the examiners who interacted with the examinee, conducted the polygraph session, may have known the study hypothesis, and did the manual scoring. Conclusions based on these manual scorings may be confounded by the conditions under which the scores were produced. As a final note, the Capps (1993) manual scoring data were reanalyzed by Honts, Amato, and Gordon (2000), who found that even if one accepted the original examiner scores, the effect of the symptomatic question was quite modest, accounting for only 3% of the variance.

The only other direct test of the symptomatic question was conducted by Honts, Amato, and Gordon (2000). They polygraphed volunteer subjects in a 2 (guilty, innocent) X 2 (outside issue present, absent) X 2 (outside issue question present, absent) between-subjects mock-crime design, with 24 subjects in each condition. In his study, Honts et al found that the symptomatic question showed no validity for detecting or reducing the effects of outside issues. Statistical analyses of the raw physiological data found little useful information in the reactions to the outside issue questions. He concluded that the power of the symptomatic question to detect outside issues, or reduce their effects, was inconsequential.

The super-dampening concept is still compelling, however. Given that an examinee's attentional resources are finite, and some portion is required for the examinee to produce physiological responding to test questions, it would seem reasonable that there should be some sort of relationship between the strength of polygraph scores and the presence or absence of outside issue concerns. It would also seem plausible that symptomatic questions might identify those examinees who harbor those outside concerns. One would expect that, as response intensity increases to the symptomatic questions, differential responsivity of the diagnostic questions should decrease. Below is the expected relationship in graphic form (See figure 1).

Figure 1.



It may be their intuitive appeal that contributed to the acceptance of the super-dampening concept and symptomatic questions long before they were subjected to independent verification. Since the symptomatic question was created to provide a physiological gauge for how distracting an outside issue was for the examinee, it became our interest to assess the association between the physiological responses to symptomatic questions and polygraph scores. If field data uncovered a relationship, it could provide substantiation for the Backster super-dampening concept, and support the continued or even expanded use of symptomatic questions. The Capps et al (1993) study was designed to determine whether the mere presence of symptomatic questions affected inconclusives, and though we do not know how to interpret their conclusions for the reasons stated earlier, our immediate question pertaining to the relationship between inconclusives and responses to symptomatic questions could not be answered by their methodology. The Honts et al (2000) data suggest that symptomatic questions did nothing meaningful in that regard. However, it is not known whether their findings would generalize beyond the laboratory. We therefore set out to scrutinize field data for evidence that responses to symptomatic questions correlated with reduced differential responding to the relevant and comparison questions.

Methodology

Cases

One hundred field cases were randomly drawn from the DoDPI confirmed case database. Cases in this database were all confirmed by confession, medical tests, or other irrefutable evidence. Because evidence separate from the polygraph decision was the criterion for inclusion in the DoDPI database, there were cases in which the original polygraph decisions were inconclusive or in error. The criteria for selection for the present study were that the cases had to be single-issue DoDPI ZCT examinations in which three relevant and three probable-lie exclusionary comparison questions were used. Symptomatic questions were placed in positions 3 and 8. All were field cases conducted by federal, state or local polygraph examiners. Half were from confirmed deceptive cases, and the other half were confirmed nondeceptive cases. No demographic variables were considered in the selection process.

Human Scoring

An experienced polygraph examiner scored the 100 cases using the 7-position scoring system (DoDPI, 2001). He also recorded his subjective assessments of the presence and absence of responses to the outside issue questions, by channel and question. The examiner was kept blind to ground truth, base rates, and case facts.

Data Analysis

It was possible to test the effect of the super-dampening concept by looking at tabulations of responses to the symptomatic question and the absolute value of scores to the cases, since the concept predicts an inverse relationship, as seen in Figure 1. A Pearson's r was conducted for the absolute value of the polygraph scores by the number of reactions noted by the scorer to the symptomatic questions. Statistical significance was set at .05.

Automated Scoring

As a means of independent analysis, these same 100 cases were analyzed using a slight variant of the Rank Order Scoring System (ROSS) (Honts & Driscoll, 1987; Krapohl, Dutton & Ryan, 2001). The variation from standard procedure was that the symptomatic questions were also ranked, which is not normally done with the ROSS. The Kircher features were used in the rank assignment: respiration line length, EDA amplitude, and blood volume amplitude. The Kircher features were measured and recorded by a software package (Extract, ver 3.0). Ranking of the Kircher features was also automated, using the functions found in Microsoft Excel.

The physiological features for the symptomatic, relevant, and comparison questions were ranked in order of magnitude, within channel and within chart. For the EDA and blood volume, the largest responses received the highest ranks. For respiration line length (RLL), the shortest line lengths received the highest ranks, because RLL is inversely related to physiological arousal

(Timm, 1982). Ranks were summed separately for relevant, comparison, and symptomatic questions.

