Polygraph

VOLUME 33 2004 NUMBER 4

Contents

Published Quarterly

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

The Efficacy of Detecting Deception in Psychopaths Using a Polygraph[1]

Brett A. Stern[2] and Donald J. Krapohl[3]

Abstract

This paper addresses the efficacy of psychophysiological detection of deception (PDD), or polygraph testing, of persons who are classified under the umbrella term Antisocial Personality Disorder as defined by the Diagnostic and Statistical Manual of Mental Disorders. Specifically, the examination of polygraphing psychopaths was undertaken. The classification conundrum surrounding psychopaths is discussed, as are the ways of psychopaths, and finally the applicable research involving PDD testing of psychopaths. Coming to know the psychopath helps to understand how some of the myths surrounding the psychopath evolved, and why some have difficulty reconciling the research findings with what is believed about the psychopath. A common myth held in the law enforcement, judicial, and polygraph arenas is that the psychopath's deception is invisible to the polygraph. The relatively limited research evidence suggests otherwise.

For psychopaths, the world is a giant dispensing machine from which they obtain goodies without giving up any coins (Simon, 1996).

Robert Meyer introduces readers to a prototype psychopath in his 1992 text. His name was Charles Starkweather, and he was the inspiration behind the movies *Kalifornia* and *Natural Born Killers*. By the late 50's, the Starkweather case was the second worst case of mass murder in United States history. Starkweather began his killing spree murdering 11 people in five states (Boring, 2002), some in gruesome fashion. By June 25, 1959 the jury had heard the testimony, considered the evidence, and rendered its verdict in less than 24 hours of deliberating; Strarkweather was to die, and the electric chair was the last place he would ever sit (Bardsley, 2002).

Meyer (1992) writes: "Starkweather loved nature but loathed humans." As one defense psychiatrist said, "He is unable to experience feelings that other people do. People don't mean anything to him. They are no more than a stick or piece of wood to this boy. . . . The act of killing meant no more to him than stepping on a bug." Another defense psychiatrist said, "The thoughts and the feelings are not there like they are in the ordinary person, who has learned by being around others and has feelings for them, and in relation to them. . . I don't think he has ever learned to be a person."

[1]This article is one in a series under the heading Polygraph Myths. The opinions expressed in this article are exclusively those of the authors, and do not necessarily represent those of the Department of Defense or the U.S. Government.

[2]Brett A. Stern is a Senior Instructor at the Department of Defense Polygraph Institute. Requests for reprints should be directed to Brett A. Stern, 7540 Pickens Avenue, Fort Jackson, South Carolina 29207.

[3]Donald J. Krapohl is Deputy Director of the Department of Defense Polygraph Institute.

The research findings regarding polygraph testing of the psychopath, which will be taken up later in this article, are made far more intriguing when one has insight into the composition of the psychopath. Coming to know the psychopath helps us to understand how the myth developed that a psychopath's prevarications are impervious to discovery with a polygraph. Moreover, it also helps us to better understand why some people remain unconvinced by the research findings.

The Classification Conundrum

The first issue we must contend with toward understanding the psychopath is one of proper classification. According to David Lykken, (1955), "Classification in medicine (broadly defined) goes under the name of diagnosis. The rules of medical classification--the dimensions of similarity to be utilized in defining a diagnostic category--are characteristically inconsistent." Lykken suggests ". . . this inconsistency may be traced to the purposes of diagnosis--those disorders are to be classified together which the clinician is to treat in the same way." Lykken tells us that "The history of the concept of psychopathic personality is one long chronicle of attempts at . . . classification most of which, to date (1955), have proved abortive."

According to Meyer (1992), "The term Antisocial Personality Disorder (APD), (the term in vogue today, of which the psychopath is a subgroup), is the result of an evolution through a number of terms, the most widely known of which is undoubtedly psychopath." Meyer states " . . . there is considerable overlap among the terms APD . . . psychopath, and sociopath." Johann Koch introduced the label psychopathic inferiority in the late nineteenth century and that term became accepted for a while. Complicating matters even more, Meyer claims that ". . . many experts feel that there is reasonable evidence to further subdivide the APD, such as into 'primary' and 'secondary' psychopaths." Citing Loeber (1990), irrespective of the category you put the psychopath in, Meyer writes: they ". . . are different from individuals who are antisocial because they grew up in and adapted to a delinquent subculture. The non-psychopathic antisocial personality are conformists in that they follow the rules and mores of their subculture."

In 1976, Hervey Cleckley wrote his classic book on psychopathy, The Mask of Sanity. Commenting on the classification dilemma Cleckley said, "The term psychopath (or antisocial personality disorder) as it is applied by various psychiatrists and hospital staffs sometimes become so broad that it might be applied to almost any criminal."

Robert Hare (1993) and Karl Menninger (1942) also have addressed the problem of imprecise classification and actual mislabeling of the psychopath with Menninger going so far as to advocate a new official name for the psychopath.

According to the DSM

Even the American Psychiatric Association (APA), in its *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV, 1994), has grappled with the classification difficulty of psychopaths. According to Hammond (1980), the second edition of the APA's DSM, which was released in 1968, attempted to coordinate its classification system with that of the World Health Organization and adopted the category 301.7, antisocial personality. According to Robert Simon (1996), ". . . the term psychopath was used originally in psychiatry to refer to all personality disorders (e.g., paranoid, schizoid, anti-social, borderline, histrionic, narcissistic, avoidant, dependent, and obsessive-compulsive)." Based on Cleckley's work with psychopaths, APD was the first personality disorder recognized in psychiatry and was included in the first edition of the DSM. In 1968, the term sociopath or sociopathic personality replaced the term psychopath to emphasize the environmental factors that allegedly generated the disorder (Simon, 1996). The 1994 edition of the DSM holds the psychopath under the rather imprecise umbrella term APD but focuses more on antisocial behavior over personality traits and their motivation in the definition of APD. Interestingly, and understandable from the standpoint of the consequences of prematurely branding or labeling someone psychopathic, if the same traits and behavioral characteristics of the psychopath were found in a person

under age 18 that person would be deemed to have a conduct disorder.

Few people have devoted as much of their lives to understanding and writing about the psychopath as Robert Hare. From his extensive work with psychopaths, Hare developed a Psychopathy Checklist that is perhaps the most important assessment tool available today to clinicians, researchers, and a wide range of other people involved in assessing and dealing with psychopaths. His other more notable works include the 1993 book *Without Conscience*: *The Disturbing World of the Psychopath Among Us*. We will rely heavily upon Hare's work in helping to understand psychopaths and to answer specifically whether detecting deception in the psychopath is any more of an elusive undertaking than it is in detecting deception in "normals."

Around 1800, Philippe Pinel coined the term *manie sans delire* (i.e., insanity without delirium) (Hare, 992; Meyer, 1992) for persons who manifest extremely deviant behavior but show no evidence of delusions, hallucinations, or other cognitive disorders (Cleckley, 1976; Meyer). Similarly, Hammond, citing Fotheringham (1957), writes "The disorder is an illness without evidence of mental deficiency, structural disease of the brain, epilepsy, psychosis, psychoneurosis or intellectual impairment, and that it is primarily a disorder of behavior rather than thinking."

Hare (1993) tells us that the classification conundrum and confusion about psychopaths stems from the word itself, and also attributes misperceptions regarding the psychopath to the media who improperly label psychopaths as "crazy," or "insane" or by the more often melodramatic term "psycho." According to Hare, these terms are misnomers because the psychopath is " . . . not disoriented or out of touch with reality, nor do they experience the delusions, hallucination, or intense subjective distress that characterize most mental disorders." Hare also tells us that psychopaths are not insane in the psychological or legal sense. Hare sums up the distinction between society's response to the typical psychopath and its response to the person suffering a mental disorder of schizophrenia, for example, where they may experience auditory hallucinations directing them to kill someone. The schizophrenic person is deemed not responsible for his or her actions "by reason of insanity" and is given mental health treatment. The psychopath is judged by society as sane, and is sent to prison to receive little or no treatment.

A Case for Emulating the Psychopath

While most people knowledgeable of psychopathy, and the ways of psychopaths, view them as menacing, socially inept figures who prey on the rest of us, others have the boldness to suggest society might be best served by envying them. This notion certainly clarifies nothing about the psychopath but rather advances the confusion of classification and how to explain him. Author Alan Harrington (1971) would have us ponder whether the psychopath is to be reviled or revered. Harrington suggests the psychopath may be worthy of emulation. Consider, for example, the following from his works as cited by Cleckley (1976): "Have we come to the hour of the psychopath, the advent of psychopathic man . . . {when} what was once presumed to be a state of illness is abruptly declared to be a state of health." He continues, " . . can it be true that with the dramatic appearance of the psychopathic ideal, a new man has come upon us, that in order to survive the turbulent years ahead, far from seeking to treat the psychopath in clinics, we should rather emulate him, learn how to become him?" In response to Harrington, and those of the counterculture movement of the time, Cleckley writes, "A sincere choice of the real psychopath as model or leader by anyone familiar with the subject would be beyond absurdity."

What Does the Word Psychopath Mean?

While classification may continue to prove elusive, the attributes of the psychopaths are fairly well settled. What is a psychopath? First, let us dissect the word itself. The first part of the word "psycho-" comes from the Greek word *psyche* meaning soul, spirit, or mind (Becker, 1989). The second part of the word "path" originates from

the Greek word *pathos* (i.e., from *paschein* to undergo, be affected) meaning an incident, experience, sensation, emotion, mishap: trouble, and suffering. In combined form, pathos translates to disease or pathologic (Becker, 1989). In psychological parlance, when both words (i.e., psycho & path) are used in combination it connotes mental illness (Hare, 1993).

Antisocial Personality Disorder and Psychopathy Defined

Churchill's Medical Dictionary (Becker, 1989) defines a psychopath as someone who manifests characteristics of the antisocial personality. A sexual psychopath, for example, is a person whose manifestations of an APD are predominantly in the sexual area (Becker). "The psychopath can have lustful sex, but for them the experience is devoid of any intimacy or commitment; the partner is essentially an instrument of masturbation" (Simon, 1996). The term "antisocial" refers to the characteristic of avoidance of interpersonal relationships. It is also reflective of behavior that violates the laws, rules, or moral or ethical code of one's culture (Becker).

Generally, a personality disorder is characterized as an enduring pattern of inner experience and behavior that deviates markedly from the expectation of the individual's culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment (DSM IV, 1994).

"Antisocial Personality Disorder (often referred to as psychopathy, sociopathy, or dissocial personality disorder) is a particular type of personality disorder the essential feature of which consists of a pervasive pattern of disregard for, and violation of others that begins in childhood or early adolescence and continues into adulthood" (DSM IV, 1994).

Pathological Lying

While the term pathological liar appears in medical dictionaries and the research literature (Davidoff, 1942; Deutsch, 1982; Hare, 1989) it is not a specifically referenced mental disease recognized within the DSM IV. However, lying, deception, and manipulation of others is central to individuals diagnosed with APD. Although by supposition, the term pathological liar may have been borne out of the fact that psychopaths engage in mendacious behavior that is, according to Hare, ". . .habitual and blatant and do so with considerably more panache, than do most people" (Hare, Forth, & Hart, 1989). While all of us lie and deceive it is the extent to which the psychopath deviates from societal norms, regarding deception and lying, that gives rise to the notion that they are diseased--hence pathological.

Is There Such a Thing as a Lying Disease?

One could argue there is no such thing as a "pathological liar," for the term pathological denotes an abnormal finding, particularly a morphological (an organism's structure and form, excluding its functions) alteration resulting from disease (Becker, 1989). Therefore, it is questionable that a person possesses a lying disease, per se. "There is some evidence that psychopaths differ from normal people in the processing, use, and cerebral organization of language….The psychopath's words and actions often appear to reflect some sort of affective deficit (Hare, Forth, & Hart, 1989)."As Hare, Forth and Hart report, consider the following:

> On language of the psychopath, Cleckley writes, "He can learn to use ordinary words... (and) will also learn to reproduce appropriately all the pantomime of feeling . . . but the feeling itself does not come to pass." Grant (1977) writes, "Ideas of mutuality of sharing and understanding are beyond his understanding in an emotional sense; he knows only the book meaning of words." Johnson (1946), states, "(He) exhibits a facility with words that mean little to him, form without substance. . .His seemingly good judgment and social sense are only word deep." Some researchers believe there may be a unique organic component

> (brainwave disorder) found in psychopaths (Doren, 1987). Heredity, brain dysfunction, individual developmental experiences, and subcultural conformity are all promoted as generic to the antisocial personality (Meyer, 1992).

Reid (1978), citing Eissler and Aichom, stresses the importance of early mother-child relationships and later oedipal identifications as influencing psychopathy. Greenacre, according to Hammond (citing Cleckley), "... concluded that the confusing influence of a stern, authoritative father and an indulgent or frivolous mother is common in the early background of the psychopath."

Kegan (1986) attributes a developmental delay in psychopaths to help explain the psychcopath' s mendacity. The psychopath's "... manipulation can be understood as a developmental delay in which his cognitive, affective, and interpersonal processes are like that of a normal child around ten years old." Kegan adds, ". . . there is greater concern for one's own needs than with the needs of others. These needs are satisfied by manipulating and controlling the behavior of others."

Another explanation for the psychopath's reliance on lying and deception might be behavioral rather than biological. Lying is a learned behavior and as that behavior proves fruitful it becomes reinforced. Consequently, the liar continues to lie and manipulate people with greater frequency to the extent that lying neither carries the emotional baggage most of us experience when we tell a lie nor the stigma that society attaches to it. Not surprisingly, the habitual liar learns to embrace lying as way of life--for successful lying brings with it the rewards that prompted the lie in the first place. Moreover, as the habitual liar continues his manipulative behavior he becomes progressively desensitized to lying. If the liar becomes desensitized, it is argued, that he would have a correspondingly diminished autonomic nervous system response when telling a lie and therefore more difficult to detect through use of the polygraph. We suspect, in part, it is this intuitive perspective that may have given rise to the myth held by many people that the pathological or habitual liar cannot be satisfactorily tested through use of the polygraph

On the other hand Eugene Davidoff (1942) writing on *The Treatment of Pathological Liars* stated: "Except in the very young children, pathological lying rarely appears as an isolated phenomenon. It is in general a function of the integration of the child's personality. As such, it is frequently found as a sub-group of other neurotic (personality) or psychopathic (conduct) disorders. . . ." To understand pathological lying it is helpful to address it in terms of classification (i.e., normal/abnormal) and severity (i.e., mild/severe). Davidoff chooses to classify liars according to their prognosis and response to therapy. None of us would dispute that we lie. When we lie we generally do so occasionally, it is situation driven, it has some pseudo-constructive purpose, and is a byproduct of conscious thought. We have an insight into why we are lying. We are not psychotic (i.e., insane or suffering from severe mental illness). Davidoff would likely classify us "normal" possessing at worst a mild form of pathology that is responsive to therapy. At the other end of the continuum is the abnormal or pathologic liar. They lie continuously, are compulsive, and their lies are often destructive to themselves and others. Their lie originates from fantasy, not reality. They often manifest characteristics of paranoia and psychosis. Davidoff believes they have little or poor insight into their lying, their pathology is severe and they, as a general rule, do not respond well to therapy.

Psychologists may also have inflamed the belief about the suitability of psychopaths for detection of deception testing given the psychopath's alleged propensity to show diminished responsivity during electrodermal trials (Ansley, n.d.): According to Ansley, one such trial was conducted by Lykken (1955) wherein he administered a peak-of-tension test (numbers test) to psychopaths and non-psychopaths and found, using only a galvanic skin resistance measure, non- psychopaths displayed greater response differentiation between the number lied about than those they were truthful about than did psychopaths.

Aren't We All a Little Psychopathic?

Dr. Simon states, "Everyone has antisocial impulses and the vast majority of us would reflect so on various personality tests designed to identify psychopathy. The good news is that the extent of our psychopathy doesn't trespass over the line of demarcation where we would be classified as possessing a psychopathic personality. On Hare's Psychopathy Checklist (PCL) (PCL-1 appeared in 1980; PCL-2 appeared in 1985) non-psychopaths might score a five out of a maximum score of 40 points. The psychopath might score anywhere above 30, for example. Having committed a criminal act does not make one a psychopath nor are all psychopaths criminals.

"Psychopaths exist in all levels of society, in all walks of life. No profession, however noble, is spared their cadre of them. We know them, if we know them at all, by their acts (Simon, 1996)." The criminal non-psychopath typically has standards or boundaries within which he operates. If he kills during the commission of a crime it is viewed as the inherent cost of doing business (Simon, 1996). However, he regrets doing so and will often reflect upon his act. To the psychopath, he could care less that he had to kill you. After all, it's your fault for you shouldn't have been there in the first place.

The psychopath, according to Hare (1996), is a natural predator. While prisons are filled with clinically diagnosed psychopaths, they may exist in greater numbers among the general population, and may be cloaked in such benign titles as grandfather, mother, father, sister, brother, teacher, supervisor, boss, and the like. Hare estimates there may be as many as three million psychopaths in North America.

Psychopath or Entrepreneur?

Person (1986) gives us an interesting and good sense of what differentiates the psychopath from the successful businessman, both of whom engage in " . . . risk-taking and manipulative behaviors . . . to control events and people and they likewise receive . . . tangible and psychological rewards for doing so. The distinction is that the businessman's manipulation " . . . is more rational and goal oriented than is the psychopath's." While the businessman may be "ruthless in his business dealings" he is capable of ". . . developing warm affectionate bonds with others" while the psychopath cannot (Person). Moreover, the entrepreneur's manipulation is more geared toward attainment of wealth, prestige, and power whereas the psychopath uses manipulation as a means of ". . . dominating and humiliating others." About the psychopath, Cleckley (1976), writes: "There is nothing odd or queer about him, and in every respect he tends to embody the concept of a well-adjusted, happy person.... He looks like the real thing.... More than the average person, he is likely to seem free from minor distortions, peculiarities, and awkwardness so common even among the successful."

Is The Psychopath Responsive to Treatment?

Cleckley (1976) and Hare (1996), question the efficacy of rehabilitating the psychopath, for psychopaths do not see themselves as possessing a mental disorder. To rehabilitate psychopaths it would be necessary to alter their behavior, and their perception of their behavior. Hare believes many treatment programs only provide a breeding ground for the psychopath. They learn the appropriate psychological vernacular, they learn what makes people tick, and they use this newly acquired knowledge to advance their exploitive behavior. What they believe is right for them, irrespective of what society believes. Psychopaths operate according to their own rules and pick and choose which rules to violate and when to violate them. They view people as objects--either roadblocks or gateways to their desires. They do not internalize society's norms or rules (Hare). Contemptuous of the feelings, rights, and sufferings of others; impulsivity; lack of empathy; remorselessness; callus, cynical, inflated and arrogant self-appraisal; glib; superficial charm, self-assured and exploitive; and lack of individual concern are just some of the associated features characteristic of persons with APD (DSM IV, 1994). It remains unsettled whether the psychopath is responsive to treatment. The question that must first be answered definitively is whether psychopathy is either

an all-or-nothing proposition or are there different classifications of psychopathy, and where is the line of demarcation separating those classifications?

Are They Really That Smart?

Nowhere in this article have we said that the psychopath is stupid or unable to function in every day life. Writing about the psychopath, Cleckley (1976) states, "Psychometric tests also very frequently show him (the psychopath) of superior intelligence." Harrington (1971) writes that there are "Brilliant individuals among us that are basing their own lives on the psychopathic model." Meyer (1992) has taken issue with Cleckley on this point, even asserting that Cleckley's findings only applied to a small subset of patients within his private clinic. Meyer writes, "As a whole antisocial personalities show lower-than-average scores on intelligence tests." Notwithstanding the intellect controversy, many psychopaths function quite well and may rise to enjoy a professional status many of us can only aspire to achieve, and yet never cross the line into criminality.

The Psychopath's Achilles Heel?

After researching the psychopath, it is understandable and reasonable how one could hold the opinion that the psychopath would be an unsuitable candidate for polygraph testing. If the psychopath lies with effortless skill, is supposedly indifferent to having his lies detected, is purportedly electrodermally hyporeactive, is regarded as a master manipulator of people, internalizes no guilt about his or her acts no matter how heinous we might view them, how then could their body betray their tongue during the course of a polygraph examination? This question currently lacks a definitive answer.

Psychopaths are in charge of their faculties, know what they are doing, and why they are doing it. They simply are unaffected about the impact of their actions on others. While they may not harbor any concern about their criminal acts, the psychopath is highly motivated and doesn't want to get caught. Their motivation in a polygraph setting is affected by the challenge of attempting to control physiological responses during deception. They find themselves in a unique environment with a difficult task of controlling the decision outcome. As Raskin and Hare (1978) and Hammond (1980) reported, the psychopath is essentially in competition with an inanimate object and is, therefore, perhaps more challenged than when placed in a face-to-face encounter with a person or people who they have made a habit of duping. They care if something affects them immediately, according to Hare (1996). The fact psychopaths are highly motivated, challenged, find themselves in a novel environment, care about being caught, and will attend to things that have an impact on their immediate well-being may be their Achilles' heel, affording the polygraph examiner an opportunity to exploit them.

What Does the Research Say?

The psychophysiological detection of deception research into the susceptibility of psychopaths to polygraph testing is limited; however, the results are consistent. We will address the following research in the remainder of this article: Raskin, Barland, and Podlesny (1977), Raskin and Hare (1978), Hammond (1980), and Patrick and Iacono (1989).

Raskin, Barland, and Podlesny (1977)

In 1977, Raskin, Bar1and, and Podlesny completed a project concerning the validity and reliability of polygraph techniques in the detection of truth and deception with criminal suspects. They also conducted laboratory experiments that addressed general problems of accuracy and reliability that could not be easily studied in field situations. Of the eight studies and experiments conducted, one addressed the issue as to whether psychopaths can "beat" a polygraph.

Of 24 subjects classified as psychopathic, decision accuracy was 96%. There was only one misclassification decision and that was a false positive. There were no inconclusives and not one guilty psychopath was able to produce a false negative. Of the 24 subjects comprising the non-psychopathic group, there were 19 of 24 (79%) correct

decisions, one misclassification, and four inconclusives.

Raskin and Hare (1978)

Raskin and Hare (1978) set out to answer the question, how effective is the "control question" (now known as comparison question) test in detecting deception in psychopaths when standard measures of respiration, electrodermal, and cardiovascular activity are used. They published the results of their study in an article *Psychopathy and Detection of Deception in a Prison Population* (1978). Forty-eight subjects, half of whom were diagnosed as psychopathic, were obtained from the inmate population in Burnaby, British Columbia. The subjects were instructed to enter an otherwise off-limits room when no one would see them and steal $20.00 from an envelope and secrete it on their person. They were subsequently escorted to another room where they were subjected to polygraphic examination. If a guilty subject could produce a false negative (i.e., a guilty person adjudged as innocent) and the criterion-innocent subjects a true negative (i.e., an innocent person adjudged as innocent) they would each receive a $20.00 reward. A $20.00 reward, at that time, represented the equivalent of about 27 days pay for prison labor. Raskin crafted the questions, administered the test, evaluated the data, and ultimately rendered a diagnostic opinion. Raskin's evaluation of the data was done while blind to the programming status of subjects. Between-chart comments were used (e.g., "Do any of the questions bother you?" and "Would you like to change the wording of any question?") and directed all subjects' attention toward the comparison question, a practice that has remained highly controversial, and not widely adopted. If a subject expressed sensitivity to a comparison question the question of focus was changed; however, no relevant question was changed irrespective of whether a subject voiced concern to a relevant question.

When using the full complement of charts (anywhere between three and seven) 88% of decisions were correctly categorized, 4% were incorrect, and 8% were inconclusive. An overall accuracy of 96% was reported, excluding inconclusive opinions. Raskin and Hare concluded that there was no significant difference in accuracy rates for psychopaths and non-psychopaths. In other words, psychopaths were as easily detected as non-psychopaths. It was also noted that the psychopaths showed stronger electrodermal responses and heart rate decelerations.