Data Analysis

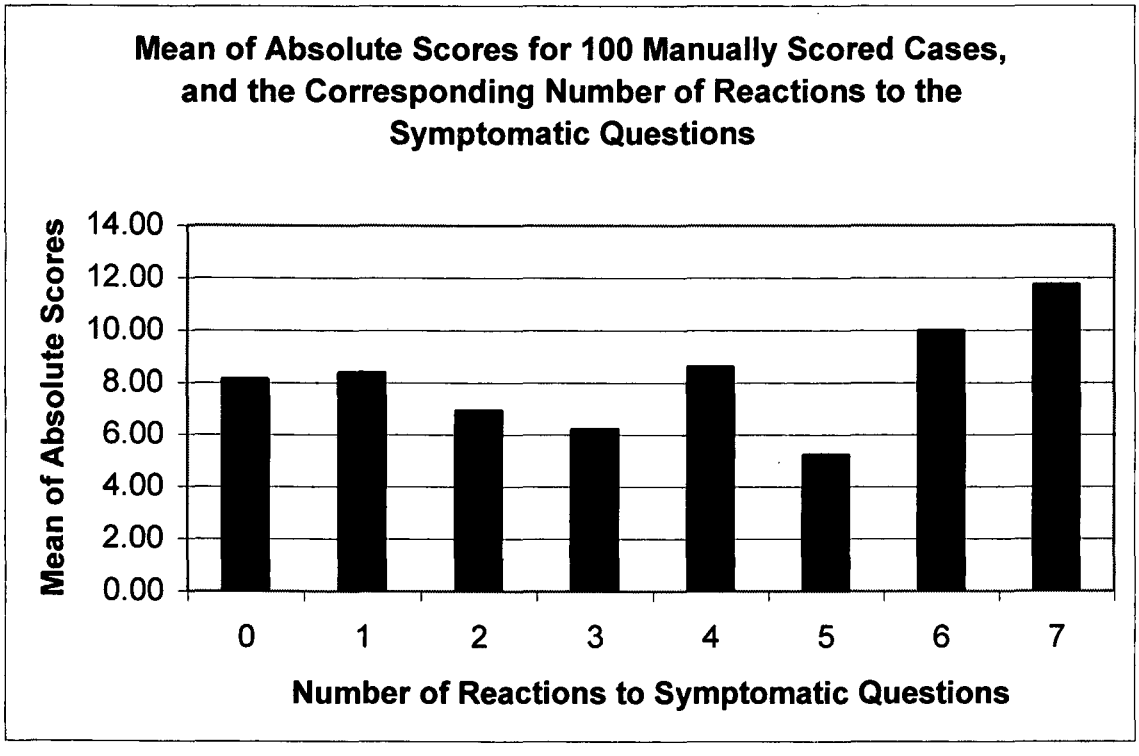
The sum of the ranks of relevant questions was subtracted from the sum of the ranks of the comparison questions, creating a measure of differential arousal. The super-dampening concept holds that larger responses to symptomatic questions should predict dampening of differential responses. A Pearson's r was applied to the absolute value of the difference scores and the summed ranks of the symptomatic questions. Statistical significance was set at .05.

Results

Human Scorer

The correlation for the number of reactions to symptomatic questions and the total of absolute scores for all cases was weak (r=0.07), and not significantly greater than chance. Figure 2 shows the absolute values of scores across the number of reactions to symptomatic questions.

Figure 2.



Automated Scoring

The correlation between the absolute difference scores of the ROSS and the average ranks of the symptomatic questions was also quite weak (r=.06), which was not significantly greater than chance. Similar analyses were conducted for the nondeceptive and deceptive cases separately. Nondeceptive cases showed a poor and non-significant relationship between the two variables (r=.04). Deceptive cases also showed a small correlation (r=-.18), but that coefficient also failed to achieve statistical significance.

Discussion

The human scorer and the automated analyses of the present data converge on the finding that the symptomatic question did not predict suppressed polygraph scores in field

cases. The present data also corroborate the laboratory findings of Honts, Amato and Gordon (2000), who found no value in the symptomatic question. Notwithstanding Capps' (1993) conflicted data, there exists no published evidence that the symptomatic question has demonstrated empirically it can do as advertised. Against this backdrop, it is our view that the existing evidence reopens the debate as to whether chart time is being well invested with these questions.

There is a strong sentiment among advocates of symptomatic questions that they provide some value. Many field examiners can point to individual cases where symptomatic questions did uncover an outside issue, and once the issue was resolved, the case concluded with a valid decision. Of course, one can never know from those isolated cases whether the symptomatic questions only identified the presence of an outside issue, or whether, in a circular manner, they actually caused the concern that gave rise to the reactions to the symptomatic questions in the first place. Some experienced and competent field examiners have complained that the inclusion of symptomatic questions raises suspicion in many examinees rather than the alleviating it. This is especially true in non-US cultures, where some examiners have abandoned symptomatic questions because of the complications they introduce into an otherwise straightforward examination process. And when one considers what they are designed to do, it might be argued that the symptomatic question changes an explicitly single-issue (one crime) polygraph examination into an implicitly multiple-issue (this crime plus other crimes) polygraph examination.

To provide balance, let us also state that no published data indicate that symptomatic questions significantly interfere with polygraph examinations. We note that there may be isolated cases where examiners perceived some benefit from using symptomatic questions, just as others perceive they create problems. Reasonable people can disagree on this issue, but we must acknowledge that anecdotes or selective personal recollections will not provide the best answer, nor will reference to authority: Only data can move us forward. The preponderance of the independent evidence now suggests that symptomatic questions probably do nothing reliably, neither good nor bad, across large numbers of examinations. As such, there are implications for the polygraph community. A collective decision is in order as to whether there is greater value in shortening the test question sequence, continuing with the traditional practice, using the symptomatic question on a case-by-case basis, or developing and validating a question that can demonstrate value. One type of question with at least face validity in the current environment is a countermeasure question. We would suggest, from the trends in the published literature, that counter-measure attempts are more problematic to polygraph validity than are outside issues, and they may be detected or deterred with a direct question. Considerable validation work is needed for whatever question might supercede the symptomatic question before we would be prepared to make a recommendation.

The conclusions of this paper will no doubt challenge long-accepted, if not empirically established doctrine for polygraph examiners trained to use symptomatic questions. Because of its controversial nature, the present authors provided an advance copy of this paper to Cleve Backster, the innovator of many polygraph concepts including those investigated here, and invited him the opportunity to comment. His response is published contiguous to this article. In addition, we will take the additional step of making our raw data available to anyone wishing to conduct a reanalysis. It is our goal to use the best scientific information to help the discipline develop a "best practices model" for polygraphy.

References

Arther, R.O. (1977). Non-reactors. *The Journal of Polygraph Science*, 12(2).

Backster, C. (1962). Methods of strengthening our polygraph technique. *Police*, 6(5), 61-68.

Capps, M.H., Knill, B.L., & Evans, R.K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22(4), 285-298.