Criticisms

In his paper, *The Psychopath and the Lie Detector* David T. Lykken (1978) analyzed Raskin and Hare's 1978 study. Lykken states, "It is my opinion that all of these important implications, claims, and conclusions are unsupported by the evidence at hand, and may have adverse and serious social consequences." Lykken maintained that the experiment did not definitively demonstrate that deception employed by psychopaths was any more or less easily detected than in non-psychopaths with the polygraph. Lykken argued that psychopaths and non-psychopaths alike should have little difficulty in thwarting the lie detector because all they need to do is augment their responses to selected questions. Virtually anyone can be taught to recognize comparison questions, argues Lykken. Lykken maintains the reason why psychopaths are believed to be able to defeat a polygraph test is because some hold that "psychopaths are habitual or practiced liars and seem to feel relatively little guilt or fear about these actions (referring to the mock crime paradigm) or their consequences." Lykken holds that innocent or guilty, a normal subject will experience apprehension about the relevant question and will thus fail the comparison question test. According to Lykken, "Because he is less disposed toward anxious apprehension, the psychopath might be expected to respond relatively less to the critical questions whether he is innocent or guilty of the real criminal act of which he is suspected." Moreover, the psychopath's responses to comparison questions would also ". . . be relatively more attenuated so that the most plausible expectation might be that the psychopath would produce relatively more "inconclusive" outcomes and fewer "deceptive verdicts than would a normal subject." While responses, in general, may arguably be more attenuated it is doubtful that such attenuation would be selective to one type question to the exclusion of another. "The only reason for

expecting the psychopath to be better able to avoid failing the lie test, even though deceptive, is that he might be expected to be less frightened or guilty. . . than will be the normal subject," says Lykken. With this in mind, Lykken maintains that Raskin and Hare's mock crime experiment ". . . would not have anything to do with genuine fear or guilt." Of Raskin and Hare's experiment Lykken said, " . . . I cannot imagine that I would have found the experience frightening or guilt-provoking in any way. On the contrary, I should think it would have seemed like an interesting game in which I stood a chance of winning a $20.00 prize plus the admiration of my colleagues. . . ." What Lykken is touching upon, with respect to gaining ". . . the admiration of my colleagues," is the reference to "duping delight" (Ekman, 1992). In addressing duping delight, Paul Ekman writes, "The liar may feel excitement, either when anticipating the challenge or during the very moment of lying, when success is not yet certain. Afterward there may be the pleasure that comes with relief, pride in achievement, of feelings of smug contempt toward the target." Lykken goes on to say, "What is different about the psychopath is his attenuated capacity for fearful or guilty apprehension; no psychopath of my acquaintance is deficient in his interest in games, in opportunities to 'show off,' or in winning money prizes."

In addressing responses to relevant questions posed in Raskin and Hare's experiment, they should not have produced ". . . the kind of fear or apprehension that the lie test elicits in real life." This is the external validity argument (i.e. generalizability of laboratory test results to a real-world situation) that polygraph laboratory studies often suffer. Thus, according to Lykken, responses to the relevant question should have been interpreted merely as orienting responses that the psychopath displays as frequently as non-psychopaths.

The stronger electrodermal responses noted in this study are of particular interest, because previous work by Lykken (1955) showed that psychopaths were electrodermally hyporeactive (i.e., less reactive). This lack of electrodermal responsiveness, coupled with the associated features of persons diagnosed with APD, gave rise, within the scientific community, to the hypothesis that psychopaths should be able to defeat the polygraph examination process. The Raskin and Hare study arguably demonstrated otherwise.

Hammond (1980)

Hammond's (1980) dissertation involved polygraph laboratory research into the responding of normals, alcoholics, and psychopaths. The purpose of Hammond's research was to test the hypothesis of atypical responding by alcoholics and psychopaths as compared with normals who undergo a polygraph experiment. Psychopaths were examined because of purported deficits in the area of conscience development that could theoretically make them more difficult to detect when lying. Hammond was also interested in substantiating claims made by previous investigators who had studied psychopaths and found them to be "…adequate responders and therefore amenable to the polygraph test."

Sixty-two subjects participated in Hammond's study and were placed in one of three groups (i.e., normals (21), alcoholics (20), and psychopaths (21)). The mock crime paradigm involved the theft of $10.00 from a pair of coveralls hanging in a closet of a room. Thirty-two subjects were guilty of stealing the money and 30 subjects were innocent. All subjects were promised $7.00 for participation and were told of the possibility of earning a $10.00 bonus for producing a false negative result. A probable-lie comparison question test was administered using the Zone Comparison Test format. Subjects met the following profile: white males, ages 21 through 55, prison record, moderate to low income, and living in the greater San Diego, CA area. Polygraph examiners were in week five and six of a seven-week training program. Hammond's study capitalized on shortcomings he believed plagued other studies (e. g. inadequate control groups; use of incarcerated subjects; questionable participant motivation; participant personality group disclosure).

No significant differences were found regarding the suitability of the three categories of subjects who underwent a control-question polygraph examination. All were detected

above chance level. Hammond reported an accuracy rate of 93% when inconclusives (n=35) were eliminated from the computation, with 7% error. The inconclusive range was set at +/-8, which may have created a larger proportion of inconclusive decisions that would the more orthodox cutting scores of +/-6. Hammond concluded, "The data provides no support to the myth that psychopaths can lie without producing some psychological, emotional, or physiological concomitants that are associated with deception."

Patrick and Iacono (1989)

With the challenges and controversy that Lykken surfaced surrounding the Raskin and Hare study it was inevitable that additional research into the efficacy of detecting deception in the psychopath, though use of a polygraph, would need to be undertaken. Eleven years later it was.

Patrick and Iacono (1989) responded to the challenge and embarked on a study, with "procedural refinements" to reevaluate the accuracy of the comparison question test with psychopathic individuals. The study utilized a mock crime scenario involving 48 inmates from the same British Columbia prison as in the Raskin and Hare study. Half of the subjects were classified as psychopathic. Psychopathy was determined by using the Psychopathy Checklist developed by Hare. The Checklist is an " . . . instrument whose psychometric properties and relevance to Cleckley's conception of psychopathy are well-established." Other measures of evaluation were also employed; lengthy structured interview; and case record review by two trained diagnosticians, and only where consensus existed as to inmate classification were inmates selected for study. A total of 107 inmates were screened before the final sample of 48 was secured.

The scenario involved the theft of $20 from a doctor's jacket coat pocket that was in a room normally off-limits to prisoners. One of the keys of this study is that the scenario was set up to capitalize principally on conditions of threat (i.e., failure and the consequences) versus reward. There was an initial monetary incentive for inmates to undergo initial screening (i.e., $2.00) and an additional $10.00 bonus if they were ultimately selected as a participant. All subjects were instructed that the study was designed to see if they had the ability to beat the polygraph test when there was something important at stake. It is important to understand that no "individual incentive" was offered per se. However, each individual group member (48 in all) stood to gain an additional $20.00 bonus if no more than 10 of the 48 inmates were classified as deceptive. The experimenter stressed to each subject how important their individual performance was to the group and that his peers were counting on them. If more than 10 inmates were classified as deceptive, the inmate participants were told they would lose the bonus and the name of the participants responsible would be made known to the prison population--who presumably would deal with those responsible in the customary prison manner. Although unknown to the study participants at the time, each subject would ultimately receive the $20.00 bonus irrespective of their test results.

Results and Conclusions of the Patrick and Iacono Study

Excluding the inconclusive rate, the overall hit rate for both groups was 87%. Only 2 of 12 guilty psychopaths and only 1 of 12 innocent non-psychopaths were misclassified (false negative v. false positive rate). With respect to the innocent subjects, group differences in accuracy were reported as non-significant (the actual data were not provided). Excluding the inconclusive rate the overall hit rate for innocent subjects was only 56%.

To test for the unknown influence of extra-polygraphic cues or contamination in the overall accuracy decision, Patrick and Iacono (1989) had the examiner, just prior to going into data collection, register an opinion on a weighted scale as to the examinee's guilt or innocence. The study found that pretest guilt judgments did not exceed chance.

The results of this study supported Raskin and Hare's (1978) study where psychopaths were no more likely to defeat a comparison question polygraph test than were non-psychopaths. This is still the case with blind numerical evaluations by an independent evaluator. Moreover, the inter-

rater agreement between the examiner of record and the independent evaluator was 87% permitting only a limited biasing influence from extra-polygraphic information. Some may challenge the studies we have presented in this article from the standpoint that they may not accurately reflect how psychopaths truly perform in real-life polygraph examinations (Hare, Forth, & Hart, 1989), however, this is the same argument that could be made for virtually any laboratory study involving the detection of deception.

Conclusion

You were introduced to a psychopath--Charles Starkwheather. The evolutionary and continuing dilemma of proper classification was addressed. Yet one cannot help being left with the impression that as more is learned about the psychopath other terms shall find their way on the ever-expanding list of terms. The origin and definition of psychopathy, as it is known today, was explored. However, a new more precise definition is surely to come. The psychopath is better understood today, because of the work of Dr. Hare and others who have given us an insight into the personality traits and behavioral characteristics of psychopaths. As predatory as psychopaths are there are some who incredibly advocate that they be emulated. Thankfully, those who know psychopathy best find this suggestion perverse. Whether psychopathy has a biological, behavioral, or other component we know that we must contend with the millions of Americans who fit the definition of psychopath yet may never cross the line into criminality. Society cannot afford to remain indifferent to the psychopath's manipulative and predatory ways for in its grossest form they can wreak devastation on our way of life. Finally, the limited but rather convincing research on the efficacy of detecting deception in psychopaths, whose deception is purportedly impervious to detection, was reviewed.

In the absence of other evidence, one may conclude that the collective and consistent findings in the Raskin, Barland, and Podlesny (1977), Raskin and Hare (1978), Hammond (1980), and Patrick and Iacono (1989) studies debunk the myth that the psychopath's deception is impervious to discovery through the use of a polygraph. Moreover, when the psychopath engages in deception, his or her deception is no more difficult to detect than deception practiced by non-psychopaths. The question is why.

Earlier it was mentioned that the psychopath's Achilles heel may stem, in part, from the fact that they, like the rest of us, don't want to have their prevarications found out--particularly when the consequence of disclosure might impact them legally. We learned that psychopaths care about things that affect them immediately. They are motivated to pass their polygraph test, if for no other reason than to simply dupe the examiner. They are challenged not only by the opposition, who they view as merely a roadblock, but by an inanimate object with which they likely have had little or no exposure and likely have never defeated in battle. The psychopath finds himself in a unique setting that is highly controlled--by others.

The fact that the psychopath's deception, in the studies that were reviewed, was no more difficult to ferret out than the non-psychopath suggests the possibility that they might, in fact, become as emotionally aroused by relevant questions as non-psychopathic people do when engaged in deception. It is well established that guilt, one of the emotion-based theories, is not a necessary precondition for polygraph detection efficacy. It is also known that cognitive-based theories, such as cognitive awareness, offer a plausible explanation for why polygraph subjects respond to critical items in Concealed Information Tests, Peak-of-Tension, and other tests. Finally, in explaining the rationale behind the research results involving detection of deception efficacy of psychopaths Hare, Forth, and Hart (1989), write " . . . it is more a reflection of perceptual-cognitive demands than of fear or anxiety. That is, a psychopath who is not at all anxious or fearful during the examination may nevertheless respond physiologically to the critical questions because he sees the examination as a game or challenge and because he is highly motivated to beat the test."

Acknowlegdements

The authors wish to acknowledge the support of Mrs. Genni Arledge and Ms. Stacey Davis in obtaining reference materials and providing other assistance toward the completion of this paper.

References

Ansley, N. (n.d.) *Psychopaths as polygraph subjects*. (Available from the Department of Defense Polygraph Institute, 7540 Pickens Avenue, Fort Jackson, SC 29207).

Bardsley, M. (n.d.). Charles Starkweather & Caril Fugate. Available from http://www.crime library.com/starkweather/starkmain.html.

Becker, L. et al. (1989). *Churchill's illustrated medical dictionary*. NY: Churchill Livingstone.

Boring, E. (n.d.). *The serial killer*. Available from http://members.home.net/supersatan/ killers/charlesstarkweather.html.

Cleckley, H. (1976). *The mask of sanity*. Saint Louis, MO: The C. V. Mosby Company.

Davidoff, E. (1942). The treatment of pathological liars. *Nervous Child, 1*, 358-388.

Deutsch, H. (1982). On the pathological lie (Pseudologic Phantastica). *The Journal of the American Academy of Psychoanalysis, 10*(3), 369-386.

Doren, D. (1987). Understanding and treating the psychopath. In R.G. Meyer (Ed), *Abnormal behavior and the criminal justice system*. NY: Lexington Books.

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

Fotheringham, J. B. (1957). Psychopathic personality: A review. In D. L. Hammond (1980) The responding of normals, alcoholics, and psychopaths in a laboratory lie detection experiment. *Dissertation Abstracts International*, *41 (6-B)*, 2374 (UMI No. AAD8-28678).

Grant, V. (1977). The menacing stranger. In R. D. Hare, A. E. Forth & S. D. Hart (Eds), T*he psychopath as prototype for pathological lying and deception*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Hammond, D. L. (1980). The responding of normals, alcoholics, and psychopaths in a laboratory lie detection experiment. *Dissertation Abstracts International, 41 (6-B)*, 2374 (UMI No. AAD8-28678).

Hare, R. D. (1993). *Without conscience: The disturbing world of the psychopaths among us.* NY: Pocket Books.

Hare, R. D. (1996). *The psychopathic mind*. {Videotape}. {Interview with Dr. Robert D. Hare} Available from Films for the Humanities and Sciences. The National Canadian Broadcasting Company News. Princeton, NJ.

Hare, R. D., Forth, A. E., & Hart, S. D. (1989). The psychopath as prototype for pathological lying and deception. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 25-49). Dordrecht, Netherlands: Kluwer Academic Publishers.

Harrington, A. (1971, December). The coming of the psychopath. *Playboy*, 200-204.

Johnson, W. (1946). People in quandaries: The semantics of personal adjustment. In R. D. Hare, A. E. Forth, & S. D. Hart (Eds.), *The psychopath as prototype for pathological lying and deception.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Kegan, R. G. (1986). *The child behind the mask: Sociopathy as developmental delay*. In R. D. Hare, A. E. Forth, & S. D. Hart (Eds.), *The psychopath as prototype for pathological lying and deception.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Loeber, R. (1990). Development and risk factors of juvenile antisocial behavior and delinquency. *Clinical Psychology Review, 10,* 1-42.

Lykken, D. C. (1955). A study of anxiety in the sociopathic personality. *Dissertation Abstracts*, *16(4)*, 795. (UMI No. AAD 00-15944).

Lykken, D. T. (1978). The psychopath and the lie detector. *Psychophysiology, 15*(2), 137-142.

Meyer, R. G. (1992). *Abnormal behavior and the criminal justice system*. NY: Lexington Books.

Menninger, K. (1942). Recognizing and renaming "psychopathic personalities." In H. Cleckley (Ed.), *The mask of sanity.* The C. V. Mosby Company.

No Author (1994). *Diagnostic and Statistical Manual of Mental Disorders*, (4th ed.). Washington, DC: American Psychiatric Association.

No Author (1968). *Diagnostic and Statistical Manual of Mental Disorders*, (2nd ed.). Washington, DC: American Psychiatric Association.

Patrick, C. J., & Iacono, W. G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology, 74*(2) 347-355.

Patrick, C. J., & Iacono, W. G. (1986). *The validity of lie detection with criminal psychopaths.* Paper presented at the meeting of the Society for Psychophysiological Research, Montreal, Canada, October 1986.

Person, E. S. (1986). Manipulativeness in entrepreneurs and psychopaths. In R. Hare, A. Forth, & S. Hart (Eds.), *The psychopath as prototype for pathological lying and deception*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Raskin, D. C., Barland, G. H., & Podlesny, I. A. (1977). Validity and reliability of detection of deception. *Polygraph*, *6*(1), 1-39.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, *15*(2), 126-135

Reid, W. H. (1978). The psychopath: A comprehensive study of antisocial disorders and behavior. In D.L. Hammond, (1980). The responding of normals, alcoholics, and psychopaths in a laboratory lie detection experiment. *Dissertation Abstracts International, 41 (6-B)*, 2374 (UMI No. AAD8-28678).

Simon, R. I. (1996). *Bad men do what good men dream*. Washington, DC: American Psychiatric Press Inc.

A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute[1]

Stuart M. Senter[2], Andrew B. Dollins, and Donald J. Krapohl

Abstract

The accuracy for specific issue laboratory polygraph studies based at the University of Utah is approximately 12% higher than for similar studies conducted at the Department of Defense Polygraph Institute. This project was completed to determine possible sources for these differences. Four scorers, two trained at each institution, assigned values to physiological responses from polygraph examinations of 50 deceptive and 50 nondeceptive individuals. Veracity decisions were obtained for each examination and scorer by combining the assigned values according to conventions used at the two institutions. Results suggest no differences in chart evaluation ability among scorers based at the two institutions and that observed accuracy differences may be due to differences in data evaluation conventions. The highest accuracy was obtained when the University of Utah data evaluation conventions were used.

The psychophysiological detection of deception (PDD) serves a critical complementary function within the forensic sciences, especially in cases where there is little physical evidence to support an individual's involvement in a crime. In other words, in lieu of sufficient incriminating forensic evidence, the PDD represents a powerful and effective method for determining suspect case involvement. As with any methodology, attempts are constantly made to improve the accuracy and effectiveness of the PDD. Such attempts typically originate within the confines of controlled laboratory settings, using pretend or mock crimes. These mock crimes are attempts to mirror the reality of an actual criminal situation, as much as possible, with the obvious limitation that laboratory participants are not in jeopardy of arrest or imprisonment. Despite the lack of jeopardy, the mock crime is arguably the most effective method of generating PDD data for which examinee veracity is reliably known. We thus chose to use laboratory data to examine alternative PDD data evaluation methods. If methods not currently supported by federal policy prove useful in the laboratory, then the policy changes and costs necessary to investigate field data may be justified. In other words, if empirical research supports the efficacy of a given procedure in the laboratory, then such a procedure may warrants exploration in field settings.

Since 1978, investigators from only two institutions, the University of Utah (UoUt) and the Department of Defense Polygraph Institute (DoDPI), have consistently published original research on PDD. Clear accuracy discrepancies have been reported by scorers at these institutions when evaluating PDD examinations in laboratory environments.

[1]The opinions expressed in this article are exclusively those of the authors, and do no necessarily represent those of the Department of Defense or the U.S. Government.

[2]Correspondence concerning this article should be sent to Stuart Senter (senters@jackson-dpi.army.mil), Department of Defense Polygraph Institute, 7540 Pickens Avenue, Fort Jackson, SC, 29207.

These discrepancies are depicted in Table 1 which summarizes the decision accuracy obtained in laboratory specific issue studies conducted at the two institutions over the last 22 years. Excluding No Opinion (NO, i.e., undecided or inconclusive) decisions, the weighted mean accuracy produced in the DoDPI studies (79.2%) is substantially lower than that of the UoUt studies (91.1%). The same trend exists when NO decisions are included as errors (*M* = 60.5% versus 79.8% for DoDPI and UoUt, respectively).

Among the factors that could cause the discrepancy in accuracy rates obtained by scorers at the UoUt and the DoDPI are: participant characteristics, participant manipulation methods, physiological tracing quality, scoring systems, data evaluation conventions, and efficiency of applying scoring system and data evaluation conventions. As a preliminary step toward resolving the discrepancy, this study was conducted to determine the impact of the different data evaluation conventions used by the two institutions.

During a PDD examination, the participant is asked a series of questions while physiological reactions are recorded for subsequent evaluation. The questions are usually categorized as irrelevant (e.g., "Is today Thursday?"), comparison (e.g., "Before the age of 18, did you ever take anything of value from someone who trusted you?") or relevant (e.g., "Did you steal that money from that bank?"). Test format refers to question syntax, the number of questions, their presentation order, and the number of times each question is presented.

Table 1
Accuracy of Specific Issue Laboratory Studies Conducted at the Department of Defense Polygraph Institute and the University of Utah

	Deceptive Group Decisions			Nondeceptive Group Decisions			Percent Correct NO Decisions	
Study	Corr	Incorr	NO	Corr	Incorr	NO	With	Without
Department of Defense Polygraph Institute								
Blackwell (1994)	32	14	14	43	6	11	63	80
DoDPI Staff (2001)	24	2	6	15	8	9	61	78
Honts et al. (1989)	30	7	3	18	2	0	80	84
Honts (1992)	29	21	22	51	6	22	53	75
Ingram (1996a)	6	1	4	5	2	4	50	79
Ingram (1996b)	11	0	3	6	2	7	59	90
University of Utah								
Honts et al. (1987)	8	0	2	7	2	1	75	88
Honts et al. (1994)	14	4	2	15	2	3	73	83
Kircher & Raskin (1988)	65	3	6	63	5	6	87	94
Podlesny & McGehee (1987)	59	7	7	17	6	2	78	85
Podlesny & Truslow (1993)	61	4	7	9	5	10	73	89
Raskin & Hare (1978)	21	0	3	21	2	1	88	96
Raskin et al. (1988)	55	3	12	27	1	8	77	95
Rovner (1986)	12	0	0	9	1	2	88	96

Note. NO = No Opinion, Corr = Correct, Incorr = Incorrect

Most PDD examiners monitor thoracic and abdominal respiration, electrodermal activity using either resistance or conductance, and cardiovascular activity using a blood pressure cuff (the auscultatory cuff method). UoUt investigators typically monitor cardiovascular activity using a photo-plethysmograph, in addition to the blood pressure cuff.

The recorded physiological data are subsequently evaluated manually or by computer. During manual evaluation of PDD examinations that include comparison questions, scorers compare the reaction following a comparison question to that following a relevant question for each physiological measure. A comparison and relevant question pair are typically presented at least three times during a PDD examination. A score indicating the extent of the reaction difference is assigned to each question pair and each physiological measure. Scores between –3 and +3 or between –1 and +1 (inclusive) are assigned, depending on whether the 7- or 3-position scale is used. The scores are based on the amplitude, frequency, or duration, or a combination thereof, of responses. Details of the scoring systems used by the DoDPI and the UoUt are documented by Swinford (1999), Bell, Raskin, Honts, and Kircher (1999), and the Federal Psychophysiological Detection of Deception Handbook (2001).

Decisions regarding participant veracity are made using scores assigned to the physiological responses. The DoDPI and UoUt both use total score cutoff criteria. That is, scores assigned to each pair of reactions and physiological measure are added together and a decision of deception indicated (DI), no deception indicated (NDI), or NO is contingent on the total. If the total is -6 or less the decision is DI, if the total is +6 or more then the decision is NDI, if the total is between –6 and +6 then the decision is NO (Bell et al., 1999; Federal Psychophysiological Detection of Deception Handbook, 2001; Swinford, 1999).

In addition, the DoDPI uses a "spot score" rule (Federal Psychophysiological Detection of Deception Handbook, 2001) where all of the scores assigned to a comparison and relevant question pair are summed over repeated presentations of the question pair to produce a total for that question pair. This total is called a spot score. The DoDPI spot score rule is as follows. A participant must have a +1 or greater on all spot scores and a total score of +6 or greater to be classified as NDI; a participant with a –3 or less on any spot score or a total of –6 or less is classified as DI; examinations that do not meet either the DI or NDI criteria are assigned a decision of NO.

The DoDPI teaches that three question series (or charts) should be recorded during a specific issue examination. A fourth question series may be recorded if a question in the earlier series cannot be evaluated. The three-chart rule may have been adopted because it was believed that data produced after three question series were less diagnostic or useful due to habituation (Balloun & Holmes, 1979; Suzuki & Hikita, 1964). However, recent work has shown that the strength and diagnostic value of the data signals do not degrade with additional presentations (Dollins, Cestaro, & Pettit,1998; Elaad & Ben-Shakar, 1997; Nakayama & Kizaki, 1990; Yankee & Grimsley, 1987). The UoUt approach uses the data from either three or five question series. If a decision of DI or NDI is reached after the first three question series, then only those data are used. If a NO decision is reached after evaluating the first three question series then data from two additional question series are evaluated and a decision is made using the data from all five question series. The simple total +6 and –6 cutoff scores are used whether 3 or 5 question series are evaluated.