Department of Defense Polygraph Institute (2001). Test Data Analysis (Manual Scoring Process) 7-Position Numerical Analysis Scale. Ft. Jackson, South Carolina.

Honts, C.R., Amato, S., & Gordon, A. (2000). Validity of the Outside-Issue Questions in the Control Question Test. Final Report on Grant No. N00014-98-1-0725. Boise State University.

Honts, C.R., & Driscoll, L.N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16(4), 241-257.

Krapohl, D.J., Dutton, D.W., & Ryan, A.H. (2001). The Rank Order Scoring System: Replication and extension with field data. *Polygraph*, 30(3).

Matte, J.A. (1996). *Forensic Psychophysiology Using the Polygraph: Scientific Truth Verification - Lie Detection.* Williamsville, NY: J.A.M. Publications.

Matte, J.A. (2000). *Examination and Cross-Examination of Experts in Forensic Psychophysiology Using the Polygraph.* Williamsville, NY: J.A.M. Publications.

McInerney, C.A. (1961). Emotional lethargy. Paper presented at the Eighth Annual Meeting of the American Academy of Polygraph Examiners, Washington, D.C.

Reid J. E., & Inbau, F.E. (1966). *Truth and Deception: The Polygraph ("Lie Detector") Technique.* Williams & Wilkins: Baltimore.

Timm, H.W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10(1). 47-51

A Response To Krapohl & Ryan's "Belated Look At Symptomatic Questions"

Cleve Backster

Key words: rejoinder, super-dampening concept, symptomatic questions

This is in response to the article entitled "A Belated Look at Symptomatic Questions", by Donald J. Krapohl and Andrew H. Ryan appearing in this same issue of *Polygraph*, an official publication of the American Polygraph Association.

A significant portion of the commentary contained within the Krapohl-Ryan report involves criticism of research on the effectiveness of symptomatic questions conducted by Capps, Knill and Evans in 1993. Their report states the following: "These findings have been touted by proponents as evidence of the efficacy of the symptomatic question (Matte, 1996; 2000)." As I wish to devote my full attention in this rebuttal to the more recent Krapohl-Ryan field study I have asked Dr. James Allan Matte to respond to Krapohl-Ryan's criticism of the Capps, Knill and Evans study and I am requesting that his response be included as an appendix to my response.

I note in Donald Krapohl's correspondence to me dated April 12, 2001 that I am welcome to request the raw data used in the Krapohl-Ryan study. In addition, the final paragraph of the report states that the authors "will take the additional step of making their raw data available to anyone wishing to conduct a reanalysis". My concern is not their analysis of the raw data but how they selected that which comprises the raw data they analyzed.

Under the "Data Analysis" portion of their report they state the following: "The super-dampening concept holds that larger responses to symptomatic questions should predict dampening of differential responses". This is not true. Polygraph charts can show larger responses to symptomatic questions and significant reactions to either the relevant questions or the comparison questions and still allow successful analysis of the charts.

The "super-dampening" concept is better explained by completing the last sentence of the excerpt used by the authors from my article entitled "Methods of Strengthening Our Polygraph Technique" published in the May-June 1962 issue of *Police*. I have underscored the omitted portion, which also accounts for comparison question dampening.

> With such a person the outside issue, about which he is so apprehensive, is much more important to him – or more directly affects his well-being – than does the reason for the polygraph examination. [This covers the deceptive individual.] This causes a "super-dampening" of all responses that ordinarily would have occurred, <u>including the dampening of all response to the reviewed stimulation questions asked the innocent suspect.</u> [This covers the truthful individual.] (Brackets by author)

With "super-dampening" the only expected reaction to occur is to the symptomatic question under discussion. Again, these are questions designed to indicate the success of the examiner in gaining, at a minimum, the very limited trust of the examinee that no unreviewed questions will be asked. The Backster Zone Comparison Test <u>"Tri-Zone Reaction Combinations Table"</u> was created in 1962 as a school handout and published in the September 1963 issue of *Law and Order* magazine. As this table is so pertinent to my critical response to the Krapohl-Rayn study I have asked that it be published as part of my reply. On this table there are four of the eight reaction

combinations that include symptomatic question reactions but only one of the four indicate "super-dampening". That is combination "C" where the examinee reacts to symptomatic questions only. In combinations "E", "F" and "G" the symptomatic question reactions have no effect on diagnostic question differential responsivity. In combination "F" the problem is caused by competing comparative question reactivity. If the Krapohl-Ryan sample of 100 field polygraph cases do not include a significant number of examinees who reacted only to the symptomatic questions, the "super-dampening concept" is then not being properly researched.

Also troublesome is another problem relating to the Krapohl-Ryan method of pooling total question reactivity data from their entire field polygraph case sample. Each chart within a Zone Comparison polygraph examination is unique for that particular examinee and more specifically predicts the examiner's success if additional charts are collected without applying the prescribed remedy, should a problem exist. This assessment is indicated by the seven-position scale symbol located next to each combination on the "Tri-Zone" reaction combination table. Such an interim status is indicated by the examiner following each chart. Provision for this procedure has been included since 1963 at the top of the Zone Comparison Standardized Polygraph Notepack pages 12, 13, 16 and 17. As the entire Zone Comparison Examination procedure is based upon the flow of "psychological set" as uniquely affecting each individual examinee, the compiling of pooled raw data can be misleading. Additional problems may relate to the differences in size and nature of individual subject reactivity and even variations in polygraph instrument characteristics and sensitivity settings.

With those properly trained in the use of the Backster Zone Comparison Technique, symptomatic questions represent an important aspect in understanding the flow of "psychological set". When significant symptomatic question responses occur it is often an indication that the examiner needs to more effectively obtain the limited confidence of the examinee as related to the avoidance of unreviewed questions embracing outside issues.