A final difference to be explored in this study is in the number of physiological measures recorded at the two institutions. The UoUt and DoDPI both record and evaluate respiratory, cardiovascular (ascultatory cuff) and electrodermal measures. As mentioned above, the UoUt also evaluates data from a finger photo-plethysmograph (an index of peripheral vasoconstriction). Research has shown that peripheral vasoconstriction is a useful predictor of participant veracity, though inferior in terms of diagnosticity to one or more of the other data channels (Cutrow, Parks, Lucas, & Thomas, 1972; Podlesny & Raskin, 1978; Suzuki, 1965; Thackray & Orne, 1968). The impact of these procedural

differences on decision frequencies was evaluated in the present study.

Method

Laboratory data collected from 50 deceptive and 50 nondeceptive participants (Kircher & Raskin, 1988) were evaluated by four different scorers. Responses to five question series were collected from each participant and the scorers evaluated each question series once. Two of the scorers used the UoUt scoring system (Bell et al., 1999) and two used the DoDPI scoring system (Swinford, 1999). The data collected from the two scorers using the UoUt scoring system were previously described by Kircher and Raskin (1988). Each scorer completed an evaluation sheet by assigning a numerical value between –3 and +3, inclusive, to responses following each relevant and comparison question pair for each physiological measure (i.e., respiratory activity, skin conductance, cardiovascular activity, and peripheral vasoconstriction recorded via photo-plethysmograph). The scores assigned to each response were totaled and decisions made using the UoUt (+6 and –6 cutoffs, 3 or 5 question series, and inclusion of the photo-plethysmograph measure) and DoDPI conventions (spot scores, 3 question series, and no photo-plethysmograph measure).

Chi-square analyses were calculated for the initial set of comparisons, and used as a global test of differences. To best discern the differential effects of the spot score rule, number of question series, and use of a photo-plethysmograph, a 2 x 2 x 2 repeated measures analysis of variance (ANOVA) (Keppel, 1991) was conducted. An alpha level of .05 was adopted. If violations of homogeneity assumptions were suspected, the alpha level was reduced to .025 as suggested by Keppel (p. 108). Effect size measures were calculated for all inferential tests.

Results

Table 2 shows the frequencies of scorers' correct, incorrect, and NO decisions as a function of participant veracity, calculated according to the DoDPI (i.e., +6 and –6 cutoffs, spot scores, 3 charts only, and no photo-plethysmograph measure) and UoUt (i.e., +6 and –6 cutoffs, 3 or 5 charts, and including the photo-plethysmograph measure) data evaluation conventions. Correct decision frequencies for each scorer were produced by summing the number of DI decisions for deceptive participants and the number of NDI decisions for nondeceptive participants. Incorrect decision rates were produced by summing the number of NDI decisions for deceptive participants and the number of DI decisions for nondeceptive participants. NO decision rates were calculated by summing the number of NO decisions for deceptive and nondeceptive participants.

Chi-square analyses indicated that there were no statistically significant differences between the DoDPI and UoUt scorer frequencies ($X^2(6) = 7.64$, $p > .05$, Cohen's w = .05 [Cohen, 1998]) when decisions were based on the UoUt data evaluation conventions, nor were there statistically significant differences between the DoDPI and UoUt scorer frequencies ($X^2(6) = 12.5$, $p > .05$, Cohen's w = .18) when decisions were based on the DoDPI data evaluation conventions.

An ANOVA was calculated to compare the frequency of correct decisions produced by each scorer using every combination of the three factors, varied across the two data evaluation conventions. The means and standard deviations of the resulting eight sets of accuracies are shown in Table 3. The main effect of question series was significant, $F(1,3) = 20.3$, $p < .025$, eta = .93 (Rosenthal & Rosnow, 1991), reflecting the large increase in the number of correct decisions when 3 or 5 question series were used ($M = 79.4$, $SD = 1.71$) versus when 3 question series were used ($M = 68.0$, $SD = 5.89$). Significantly more correct decisions were obtained ($F[1,3] = 60.5$, $p < .025$, eta = .98) when using numerical totals ($M = 75.1$, $SD = 3.80$) compared to the DoDPI spot score rule ($M = 72.3$, $SD = 3.26$).

The interaction of question series and data channel was also significant, $F(1,3) = 15.1$, $p < .025$, eta = .91. We interpret this interaction as indicating that there is a greater increase in accuracy with the inclusion of the plethysmograph when three question series were used (70.0 vs. 66.0), than when three or five question series were used (80.0 vs. 78.8).

Table 2
Frequency of Correct, Incorrect, and No Opinion Decisions Obtained when the Department of Defense Polygraph Institute and the University of Utah Data Evaluation Conventions are Applied to Assigned Scores

Convention and Scorer	Deceptive (n=50)			Nondeceptive (n=50)		
	Correct	Incorrect	NO	Correct	Incorrect	NO
Department of Defense Polygraph Institute						
1	42	1	7	26	10	14
2	43	3	4	30	7	13
3	36	1	13	26	4	20
4	37	2	11	26	3	21
Mean (Rounded)	39	2	9	27	6	17
University of Utah						
1	39	6	5	41	3	6
2	41	5	4	42	4	4
3	39	2	9	41	2	7
4	44	3	3	43	3	4
Mean (Rounded)	41	4	5	42	3	5

Note. NO = No Opinion

A planned comparison between the accuracy produced using the UoUt data evaluation conventions (82.5) versus the accuracy produced using the DoDPI data evaluation conventions (66.5) was significant, $F(1,3) = 25.6$, $p < .025$, eta = .95 (see Table 3). No other main effects or interactions were significant.

Table 3
Means and Standard Deviations of Correct Decisions Produced by the Factorial Combination of Cutoff Rule, Question Series Approach, and the Inclusion or Exclusions of the Photo-Plethysmograph Data Channel (N=4)

Cutoff Rule	3 Question Series		3 or 5 Question Series	
	M	*SD*	*M*	*SD*
Excluding Plethysmograph				
Absolute	65.5	8.27	81.3	1.50
Spot Score	66.5	5.07	76.3	2.99
Including Plethysmograph				
Absolute	71.0	6.06	82.5	3.32
Spot Score	69.0	4.55	77.5	1.00

Scorer decision reliability was examined to further compare the two institutions' data evaluation conventions. Table 4 shows the proportion of agreement among scorers when veracity decisions were derived using the DoDPI and UoUt data evaluation conventions. Each value is the proportion of agreement between two scorers when identifying an examinee as DI, NDI, or NO using a particular data evaluation convention. Higher values indicate a greater degree of reliability across scorers. The .820 average proportion of agreement obtained using the UoUt data evaluation conventions was not significantly greater ($Z = 1.51$, $p > .05$, $\Phi = .15$ [Rosenthal & Rosnow, 1991) than the .713 average proportion of agreement obtained using the DoDPI data evaluation conventions.

Discussion

The results from the present study suggest that, at least for laboratory data, the differences between the UoUt and DoDPI veracity decision accuracies are likely not due to the institute-specific guidelines used to assign values to physiological responses. The differences are, rather, likely due to the data evaluation conventions used to make decisions after the values are assigned. There was a clear 16.0% increase in correct decisions when using the UoUt evaluation conventions of: (a) +6 and –6 cutoff criteria, (b) no spot scores, (c) 3 or 5 question series, and (d) inclusion of the photo-plethysmograph measure, relative to the DoDPI evaluation conventions. Other factors that were not investigated, such as participant characteristics, participant manipulation methods, and physiological tracing quality could also contribute to the observed discrepancies found in other studies.

These results have two implications. First, the accuracy differences between the DoDPI and UoUt may be attributable to decision conventions following the scoring of the physiological data, and not differences between the scoring systems used at the two institutions. Second, resolving NO decisions by increasing the number of question series asked produced the greatest increase in accuracy, with the 3 or 5 question series approach producing an 11.4% accuracy increase over the 3 question series approach. Use of the total scoring rule versus the spot score rule also produced a significant increase in accuracy, though to a smaller degree (2.8%). The inclusion or exclusion of the photo-plethysmograph did not produce significant accuracy differences. The 16% accuracy difference between the two sets of scoring conventions is comparable to the 12% accuracy difference produced by the weighted means of the studies displayed in Table 1.

We point out that conclusions drawn from this study should be accepted with caution. The demonstration of differential effectiveness between the two scoring conventions was accomplished using a data set collected by the UoUt, and hence represent a possible bias in sampling. Further work comparing these rule conventions should examine data collected by other sources, including DoDPI. This will help to determine whether the differences produced by the two scoring conventions generalize to other data sets.

Table 4
Proportion of Agreement Between Scorers Using the DoDPI and UoUt Data Evaluation Conventions

	DoDPI Conventions Scorer			UoUt Conventions Scorer		
Scorer	1	2	3	1	2	3
2	.66			.75		
3	.72	.67		.86	.79	
4	.66	.68	.89	.88	.80	.84

Note. DoDPI = Department of Defense Polygraph Institute, UoUt = University of Utah

In conclusion, the results suggest that a change in DoDPI data evaluation conventions may improve laboratory-based veracity decision accuracy and remove the historical decision accuracy difference between the DoDPI and UoUt. Furthermore, the examination of scorer agreement using the two data evaluation conventions indicates that shifting to the UoUt data evaluation conventions may improve reliability across scorers. Because these results were produced using laboratory based data, with a relatively small number of scorers, further research is necessary to determine the replicability of these effects, both in a laboratory context, and ultimately with field data. The results of this project supply the impetus for further investigation of the evaluation conventions.

Acknowledgments

The authors wish to thank the scorers who contributed the data for this study: William Gary, Shirley Sturm, Sheila Thomas, Dr. John Kircher, and Dr. David Raskin. We would also like to thank Charlene Stephens for assisting with the data collection. Funds for this research were provided by the Department of Defense Polygraph Institute under project number DoDPI97-P-0005. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

References

Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology, 64*, 316-22.

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

Blackwell N. J. (1994). An evaluation of the effectiveness of the Polygraph Automated Scoring System (PASS) in detecting deception in a mock crime analog study. Fort McClellan (AL): Department of Defense Polygraph Institute. Report No.: DoDPI94-R-0003, DTIC# ADA305755.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cutrow, R. J., Parks, A., Lucas, N., & Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology, 9*, 578-88.

DoDPI Staff (2001). Test of a mock theft scenario for use in the psychophysiological detection of deception: IV. *Polygraph, 30*(4), 244-253.

Dollins, A. B., Cestaro, V. L., & Pettit, D. J. (1998). Efficacy of repeated physiological detection of deception testing. *Journal of Forensic Sciences, 43*, 1016-1023.

Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology, 34*, 587-596.

Federal Psychophysiological Detection of Deception Examiner Handbook (2001). Fort Jackson (SC): Department of Defense Polygraph Institute.

Honts, C. R., Barland, G. H., & Barger, S. D. (1989). The relative validity of criminal and screening approaches to the detection of deception [Abstract]. *Psychophysiology, 26.* Suppl 4A:533.

Honts, C. R. (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. *Forensic Reports, 5*, 215-18.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Psychophysiology, 1*, 241-47.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology, 79*, 252-59.

Ingram, E. M. (1996a). Test of a mock theft scenario for use in the psychophysiological detection of deception: I. Fort McClellan (AL): Department of Defense Polygraph Institute. Report No.: DoDPI96-R-0003, DTIC# ADA318988.

Ingram, E. M. (1996b). Test of a mock theft scenario for use in the psychophysiological detection of deception: III. Fort McClellan (AL): Department of Defense Polygraph Institute. Report No.: DoDPI97-R-0003, DTIC# ADA321138.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). New Jersey: Prentice-Hall.

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

Nakayama, M., & Kizaki, H. (1990). Usefulness of the repeated presentation of questions on the psychophysiological detection of deception. *Japanese Journal of Psychology, 60*, 390-93.

Podlesny, J. A., & McGehee, C. M. (1987). Investigative detection of deception: Role discrimination with a general question technique [Abstract]. *Psychophysiology, 24*, 605.

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology, 15*, 344-59.

Podlesny, J. A., & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology, 78*, 788-97.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15*, 126-36.

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). Validity of control question polygraph tests in criminal investigation [Abstract]. *Psychophysiology, 25*, 476.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis* (2nd ed.). New York: McGraw-Hill.

Rovner, L. I. (1986). The accuracy of physiological detection of deception for subjects with prior knowledge. *Polygraph, 15*, 1-39.

Suzuki, A. (1965). The plethysmograph as an index of the lie detector. *Reports of the National Research Institute of Police Science, 18*, 288-293.

Suzuki, A., & Hikita, Y. (1964). An analysis of response on polygraph: A diminution of responses. *Polygraph, 10*, 1-7.

Swinford J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph, 28*, 10-27.

Thackray, R. I., & Orne, M. T. (1968). A comparison of physiological indices in detection of deception. *Psychophysiology, 4*, 329-339.

Yankee, W. J., & Grimsley, D. L. (1987). The effect of a prior polygraph test on subsequent polygraph tests [Abstract]. *Psychophysiology, 24*, 621-22.

Comparison of Question Series and Decision Rules: A Replication[1]

Stuart M. Senter[2] and Andrew B. Dollins

Abstract

Senter, Dollins, and Krapohl (2004) found evidence that the Utah data evaluation conventions produced more correct decisions than the DoDPI data evaluation conventions. In this study, four evaluators assigned scores to data collected during a laboratory study of 16 deceptive and 16 nondeceptive participants. Following score assignment, decisions were coded using the factorial combination of three or three to five question series and spot scores or no spot scores. Analyses of correct decisions showed that decisions coded using three to five question series produced significantly higher accuracy than decisions using three question series. Thus, the results of the present study replicated those of Senter et al. (2004), and suggest that using three to five question series produces a substantial (e.g., 7.8%) increase in the number of correct decisions, at least within the context of laboratory-based physiological detection of deception studies.

Senter, Dollins, and Krapohl (2004) investigated data evaluation conventions used by investigators at the Department of Defense Polygraph Institute (DoDPI) and the University of Utah (UoUt) to determine the source of accuracy differences reported by researchers from the two institutions. We found that veracity decisions obtained using data evaluation conventions described by Kircher and Raskin (1988) were, on average, 16% more accurate than decisions obtained using the DoDPI data evaluation conventions. The differences were not due to rules used to evaluate specific physiological reactions, but instead were due to conventions used to make decisions after the physiological data were scored. The data evaluation conventions of the two institutions differed in three areas: the use of 'spot scores'; the use of additional data to resolve no opinion (NO; inconclusive or undecided) decisions; and, the use of an additional data channel. Senter et al. (2004) found that the use of additional data by the UoUt was the greatest contributor to the difference in accuracy.

Typically, during a psychophysiological detection of deception (PDD; e.g., polygraph or lie detection) examination, an examinee's physiological responses to questions are evaluated. Values, ranging from -3 to +3, using a seven position scale, or -1 to +1 using a three position scale, are assigned to each data channel for each relevant-comparison question pair. Relevant questions are directed at the examinee's involvement in the event in question (i.e., 'Did you steal any of that money?'). Comparison questions are directed at examinee's other past behaviors (i.e., 'Prior to this year, did you ever steal anything from someone who trusted you?'). In theory, deceptive examinees will produce stronger responses to the relevant questions and nondeceptive examinees will produce stronger responses to the comparison questions. If the response following the relevant question is larger than that following the comparison question, a negative value is assigned to that pair. If the response following the comparison question is larger than that following the relevant, a positive value is assigned. If no differences could be detected between the responses following the two types of questions. a value of zero is assigned. When using the Zone Comparison Test format, each question series typically includes three relevant-comparison question pairs (DoDPI, 1992;

[1] The opinions expressed in this article are exclusively those of the authors, and do no necessarily represent those of the Department of Defense or the U.S. Government.

[2]Correspondence concerning this article should be sent to Stuart Senter (senters@jackson-dpi.army.mil), Department of Defense Polygraph Institute, 7540 Pickens Avenue, Fort Jackson, SC, 29207.

Federal Psychophysiological Detection of Deception Handbook, 1999). For reviews of response features used in assigning scores, see Bell, Raskin, Honts, and Kircher (1999) and Swinford (1999).

The DoDPI and UoUt differ procedurally in how veracity decisions are made using the scores assigned to physiological responses. The first procedural difference between the two institutions involves the use of score totals to produce veracity decisions. The UoUt convention is to use a total cutoff rule whereby all of the assigned scores are summed and a decision of deceptive is made if the total is –6 or lower, a decision of nondeceptive if the total is +6 or greater, and a decision of NO if neither cutoff is reached (Bell et al., 1999). The DoDPI convention is to total the assigned scores for each of the three relevant-comparison pairs across each question series and data channel to obtain 'spot' scores (Federal Psychophysiological Detection of Deception Handbook, 1999; Swinford, 1999). A decision of deceptive is rendered if any spot score is –3 or less or if the totaled spot scores are –6 or less. A decision of nondeceptive is rendered if the totaled spot scores produce a value of +6 or greater and if the value for each spot is positive. A decision of NO is rendered if none of these criteria are met.

The second procedural difference involves question series. The DoDPI uses three question series, after which a decision of deceptive, nondeceptive, or NO is rendered. The UoUt uses three question series if a decision of deceptive or nondeceptive can be rendered, and two additional question series (for a total of five question series) if a NO decision is made after three question series.

Third, the UoUt and the DoDPI also differ in the number of data channels monitored. The DoDPI uses respiratory (thoracic and abdominal sensors), cardiovascular, and electrodermal measures. The UoUt uses these measures plus a photo-plethysmograph, a measure of peripheral blood flow. In the present study, only respiratory, cardiovascular, and electrodermal measures were collected. Thus, number of data channels was not considered as a variable.

The purpose of the present study was to determine whether the accuracy advantage found by Senter et al. (2004) using the UoUt data evaluation conventions would replicate with a different data set and different scorers. We expected the present study to replicate Senter et al. In addition, we predicted that the greatest contributor to the accuracy difference would be the use of additional question series.

Method

Four federally certified PDD examiners scored the physiological data from 16 deceptive and 16 nondeceptive examinees. Physiological data were recorded during a previous study (DoDPI Research Division Staff, 2001) using the Zone Comparison Test format. Examinee responses to five presentations of a question series were recorded.

Scorers used the DoDPI scoring system (Swinford, 1999) to independently evaluate each examinee's physiological responses following the repeated presentation of relevant-comparison question pairs using. Our previous research indicated that the DoDPI and UoUt do not differ significantly in the assignment of scores to physiological responses despite differences in scoring rules (Senter et al., 2004). Scorers were required to indicate their reasons for assigning values on a score sheet containing a checklist (Appendix A) developed by Capps (1993). Examiners in a previous study expressed the belief that that the checklist requirement caused them to more carefully consider the reason specific scores are assigned (Capps, 1993). Data collected with these checklists will be examined in a separate study. Scorers were also given an instruction sheet with definitions for each physiological feature (Appendix B). Three scorers assigned numerical values between –3 and +3, inclusive, to each data channel (i.e., respiration, skin conductance, and cardiograph), for each relevant-comparison pair. One scorer used a three-position scale (+1 to –1) because this was the approach this individual typically used in the field. The numerical scores collected from each scorer were totaled to produce decisions using either three question series or three to five question series, and using either the total cutoff rule or the spot score rule. A

significance criterion of .05 was used for all statistical tests.

Results

Table 1 shows the means and standard deviations as a function of decision rule, number of question series, and participant veracity. Decision frequencies were analyzed using 2 (3 vs. 3 to 5) x 2 (total cutoff vs. spot) x 2 (deceptive vs. nondeceptive) analyses of variance (ANOVA). For correct decisions, the main effect of Question Series was significant, $F(1,3) = 50.0$, $p < .05$, $\omega^2 = .86$, indicating that there were significantly more correct decisions when three to five question series were used (M = 23.9) than when only three question series were used (M = 21.4), a difference of 7.8% overall. The Decision Rule x Veracity interaction effect was also significant, $F(1,3) = 56.8$, $p < .05$, $\omega^2 = .78$, reflecting differences in correct decisions as a function of decision rule and participant veracity. The spot score rule produced more correct decisions with deceptive participants than with nondeceptive participants (M = 11.9 vs. M = 10.9, respectively), and the total cutoff rule produced more correct decisions with nondeceptive participants than with deceptive participants (M = 13.9 vs. M = 8.6, respectively). No other effects were significant.

Table 1
Means and Standard Deviations for the Number of Correct, Incorrect, and No Opinion Decisions Produced by a Factorial Combination of Decision Rule, and Question Series Usage (N=4).

	3 Series		3 to 5 Series	
Veracity Decision Rule	*M*	*SD*	*M*	*SD*
Deceptive Participants				
Correct Decisions				
Total Cutoffs	8.3	0.50	9.0	1.16
Spot Score	11.3	0.96	12.5	1.00
Incorrect Decisions				
Total Cutoffs	1.8	0.50	2.8	0.96
Spot Score	1.0	1.16	1.5	1.00
No Opinion Decisions				
Total Cutoffs	6.0	0.82	4.3	1.26
Spot Score	3.8	1.26	2.0	2.00
Nondeceptive Participants				
Correct Decisions				
Total Cutoffs	13.3	0.96	14.5	1.29
Spot Score	10.0	1.16	11.8	1.26
Incorrect Decisions				
Total Cutoffs	0.0	0.00	0.3	0.50
Spot Score	1.3	0.96	2.0	0.82
No Opinion Decisions				
Total Cutoffs	2.8	0.96	1.3	1.50
Spot Score	4.8	1.26	2.3	0.96

Overall, more correct decisions were produced using the UoUt conventions (M = 23.5) than the DoDPI conventions (M = 21.3). A planned comparison indicated that this difference was significant, $F(1,3) = 81.0$, $p < .05$, $\omega^2 = .91$.

Additional Question Series x Decision Rule x Veracity ANOVAs were calculated for the frequency of incorrect and NO decisions to determine the impact of these variables. The Decision Rule x Veracity interaction effect was significant for incorrect decisions, $F(1,3) = 12.5$, $p < .05$ $\omega^2 = .42$. The total cutoff rule produced more errors with deceptive participants (M = 2.3) than with nondeceptive participants (M = 0.1), while the spot score rule showed only small differences across deceptive and nondeceptive particpants (M = 1.3 vs. M = 1.6).

The main effect of Question Series for NO decisions was significant, $F(1,3) = 51.9$, $p < .05$, $\omega^2 = .86$. This reflected the decrease in the frequency of NO decisions when three to five question series were used (M = 4.9), relative to when three question series were used (M = 8.6). The Decision Rule x Veracity interaction effect was also significant, $\underline{F}(1,3) = 51.9$, $p < .05$, $\omega^2 = .76$, reflecting differences in NO decisions as a function of decision rule and participant veracity. The spot score rule produced fewer NO decisions with deceptive participants than with nondeceptive participants (M = 2.9 vs. M = 3.5, respectively), and the total cutoff rule produced fewer NO decisions with nondeceptive participants than with deceptive participants (M = 2.0 vs. M = 5.1, respectively). No other effects were significant.

Table 2 shows the pairwise proportion of agreement between scorers, in addition to the proportion of correct decisions for each scorer. The average proportion of agreement was higher when the UoUt rules were used (M = .82) than when the DoDPI rules were used (M = .69). While these levels of agreement were not significantly different from each other, Z = 1.05, $p > .05$, each pairwise proportion of agreement is higher for the UoUt rules than for the DoDPI rules. The levels of agreement are also comparable to those reported by Senter et al. (2004).