Also fundamental to the Zone Comparison Technique is a thorough understanding of "anti-climax dampening" phenomena as it relates to differential responsivity between the relevant question and the comparison question being utilized. Any student having attended one of over 150 basic polygraph examiner courses I have directed since 1962 will verify the amount of time I have devoted, through daily student recitation, to the Tri-Zone Reaction Combination table. I credit this emphasis for much of the success our graduates' later experience.

In conclusion, I strongly disagree with the Krapohl-Ryan methodology in creating pooled raw data for their study. What appears to have been completely absent is data reflecting charts exhibiting reaction only to symptomatic questions. Such examples of necessity would involve inconclusive examination results. It may well be that 40 years of symptomatic question usage, in addition to more recent general agreement within the polygraph profession that unreviewed questions should not be asked, has failed to provide researchers with an adequate sample of isolated symptomatic question reactivity. To use this as an excuse to eliminate the symptomatic question would make as much sense as removing all smoke detectors from a large apartment building complex because of a lack of fires. Until one can determine the difference between a non-reactor and an examinee focusing their "psychological set" externally, because of the fear of being asked an unreviewed question, I do suggest that we allow the symptomatic question to survive the Krapohl-Ryan "belated look".

"TRI-ZONE" REACTION COMBINATIONS
(Backster Zone Comparison Test)

COMBINATION			INDICATION		REMEDY	
A	r	r	A1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	r	A2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
		g	A3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS BECAUSE OF DAMPENING BY RED ZONE QUESTION RESPONSES INDICATES DECEPTION REGARDING TARGET ISSUE	g	A4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED	g
		b	A5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES THAT NO OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	A6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b
B	g	r	B1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE	r	B2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
		g	B3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE, AS NO OTHER ZONE IS DAMPENING OUT GREEN ZONE	g	B4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED	g
		b	B5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES THAT NO OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	B6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b
C	b	r	C1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS USUALLY INDICATES TRUTHFULNESS REGARDING TARGET ISSUE; THIS RULE NULLIFIED BY BLACK ZONE QUESTION RESPONSE	r	C2 NO REMEDY NECESSARY; RED ZONE QUESTIONS WILL BE FUNCTIONING AS DESIGNED AFTER BLACK ZONE QUESTION RESPONSE SUBSIDES	r
		g	C3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS USUALLY INDICATES DECEPTION REGARDING TARGET ISSUE. THIS RULE NULLIFIED BY BLACK ZONE QUESTION RESPONSE	g	C4 NO REMEDY NECESSARY; NO CAUSE TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; RECHECK AFTER RESPONSE TO BLACK ZONE QUESTION SUBSIDES	g
		b	C5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	C6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b
D	r	r	D1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	r	D2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
	g	g	D3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS IN ADDITION TO RED ZONE QUESTION INDICATES SERIOUS GREEN ZONE QUESTION DEFECT	g	D4 REDUCE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING SUBJECT AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS	g
		b	D5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES NO OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	D6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b

continued

(1 of 2 parts)

continued

"TRI-ZONE" REACTION COMBINATIONS
(Backster Zone Comparison Test)

COMBINATION			INDICATION		REMEDY	
E	r	r	E1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	r	E2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
		g	E3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS BECAUSE OF DAMPENING BY RED ZONE QUESTION RESPONSE; INDICATES DECEPTION REGARDING TARGET ISSUE	g	E4 NO REMEDY NECESSARY; NO REASON TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED	g
	b	b	E5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	E6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b
F	r	r	F1 PRESENCE OF RESPONSE TO ONE OR BOTH RED ZONE QUESTIONS INDICATES DECEPTION REGARDING TARGET ISSUE	r	F2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
	g	g	F3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS IN ADDITION TO RED ZONE RESPONSE INDICATES SERIOUS QUESTION DEFECT IN GREEN ZONE QUESTIONS	g	F4 REDUCE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS	g
	b	b	F5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	F6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING AN OUTSIDE ISSUE	b
G	g	r	G1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE	r	G2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
		g	G3 PRESENCE OF RESPONSE TO ONE OR BOTH GREEN ZONE QUESTIONS INDICATES TRUTHFULNESS REGARDING TARGET ISSUE; NO OTHER ZONE IS DAMPENING OUT GREEN ZONE	g	G4 NO REMEDY NECESSARY; NO CAUSE TO BELIEVE GREEN ZONE QUESTION STRUCTURE INADEQUATE; GREEN ZONE QUESTIONS FUNCTIONING AS DESIGNED	g
	b	b	G5 PRESENCE OF RESPONSE TO ONE OR BOTH BLACK ZONE QUESTIONS INDICATES OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	G6 EXAMINER MUST GAIN SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING AN OUTSIDE ISSUE	b
H		r	H1 LACK OF RESPONSE TO BOTH RED ZONE QUESTIONS STILL INDICATES TRUTH REGARDING TARGET ISSUE; THIS SYSTEM BASED ON SUBJECT CAPABILITY OF RESPONSE	r	H2 NO REMEDY NECESSARY; RED ZONE QUESTIONS HAVE BEEN FORMULATED AS IDEALLY AS POSSIBLE; RED ZONE QUESTIONS FUNCTIONING AS DESIGNED	r
		g	H3 LACK OF RESPONSE TO BOTH GREEN ZONE QUESTIONS IN ADDITION TO LACK OF RESPONSE TO RED ZONE QUESTIONS INDICATES SERIOUS GREEN ZONE QUESTION DEFECT	g	H4 INCREASE INTENSITY OF GREEN ZONE QUESTIONS BY ALTERING AGE CATEGORIES OR CHANGING SCOPE OF GREEN ZONE QUESTIONS	g
		b	H5 LACK OF RESPONSE TO BOTH BLACK ZONE QUESTIONS INDICATES NO OUTSIDE ISSUE BOTHERING SUBJECT DUE TO MISTRUST OF EXAMINER	b	H6 NO REMEDY NECESSARY; EXAMINER HAS SUBJECT'S CONFIDENCE REGARDING AVOIDANCE OF UNREVIEWED QUESTIONS EMBRACING OUTSIDE ISSUE	b