Table 2
Pairwise Proportion of Agreement Between Scorers, Total Accuracy, and Accuracy Excluding NO Decisions

Convention and Scorer	Scorer			Accuracy with NO	
	2	3	4	Included	Excluded
DoDPI Rules					
1	.84	.63	.63	.66	.91
2		.66	.66	.69	.92
3			.72	.66	.88
4				.66	.91
UoUt Rules					
1	.91	.75	.81	.72	.82
2		.75	.81	.75	.92
3			.88	.72	.89
4				.75	.92

Note. NO = No Opinion

Discussion

The results of this study provide further evidence that decision accuracy differences reported by the UoUt and the DoDPI can be largely attributed to differences in the use of additional question series. In addition, the results provide no evidence that use of spot scores versus total score cutoffs contributes to overall decision accuracy differences. However, the results do indicate that use of the spot score rule produces higher accuracy with deceptive participants than with nondeceptive participant, while the reverse is true for the total cutoff rule. The three to five question series contingency rule produced a substantial increase in the number of correct decisions reported in our earlier (11.4%, Senter et al., 2001) and present study (7.8%). In both studies, this increase in the number of correct decisions is mostly due to the resolution of NO decisions, with no significant change in the number of incorrect decisions.

The percentage of correct, incorrect, and NO decisions produced using the UoUt conventions (total cutoffs, three to five question series) in the present study varied greatly across deceptive and nondeceptive participants (see Table 1). The percentage of correct decisions, while quite high for nondeceptive participants (90.6%), was substantially lower for deceptive participants (57.8%). In addition, the percentage of incorrect and NO decisions was much lower for nondeceptive participants (1.6% and 7.8%, respectively) than for deceptive participants (17.2% and 25.0%, respectively) when the UoUt conventions were used. While the percentages of correct, incorrect, and NO decisions did fluctuate across deceptive and nondeceptive participants with the DoDPI conventions, all varied by less than 10%. As the significant Decision Rule x Veracity interactions for correct and NO decisions suggest, the difference in decision frequencies across participant veracity is largely attributable to the fact that the UoUt conventions employ the total cutoff rule and the DoDPI conventions employ the spot score rule. Further research should be conducted to identify new scoring approaches. It might be possible to find rule combinations that produce an overall increase in veracity decision accuracy, but not at the expense of compromising a high level of accuracy for deceptive participants.

A possible criticism of the UoUt conventions is the potential cost of collecting two additional charts beyond the standard three collected within the federal government. While this approach clearly increases decision accuracy, it adds additional time to the data collection phase of the polygraph process. The prolonged data collection phase could result in examiners missing a window of opportunity for obtaining a confession.

The results of the present study differed from those of Senter et al. (2004) in that decision rule (i.e., spot score versus total cutoff) did not produce any significant differences in overall decision accuracy, though decision accuracies produced by the decision rules differed as a function of participant veracity. Senter et al. found that total cutoffs produced a small (2.8%) but significant increase in the number of correct decisions. This increase was meager (0.8%) and nonsignificant in the present study. Excepting this discrepancy, the present study, did replicate the findings of Senter et al. However, there were a number of limitations to this replication. First, the inclusion/exclusion of the plethysmograph data channel was not included as a variable in the present study. Second, all four scorers in the present study assigned values using the DoDPI features. In the Senter et al. study, two scorers assigned values using the DoDPI features and two scorers assigned values using the UoUt features. Finally, the question series used in this study were collected by DoDPI examiners using Axciton computerized polygraph instruments, while the questions series used by Senter et al. were collected by UoUt researchers using a Beckman Type R Dynograph which is an analogue instrument. The impact of the different data collection personnel and instruments is unknown.

There were other general limitations to this study. Data were collected from only four scorers, a number that should be increased in future studies to increase statistical power and determine the extent to which these effects generalize across different examiners. In addition, one of the four scorers assigned values using a three-position scale while the

other scorers used a seven-position scale. While the impact of this inconsistency is unknown, future designs should emphasize the consistency of the scale used for score assignment.

In summary, the results of the current study indicate that substantial increases in the number of correct decisions can be achieved through the use of three to five question series, relative to that obtained using three question series. This increase in accuracy is primarily due to the correct resolution of NO decisions, with little change in the number of errors. The combined results of Senter et al. (2004) and this study suggest that the use of the three to five questions series rule could be beneficial within the federal government. One limitation of this study in terms of its applied value is that this robust effect has been produced only with laboratory data. Future efforts should examine the impact of the different question series usage rules with data collected during actual cases.

Acknowledgements

The authors would like to thank David Cameron of the Defense Intelligence Agency, D. Perry Sell and Ronald Sharetts of the Air Force Office of Special Investigations, and Terry Wachtel of the Department of Defense for their work on this project. The physiological data used in this project were collected by N. Joan Blackwell as specified in Department of Defense Polygraph Institute Report DoDPI00-R-0002. This research was funded by the Department of Defense Polygraph Institute, Fort Jackson, South Carolina, as project DoDPI01-P-0008.

References

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

Capps, M. H. (1993). *The contribution of physiological parameters and criterion frequency in polygraph evaluation.* Unpublished master's thesis, University of Baltimore, Baltimore.

Department of Defense Polygraph Institute. (1992). *Zone comparison test.* Fort McClellan, AL: Author.

DoDPI Research Division Staff (2001). Test of a mock theft scenario for use in the psychophysiological detection of deception: IV. *Polygraph, 30*(4), 244-253.

Federal Psychophysiological Detection of Deception Examiner Handbook. (1999, May). Fort Jackson, SC: Department of Defense Polygraph Institute.

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.

Senter, S. M., Dollins, A. B., & Krapohl, D. J. (2004). A comparison of polygraph data evaluation conventions used at the University of Utah and the Department of Defense Polygraph Institute. *Polygraph, 33*(4), 214-222.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph, 28,* 10-27.

Appendix A
Score Sheet

CHART # _____

CRITERIA

Pneumo	**R1**	**R2**	**R3**
Hyperventilation	____	____	____
I/E Ratio Change	____	____	____
Suppression	____	____	____
Base Line Change	____	____	____
Base Line Loss	____	____	____
Apnea	____	____	____
Change in Amplitude	____	____	____
Change in Rhythm/Regularity	____	____	____
Other _____	____	____	____

GSR

Degree of Reaction	____	____	____
Duration of Reaction	____	____	____
Other _____	____	____	____

Cardio

BP Increase & Decrease	____	____	____
BP Increase Only	____	____	____
BP Decrease Only	____	____	____
Pulse Rate Increase	____	____	____
Pulse Rate Decrease	____	____	____
Amplitude Increase	____	____	____
Amplitude Decrease	____	____	____
Other _____	____	____	____

CHART #1

	R1	R2	R3	TOTAL
CHART				
PULSE				
GSR				
# CRITERIA				

CHART # _____

CRITERIA

Pneumo	**R1**	**R2**	**R3**
Hyperventilation	____	____	____
I/E Ratio Change	____	____	____
Suppression	____	____	____
Base Line Change	____	____	____
Base Line Loss	____	____	____
Apnea	____	____	____
Change in Amplitude	____	____	____
Change in Rhythm/Regularity	____	____	____
Other _____	____	____	____

GSR

Degree of Reaction	____	____	____
Duration of Reaction	____	____	____
Other _____	____	____	____

Cardio

BP Increase & Decrease	____	____	____
BP Increase Only	____	____	____
BP Decrease Only	____	____	____

CHART #2

	R1	R2	R3	TOTAL
CHART				
PULSE				
GSR				
# CRITERIA				

Pulse Rate Increase	____	____	____
Pulse Rate Decrease	____	____	____
Amplitude Increase	____	____	____
Amplitude Decrease	____	____	____
Other _____	____	____	____

CHART # _____

CHART # _____

CRITERIA

Pneumo	**R1**	**R2**	**R3**
Hyperventilation	____	____	____
I/E Ratio Change	____	____	____
Suppression	____	____	____
Base Line Change	____	____	____
Base Line Loss	____	____	____
Apnea	____	____	____
Change in Amplitude	____	____	____
Change in Rhythm/Regularity	____	____	____
Other _____	____	____	____

GSR

Degree of Reaction	____	____	____
Duration of Reaction	____	____	____
Other _____	____	____	____

Cardio

BP Increase & Decrease	____	____	____
BP Increase Only	____	____	____
BP Decrease Only	____	____	____
Pulse Rate Increase	____	____	____
Pulse Rate Decrease	____	____	____
Amplitude Increase	____	____	____
Amplitude Decrease	____	____	____
Other _____	____	____	____

CHART 3 P G C

R1	R2	R3	TOTAL

CHART # _____

CRITERIA

Pneumo	**R1**	**R2**	**R3**
Hyperventilation	____	____	____
I/E Ratio Change	____	____	____
Suppression	____	____	____
Base Line Change	____	____	____
Base Line Loss	____	____	____
Apnea	____	____	____
Change in Amplitude	____	____	____
Change in Rhythm/Regularity	____	____	____
Other _____	____	____	____

		R1	R2	R3	TOTAL
C					
H	P				
A					
R					
T	G				
#	C				
4					

GSR			
Degree of Reaction	____	____	____
Duration of Reaction	____	____	____
Other _____	____	____	____

Cardio			
BP Increase & Decrease	____	____	____
BP Increase Only	____	____	____
BP Decrease Only	____	____	____
Pulse Rate Increase	____	____	____
Pulse Rate Decrease	____	____	____
Amplitude Increase	____	____	____
Amplitude Decrease	____	____	____
Other _____	____	____	____

CHART # _____

CRITERIA

Pneumo	**R1**	**R2**	**R3**
Hyperventilation	____	____	____
I/E Ratio Change	____	____	____
Suppression	____	____	____
Base Line Change	____	____	____
Base Line Loss	____	____	____
Apnea	____	____	____
Change in Amplitude	____	____	____
Change in Rhythm/Regularity	____	____	____
Other _____	____	____	____

		R1	R2	R3	TOTAL
C					
H	P				
A					
R					
T	G				
#	C				
5					

GSR			
Degree of Reaction	____	____	____
Duration of Reaction	____	____	____
Other _____	____	____	____

Cardio

BP Increase & Decrease	____	____	____
BP Increase Only	____	____	____
BP Decrease Only	____	____	____
Pulse Rate Increase	____	____	____
Pulse Rate Decrease	____	____	____
Amplitude Increase	____	____	____
Amplitude Decrease	____	____	____
Other _____	____	____	____

Appendix B

CRITERIA DEFINITIONS

Pneumo

Hyperventilation is exhibited when cycles exceed the normal tracing average in amplitude for a continued period.

I/E Ratio Change is exhibited by a change in the ratio between the inhalation stroke of the breathing scale. The ratio is based on the distance from the center line drawn from the peak of the cycle to the beginning on the inhalation stroke as compared to the distance from the same line to the end of the exhalation stroke.

Suppression is exhibited when the respiratory cycles are subdued below the normal tracing average.

Base Line Change occurs when the base line goes up or down for a continued period and then returns to its approximate previous position.

Base Line Loss occurs when the base line goes up or down and does not return to its previous position but maintains its new position.

Apnea is the continued cessation of breathing including holding or blocking.

Change in Amplitude is a continued increase or decrease in the overall amplitude of the cycles during that portion of the tracing under comparison. Includes staircases.

Change in Rhythm/Regularity is exhibited when the pattern changes frequency of repetition or in other ways no longer conforms to the normal established pattern not inclusive of the other criteria.

GSR

Degree of Reaction is measured from the baseline where the vertical rise begins to the highest point of that rise. This is not a diagonal measure.

Duration of Reaction is measured from the beginning of the vertical rise to the point where the pattern has stabilized and a new baseline has been established. This would be inclusive of 'saddle' reactions.

Cardio

Blood Pressure Increase and Decrease is identified by an upward trend in the cardio tracing followed by a downward trend in the cardio tracing.

Blood Pressure Increase is identified by an upward trend in the cardio tracing.

Blood Pressure Decrease is identified by a downward trend in the cardio tracing.

Pulse Rate Increase is identified by an increase in the number of cycles in the cardio pattern.

Pulse Rate Decrease is identified by a decrease in the number of cycles in the cardio pattern.

Amplitude Increase is identified by an increase in the tracing size as measured between the tip of the systolic stroke and the diastolic stroke.

Amplitude Decrease is identified by a decrease in the tracing size as measured between the tip of the systolic stroke and diastolic stroke.

Exploration into the Effect of Race on Polygraph Scores and Decisions[1]

Donald J. Krapohl and William B. Gary, Jr.