(2 of 2 parts)

Comments on Krapohl & Ryan Criticism of Capps, Knill & Evans Research on Symptomatic Questions

James Allan Matte

Key words: rejoinder, super-dampening concept, symptomatic questions

Krapohl criticizes the research by Capps et al (1993) stating that "there were only a relative handful of inconclusive findings by the original examiners, merely 4 and 12 and as such, subject to more variability than if the sizes had been more substantial." However, Krapohl fails to recognize that Capps used a respectable sample containing 150 field cases; 75 with symptomatic questions and 75 without symptomatic questions. The number of inconclusives must be considered in relationship to the total number of cases from which they were drawn. If we were for example attempting to determine the error rate between two techniques and used a sample of 500 cases for each technique and found only 5 errors in one and 15 in the other, the small number of errors should and would not disqualify the study from serious consideration. The Capps et al study offers a compelling argument for the retention and use of symptomatic questions in zone comparison tests.

Krapohl's conduct of a chi-square on PolyScore decisions in lieu of the manual examiner scores to show that there was not "a significant difference in decisions for the presence or absence of the symptomatic question" ignores Capps' raw data which shows that "Of the 75 zone comparison tests that used symptomatic questions, there were four inconclusive decisions made by the examiners, and five by PolyScore, the analytic algorithm. Of the 75 zone comparison tests that substituted irrelevant questions for symptomatic questions, there were twelve inconclusive decisions made by the examiners, and eight by the algorithm. Of sixteen inconclusive decisions by examiners, four occurred with zone tests with symptomatic questions, eight occurred with zone tests without symptomatic questions." Krapohl must know that the use of statistical formulas may support an argument but they do not prove anything. Krapohl favors the use of decisions by PolyScore in lieu of decisions made by the original examiners because of the former's absence of non-polygraphic data, suggesting that the scores from the original examiners were somehow affected by their knowledge of the case facts and their interaction with the examinee. However, it must be recognized that PolyScore's criteria for deception is limited to a total of 18 features (Olsen, 1999), whereas the Backster criteria for manual scoring totals 45 features (Matte, 2000), which provides the examiner more than twice the data in his decision making process, hence a viable explanation for the difference in the scores for the two systems, that does not include non-polygraphic data. It should also be noted that Capps et al also conducted a goodness of fit chi square test to analyze the inconclusive diagnosis data from their research study and concluded that "These data provide sufficient evidence to reject the hypothesis that there is no difference in the number of inconclusive diagnoses in those zone comparison examinations containing symptomatic questions and those that do not."

References

Backster, C. (1962). Methods of strengthening our polygraph technique. *Police,* 6(5), 61-68.

Backster, C. (1962). Backster Zone Comparison Test "Tri-Zone" Reaction Combinations Table. School Handout.

Backster, C. (1963). Polygraph technique trouble shooting. *Law and Order,* 11(9), 54-55.

Backster C. (1963), Standardized Polygraph Notepack and Technique Guide.

Capps, M. H., Knill, B. L. & Evans, R. K. (1993). Effectiveness of the symptomatic questions, *Polygraph,* 22(4), 285-298.

Matte, J. A. (2000), *Examination and Cross-Examination of Experts in Forensic Psychophysiology Using the Polygraph.* Williamsville, New York: J.A.M. Publications.

Matte, J. A. and Grove, R.W. (2001), Psychological set: Its origin, theory and application. *Polygraph* 30(3), 196-202.

Olsen, D. (1999, 1 October). Personal communication.

Final Comment on the Belated Look at Symptomatic Questions

Donald J. Krapohl & Andrew H. Ryan

Key words: rejoinder, super-dampening concept, symptomatic questions

We appreciate the thoughtful comments of Cleve Backster and James Matte, and the clarity they provide to the issue of symptomatic questions. We hope that there will be more investigation into the utility of these questions, so field practitioners will have available to them validated protocols from which to choose.

In their rejoinders, both Cleve Backster and James Matte raised important questions regarding our analytical methods, and we would like to address them briefly here.

Cleve Backster expressed concern about our use of pooled data in our statistical analysis, which is a reasonable complaint. He correctly states that effects that occur only at the ends of the distribution (such as the dampening of responsivity only when there are large reactions to symptomatic questions) can be obscured when one looks at averages of all the data lumped together. We would only point out that our histogram of the results parsed the data into sections corresponding with the number of reactions to the symptomatic question. In doing so, we provided readers an opportunity to observe the relationship of polygraph scores across several levels of reactions to symptomatic question, including the tails of the distribution. A second look at that graph will make clear that, even using only those cases with large numbers of reactions to symptomatic questions, there is a slight increase in scores, and though not statistically significant, it is in the opposite direction from what would be predicted by the super-dampening concept. We believe that this graph supplements our statistical treatments, overcoming the problems associated with pooled data.

It is important to note that the outcome of our research would have been predicted by the Honts, Amato and Gordon (2000) laboratory study, and the automated analysis results of the Capps, Knill and Evans study (1993). Honts et al's data were unambiguous, that the symptomatic questions did nothing diagnostic, and neither Cleve Backster nor James Matte has taken issue with Honts' findings. With regard to the Capps et al study, James Matte disagrees with our preference for the automated algorithm outcomes instead of the original examiner scores, contending instead that the human scorers may have had an advantage over the algorithm because they used 45 features versus the algorithm's use of only 18. We have serious reservations about this argument. First, increasing the features beyond some optimum number can decrease inter-scorer agreement. Additional features could actually end up reducing validity of a manual scoring method if the number is excessive. Moreover, unless the manual scoring features were individually validated, as the PolyScore features were, we do not know for certain whether they are doing something good, doing nothing, or doing something worse than nothing. The validation work on features is essential, and the existing research on manual scoring can support no more than a dozen features or so. More is not always better, and in the case of manual chart interpretation, very high numbers of features almost certainly indicates that unreliable or marginal elements have crept into the criteria list. Decision theory would view much more favorably the 18 validated features of the automated algorithm than 45 partially validated features used by human scorers. Also note that the examiners in the Capps study made more inconclusive decisions than the automated algorithm on the same cases. This result would be predicted if invalid features are used by the human scorers.

More to the heart of the issue of symptomatic questions, arguments for them that rely singularly on the suspect half of the

Capps et al data (non-blind manual scorings), and ignoring the growing inconvenient evidence, cannot be taken seriously. And even if one insisted on accepting the Capps manual scoring data, one must also face the fact that Capps' own data showed that the symptomatic question accounted for a tiny portion (3%) of the variance. This is hardly convincing evidence in favor of symptomatic questions. Notwithstanding the weight of longstanding tradition, at the present time we find no compelling rationale for them, and while the value of tradition may be sufficient in some quarters, it should not be confused with proof. We have no objections for use of the symptomatic questions for the former, but statements of efficacy require the latter.

We agree with James Matte's statement that statistical analyses can only be used to support an argument, not prove it. This is the essence of science. Proponents of the symptomatic question, however, are in the awkward position of not even having defensible statistical support, compounded by the accumulating empirical evidence that whatever symptomatic questions may be doing, the effect has been too small to be meaningful.

We maintain that independent research has not demonstrated that the presence or absence of reactions to symptomatic questions mean anything. To the contrary, the available relevant evidence suggests that they are irrelevant in terms of diagnosticity of the tracings. Whether symptomatic questions may serve some other purpose, we know not, nor are we aware of data that would permit us to take a position on these alternate purposes. But to repeat an admonition in our original paper, we are not warning that the symptomatic question should be immediately abandoned: They appear to be at least harmless. We do call upon those who advocate their use to provide data, reanalyses of existing data, or some concrete evidence that supports claims of efficacy of this approach. To borrow Cleve's analogy, if it can detect smoke in our home, it's a smoke detector: If not, it's decor.

Again, we appreciate the insightful replies of Cleve Backster and James Matte, and hope that there will be more open discussion on this issue.

References

Capps, M.H., Knill, B.L., & Evans, R.K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22(4), 285-298.

Honts, C.R., Amato, S., & Gordon, A. (2000). Validity of the Outside-Issue Questions in the Control Question Test. Final Report on Grant No. N00014-98-1-0725. Boise State University.

Reply to Rejoinder by Donald J. Krapohl and Andrew H. Ryan

James Allan Matte

Key words: rejoinder, super-dampening concept, symptomatic questions

To begin, it should be understood that the review and subsequent incorporation of symptomatic questions into the Zone Comparison Technique is designed primarily to reassure the examinee that no unreviewed questions will be asked during the administration of the polygraph test during which the physiological data is collected, and secondarily to identify the examinee who is not convinced that no unreviewed question will be asked during the test. It is recognized that even though the examinee may in fact be reassured by the introduction of the symptomatic questions, an orienting response to those questions may still occur, but this does not affect the purpose and result of its presence in the test structure as can be seen in Tables 1 and 2 of Capps et al's study (1993) which shows that of the 75 cases with symptomatic questions evaluated by the examiner (Table 1), only 4 inconclusives occurred, while the 75 cases without symptomatic questions evaluated by the examiner (Table 2), 12 inconclusives occurred. This data indicates that the presence of symptomatic questions reduced the inconclusives by two thirds (66 %). Even the PolyScore algorithm revealed only 5 inconclusives with the symptomatic questions versus 8 inconclusives without the symptomatic questions thus reducing the inconclusives by 37.5 percent. No application of statistical analysis can change those figures which in my view are most compelling.

Regarding the number of features in the criteria used by examiners of the Backster orientation (45) versus the number of features used by proponents of PolyScore (18), I would argue that those 45 features have been the standard chart interpretation rules of the Backster Zone Comparison Technique for the last 40 years which has withstood the test of time and many studies verifying the validity of the technique and its high accuracy (Arellano 1990, Elaad & Schahar 1985, Putnam 1983,Widacki 1982), not to mention other studies that validated components of the technique (Raskin, Barland, Podlesny 1978) and the recent study in press by this author using the files of the Virginia State Police that uses the Backster ZCT religiously, which attained a 100 percent accuracy. Statistics is the science of compiling facts and measuring the differences in data. But how one compiles those facts and data is critical to the accuracy of its results. Science is defined by Webster (1969) as "systematized knowledge derived from observation, study, and experimentation." No other polygraph technique (Backster ZCT) has been subjected to so much experimentation, observation and study.

On the question of the validity of the 45 features used by the Backster system in the interpretation of the physiological data recorded on polygraph charts, it should be recognized that several of those features such as baseline arousal when there is no change in amplitude, and changes in inhalation/exhalation ratio without a change in line length between onset of inspiration and end of expiration limb, are not within the capability of PolyScore to evaluate, yet have been recognized by the polygraph community as valid indicators of arousal/deception. Furthermore, several features were not included in PolyScore due to the low frequency of their appearance on the sample of charts used in the formation of their criteria (Matte 2000; Harris 2001). Yet those low frequency features may well be the only features manifested in a particular polygraph examination to the consternation of the polygraphist using that algorithm. Apparently, those 18 features were not adequate inasmuch as the makers of PolyScore have increased its features to 24 in their new, forthcoming version. Furthermore, there are significant differences in the way the algorithm evaluates the physiological data versus the manual

scoring using the Backster system, i.e. the algorithm does not use Backster's "Either-Or" rule in the comparison of the relevant questions to the comparison questions (Matte, in press). Finally, if the algorithms are so efficient, then why do most if not all of the United States Government agencies that use the polygraph including the Department of Defense Polygraph Institute and the Federal Bureau of Investigation require that the final determination of truth or deception be based on the manual scoring of the polygraph charts; not the algorithm (Weinstein 1997, 1998, 2000, 2001; Backster 1997, 1998, 2000, 2001; Keifer 1997, 1998, 2000, Shull 1999; Lewis 2001). Even the manufacturers of computerized polygraph systems (Lafayette, Stoelting) recommend that results of their algorithms be supported by the manual scoring of the polygraph charts as the final determination of truth or deception. Hence I will argue that the examiner decisions in the Capps, et al's (1993) study offers the more accurate diagnosis of the efficacy of the symptomatic questions in Tables 1 and 2 of aforesaid study wherein the inclusion of the symptomatic questions reduced the inconclusives by 66 percent. Less we forget, PolyScore revealed a reduction of inconclusives by 37.5 percent; not an insignificant number.

However the usefulness of the symptomatic questions does not stop there. The symptomatic questions also play another significant role. The last test question should never be a question that is used for a determination of truth or deception such as a relevant or comparison question because an examinee may relieve on the last test question. Hence the last test question should be one that ideally has orienting value unlike a neutral question, but is not used for a determination of truth or deception. Thus the symptomatic question fits that role perfectly.

References

Arellano, L. R. (1990). The polygraph examination of Spanish speaking subjects. *Polygraph*, 19(2), 155-156 (abstract).

Backster, C. (1963, 1979). Standardized Polygraph Notepack and Technique Guide. New York, N.Y.: Backster School of Lie Detection.

Backster, C. (1989, September). Backster Zone Comparison Technique: Chart Analysis Rules. Backster School of Lie Detection, San Diego, California.

Backster, C. (1962). Tri-Zone Reaction combinations: Backster Zone Comparison Test. Backster School of Lie Detection, New York, NY.

Capps, M. H., Knill, B. L., & Evans, R. K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22(4): 285-298.

Elaad, E., & Schahar, E. (1985). Polygraph field validity. *Polygraph*, 14(3): 217-223.

Harris, J. C. (2001, Aug 24). Personal communication.

Keifer, R. W. (1997 and 2000, July 13). Personal communication.

Lewis, T. (2001, Aug 24). Personal communication.

Matte, J. A. (2000). *Examination and Cross-Examination of Experts in Forensic Psychophysiology Using The Polygraph.* Williamsville, New York: J. A. M. Publications.

Matte J. A. (In press). A field study of three methods of comparison of control questions when relevant question elicits strong response. Submitted to *Polygraph.*

Putnam, R. L. (1983). *Field accuracy of polygraph in the law enforcement environment.* Unpublished manuscript.

Raskin, D. C., Barland, G. H., Podlesny, J. A. (1978). *Validity and reliability of detection of deception.* Washington, D. C.: National Institute of Law Enforcement and Criminal Justice.

Shull, K. W. (1999, October 7). Personal communication.

Webster's New World Dictionary of the American Language. Nashville, TN: The Southwestern Co.

Weinstein, D. A. (1997, 2000 and 2001, Aug 24). Personal communication.

Widacki, J. (1982). *Analiza przestanek diagnozowania w. badanich poligraficznych* (The analysis of diagnostic premises in polygraph examinations). Uniwersytetu Slaskeigo, Kkatowice. [text in polish].

"Traitors Among Us"

A book by Stuart A. Herrington

1999, Novato, CA: Presidio Press.
Traitors Among Us: Inside the Spy Catcher's World.
409 pp. Hard bound, $ 27.95.

Review by Gordon H. Barland

This is a fascinating book about how a major spy ring was detected, investigated, and prosecuted. Sergeant Clyde Conrad was arguably the worst traitor in the more than 200-year history of the U.S. Army. He betrayed NATO's detailed plans for reinforcing and defending Europe. Armed with that knowledge, a Warsaw Pact attack on Germany would likely have forced NATO to decide between losing the war or using nuclear weapons with the death of countless civilians in a crowded Europe.

Written by Colonel Herrington, then the commander of the army's Foreign Counterintelligence Activity (FCA), an organization so secret few have ever heard of it, this book reveals the immense difficulties of finding spies and even greater difficulties of building a case against them for prosecution. Despite the enormity of Clyde Conrad's betrayal and the very substantial evidence against him, the Department of Justice refused to prosecute! The FCA then turned the case over to the German government, which prosecuted him vigorously, sentencing him to life in prison without parole, the most severe punishment allowed by German law.

The FCA relies extensively on the polygraph in conducting their operations and investigations. Although Colonel Herrington is not a polygraph examiner, this book gives some insight into how the polygraph is used by a counterintelligence organization. The polygraph is mentioned in ten places (pages 36, 39-41, 59, 131, 263, 330, 371, 381, 393, and 397).

The book also describes how the FCA investigated and captured another major spy, Sgt. James Hall, who betrayed countless top secret documents containing the crown jewels of the Army's communications intercept activities in Berlin. Among other things, Hall's betrayal is believed to have caused the 1985 death of Major Arthur Nicholson, who was on a reconnaissance mission in East Germany when he was ambushed and shot by a Soviet Army guard alerted by information Sgt. Hall provided the KGB. The Soviets withheld any medical assistance while Major Nicholson bled to death at their feet.

In the Hall spy case, a missed opportunity had devastating consequences. A Turk, nicknamed der Meister, worked in the American army's auto craft shop in Berlin. In 1982 a former soldier living in Berlin accused der Meister of being a spy. He claimed der Meister had recruited Ella Pettway, an American soldier with access to top secret information, and the two were selling the secrets to the East German intelligence service. FCA investigators in Berlin interviewed der Meister. He had a plausible explanation for his activities, and claimed the soldier who was trying to get back at him out of spite for having stolen his girl friend, Ella, trumped up the allegation. The investigators asked der Meister to take a polygraph test. He unhesitatingly agreed. For reasons not mentioned in the book the investigators never scheduled him for the polygraph, closing the case as unfounded. It later developed that der Meister had recruited not only Ella Pettway, but also James Hall. We missed an opportunity to stop Hall at the very outset of his espionage career. Hall spied for the East Germans and the Soviets for five years before he was detected.

After Clyde Conrad and James Hall were convicted, the FCA turned its attention to identifying the other members of the now defunct Conrad spy ring. A number of leads were

examined on the polygraph. The author mentions that one of them, Sgt. Jeff Gregory, failed a polygraph (p. 393) and later confessed. Herrington left out some interesting details about the polygraph.

At the time Gregory was examined, the investigators believed he was probably uninvolved. The evidence linking him to the ring was largely hearsay, the investigation had turned up no corroborating evidence, and he did not fit the psychological profile of the known members of the ring. Putting him on the polygraph was a largely pro forma measure prior to closing out the investigation on him. When the polygraph turned out DI, only a half-hearted attempt was made to interrogate, for the result could very well have been in error. He had family problems and there were outside issues present. The investigation continued, and still no corroborating evidence was uncovered.

Eventually, Gregory agreed to another polygraph exam, which was conducted at Ft. Meade, Maryland by Kevin Shaw and Frank Artes. The exam again turned out DI, and this time Gregory made a detailed confession. Had it not been for the polygraph, this spy likely would have gotten clean away.

Was the polygraph instrumental in detecting the existence of either the Clyde Conrad spy ring or James Hall's activity? No. Historically, so few people are polygraphed that it is exceptionally rare for a spy to be polygraphed while he is engaged in espionage. Colonel Herrington reveals what counterintelligence officials have long known. Most major spies are not detected until somebody in the foreign intelligence service volunteers information that leads investigators to them. As Colonel Herrington put it (p. 225): "The best way to catch a spy is to recruit a spy." That's precisely what led to the uncovering of both Clyde Conrad and James Hall.

In conclusion, I recommend this book for anyone interested in espionage and the role of the polygraph in counterintelligence investigations. It presents an object lesson in the difficulties of prosecuting accused spies even when their guilt is established beyond any reasonable doubt.

The views expressed in this review are those of the author, and do not necessarily represent those of the Department of Defense or the US Government.

Closing Editorial

Donald J. Krapohl

I recall an occasion in the mid-1990s when someone said, "If you've seen one APA Editor-in-Chief, you've seen them all." This tongue-in-cheek remark was a sort of homage to Norm Ansley, who initiated the APA publications, and stayed on as Editor-in-Chief for all but a couple years since the APA publications began. When Norm turned over the reins to me in 1997 (albeit, with mixed emotions on both our parts), he had been at the helm of *Polygraph* and the *APA Newsletter* for well over a generation. He published over 100 issues of *Polygraph*, and Editor Emeritus Ansley has been a constant help to his neophyte replacement, giving advice when asked, and otherwise quietly letting trial and error run its course.

This will be my final editorial. Beginning October 1, 2001, your new Editor-in-Chief will be Dr. Dean Pollina. I've known Dean for two years, and he was the natural choice for the job. He's an extremely bright Ph.D. who has published in the peer-reviewed scientific literature on lie detection, a full-time research scientist at DoDPI, a DoDPI-trained polygraph examiner, and one of the small number of scientists who has been converted to polygraphy by seeing it in action. He will help our discipline move into the mainstream of science, challenging the critics with his research, and challenging polygraph practitioners to continually improve themselves and the profession, to be more scientific, less dogmatic. We are fortunate to have him among us, and I hope you will give Dean all the support you gave me these four wonderful years.

Over the last four years we tried a couple new ideas in *Polygraph*, and readers will judge for themselves the wisdom of our experiments. Aside from the obvious formatting changes, we published two special issues (Scoring, and Post-Conviction Sex Offender Testing), reprinted historical pieces by polygraph pioneers, even printed complete sets of polygraph charts as part of case studies. We upgraded the terminology, replacing obsolete or incorrect expressions, to bring us closer to the language as our sister sciences. We also introduced controversy, by publishing debates on important issues, such as directed lies, inter-chart stimulation, exclusionary versus non-exclusionary probable-lie comparison questions, and those found in this very issue. Some readers have found the debates unsettling. But by those debates we are permitted to take a hard look at ourselves and our beliefs, to assess whether some of our long-cherished assumptions were correct, or if perhaps there were other ways, better ways, to perform this important role for society we've undertaken. Debates, in the proper spirit, are healthy for the profession. With my newfound spare time I plan to participate in those that spark my interest, to test and report relevant data, and I hope you will, too. It is my hope that, in some small way, these changes to our publications may have helped moved us in the correct direction.

I would like to express my thanks to the scores writers who have sent in manuscripts to *Polygraph*, making it the premiere publication of the profession. My thanks also go to the contributors to the *APA Newsletter*, and especially Johnnie Rodgerson, Vance MacLaren, and Gordon Barland. The *APA Newsletter* is the central vehicle of communication in the polygraph profession, reaching more polygraph examiners than any other. It only succeeds because it is supported by the examiner community.

We all owe a debt to our Associate Editors, who are the backbone of our publications. They quietly work behind the scenes, reviewing, and sometimes rewriting, the papers you read in *Polygraph*. As most of you know, *Polygraph* is not a peer-reviewed journal, but a technical publication, supported by our editors, that provides important information unique to our profession. Many of us see the day when, as the profession upgrades, and more practitioners are trained in scientific methods, we may introduce a new publication specifically for university-grade research articles.

Many thanks go to a patient and trusting Board of Directors, whose unwavering support I could not do without. A very special thanks to my wife, Lisa, who agreed to be an interim Managing Editor in 1999, and whose keen eye for grammar and syntax has been a blessing since.

And to all my professional colleagues, wherever you may be fighting the good fight, Godspeed.