Abstract

Arther (1998) has asserted that race can influence the profile of response patterns in polygraph testing, specifically in the cardiovascular recordings. There has been virtually no evidence reported of such an effect in polygraphy. To test Arther's observation, multiple analyses were made of the cardiovascular and other polygraph channels using polygraph charts from confirmed deceptive field cases. Arther's assertion of a difference in response profiles between African-American and Caucasian examinees was not supported by any of the analyses.

The influence of race on physiological responding is an important question in the field of psychophysiological detection of deception (polygraphy). The existing body of psychophysiological literature points to some differences in autonomic responsivity among racial groups (Johnson & Landon, 1965; Juniper & Dykman, 1967; Kugelmass & Lieblich, 1968; Lazarus, Tomita, Opton, & Kodama, 1966; Lieblich, Kugelmass, & Ben-Shakhar, 1973; Murphy, Alpert, Walker, & Willey, 1988; Sternbach & Tursky, 1965). The race of the examinee has not been generally considered in polygraphy for the interpretation of the physiological recordings, however. The failure to take race into account has invited criticism from scientific groups who have issued formal reports on polygraphy (Office of Technology Assessment, 1983; National Research Council, 2002).

A common defense offered by polygraphers is that traditional polygraphy, using the Comparison Question Technique (CQT), evaluates responding within individual examinees, rather than making comparisons across examinees (Abrams, 1989; Reid & Inbau, 1977). This within-subject approach purportedly ameliorates any cross-racial differences that may exist. There is tentative support for this argument in the work of Buckley and Senese (1991), and Reed (1993), who found no significant effects on overall decision accuracy tied to the race of the examinee.

Though the little available research evidence suggests that there are no racial effects in polygraph decision accuracy, at least one notable writer in the field of polygraphy has asserted that there are differences in the profile of physiological responding between races. Arther (1998) contends that "[t]he cardio reactions of blacks are generally not only much greater but also more valid and reliable than those of whites." Because Arther's statement is taken from a training document, there were no data or citations provided to substantiate this claim. However, Arther has over 50 years of professional practice and writing in polygraphy, and his observations carry substantial weight in the field. Of more importance, if Arther's statement about race and channel-preference is true, racial information might be used to improve scoring systems or automated algorithms. Weighting of channels could be introduced to take advantage of responses that are more valid with one group over another. It became our interest to test Arther's notion on racial differences in polygraphy using the polygraph case database of the Department of Defense Polygraph Institute (DoDPI).

[1] This article is one in a series under the heading *Best Practices*. The authors are with the U.S. Department of Defense Polygraph Institute. The views expressed in this paper are those of the authors, and do not necessarily represent those of the Department of Defense or the US government. Reprint requests should be directed to: Donald Krapohl, DoDPI, 7540 Pickens Ave., Ft. Jackson, SC 29207.

Method 1

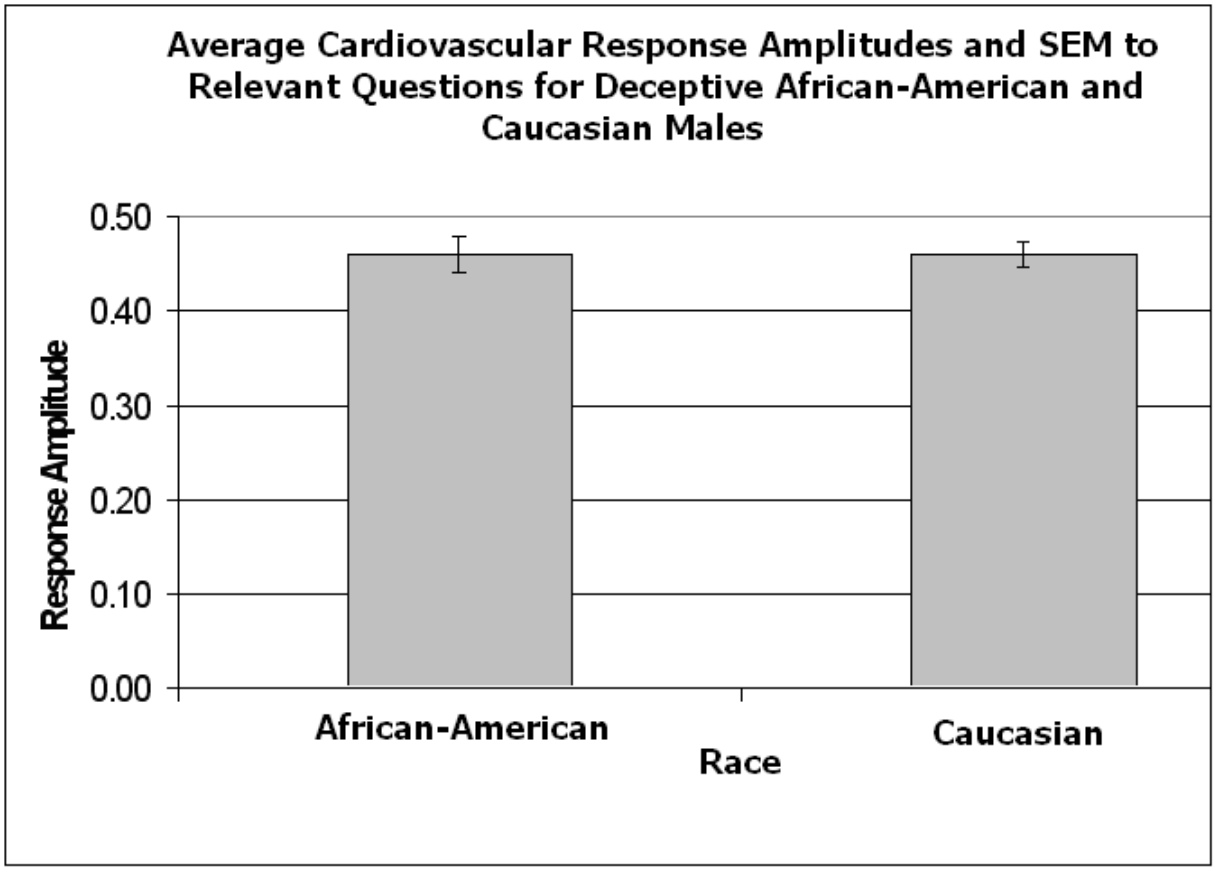
The DoDPI polygraph case database contains cases conducted on Axciton computer polygraphs (Axciton Systems, Houston, TX) by the U.S. government and several local law enforcement agencies. To standardize the sample for this effort, only those cases conducted with the DoDPI Zone Comparison Technique (ZCT) were used, provided that they were single-issue examinations with three relevant questions. From those cases, only those where both race and gender were recorded were selected. It was discovered that in this group there were insufficient numbers of females and confirmed truthful cases, so they were excluded from the present analysis. This left 19 African-American males and 34 Caucasian males for the samples, and all cases were confirmed deceptive.

The "Kircher features" (respiration line length, electrodermal response amplitude, cardiovascular response amplitude) were measured automatically, using software developed for the U.S. government (Extract, ver 3.1). These features have been found to be the most predictive within polygraph tracings (Kircher & Raskin, 1988). The measurements were exported to a spreadsheet for analysis. Significance for all statistical treatments was set at 0.05.

Results 1

The means of the raw measurements of cardiovascular response to relevant questions for the African-American and Caucasian samples were tested for significant differences. None was found ($z = 0.70$, $p > .05$). Figure 1 below depicts the average cardiovascular reactions for the two groups, with standard error of the mean (SEM) bars.

Figure 1.



.

It could be argued that measures of average responses may obscure a characteristic of more importance in polygraphy, that of differential arousal to relevant and comparison questions. Arther's (1998) statement may be taken to mean that deceptive African-Americans respond significantly greater to relevant questions than comparison questions in the cardiograph, whereas deceptive Caucasians in this same channel have less of a difference in responding to these two types of questions. Therefore, the more meaningful indicator of racial differences could be a comparative analysis of relative arousal, rather than absolute arousal.

Method 2

To test the possibility that races respond differentially to relevant and comparison questions, it was necessary to use a metric for differential arousal. The Objective Scoring System (OSS) was chosen for this task. The Objective Scoring System uses absolute measurements of the "Kircher features" (EDR amplitude, respiration line length, and blood volume amplitude) to create ratios that provide an index of relative arousal to relevant and comparison questions (Dutton, 2000; Krapohl & McManus, 1999). Those ratios are converted to scores with a 7-position scoring system. The final result, the scores, may manifest the racial dissimilarity predicted by Arther.

Results 2

Table 1 below shows the OSS scores for these samples. The differences in scores between the races were small, and not significant: pneumograph ($t[41] = 0.11$, $p > .05$; electrodermal ($t[45] = 0.87$, $p > .05$); cardiovascular ($t[47] = 1.29$, $p > .05$). Decision accuracy, using +/-6 as thresholds, are shown in Table 2.

Table 1. Average OSS scores from 53 cases of deceptive African-American and Caucasian examinees by polygraph channel. No significant differences between races.

	Average OSS Scores	
	African-American	Caucasian
Pneumograph	-5.26	-4.97
Electrodermal	-19.89	-16.47
Cardiograph	-9.53	-6.97

Table 2. Decision accuracy using the OSS with +/-6 cutting scores for deceptive African-American and deceptive Caucasian examinees. No significant differences between races.

	DI	NDI	Inc.	Accuracy w/o Inc
African-American (n=19)	17	0	2	100%
Caucasian (n=34)	30	1	3	97%

.

Though the OSS scores of the data found no differences in differential responding between African-American and Caucasian examinees, the possibility remained that racial differences in reactions might still exist in a form not captured by the objective measures used in this project. Human evaluators, using field methods, may be able to process the data in ways that would reveal the asserted racial differences in the test charts.

Method 3

An experienced polygraph examiner was tasked with evaluating the same charts with the orthodox field methods of global interpretation and numerical scoring. The evaluator was federally certified, and an instructor with DoDPI.

The test charts were printed in the hard copy form familiar to all field polygraphers. The evaluator analyzed the recordings, and provided two types of data. First, he globally assessed the charts, and based on the lability of the cardiograph channel, made forced-choice decisions that the charts were from either African-American or Caucasian examinees. He also scored the charts using the traditional 3-position scoring system (Capps & Ansley, 1992; Harwell, 2000; Krapohl, 1998; Van Herk, 1992).

Results 3

Table 3 shows the evaluator's estimate of the race of the 53 examinees. The evaluator's decision accuracy for race was 52.8%, which was not better than chance (z = 0.13, p > .05). Therefore, global evaluation of the cardiograph did not prove to be a valid indicator of examinee race for these data.

Table 3. Estimate of the race of examinees by a blind evaluator of the polygraph charts. Judgments not better than chance.

		Actual Race: African-American	Caucasian	Total
Race Decision	African-American	6	12	18
	Caucasian	13	22	35
	Total	19	34	53

Table 4. Average manual scores from 53 cases of deceptive African-American and Caucasian examinees by polygraph channel. No significant differences between races.

	Race: African-American	Caucasian
Pneumograph	-0.32	-1.12
Electrodermal	-5.21	-5.82
Cardiograph	-1.47	-1.35

x

The evaluator's scores were analyzed next. Table 4 lists the average of the manual scores for each channel. There were no significant differences found between the racial groups for pneumograph scores (t[37] = 1.39, p > .05), electrodermal scores (t[40] = 0.85, p > .05), or the cardiovascular scores (t[42] = 0.16, p > .05).

General Findings

The present analysis found that deceptive African-American and Caucasian males did not respond differently on the polygraph from members of the other race, at least in terms of Kircher features, OSS scores, manual scoring, and global interpretation. Scores from African-American examinees were not more or less predictive than those of Caucasian examinees for any of the three polygraph channels. An experienced polygraph examiner was unable to determine the race of the examinee from the polygraph data. There was no evidence that inclusion of racial information in scoring or algorithmic systems would yield any benefit in decision accuracy, nor do the data hint that such differences exist. In view of these findings, and the lack of contrary evidence elsewhere, it is premature to suggest that African-Americans and Caucasians respond differently on the polygraph.

There are limitations in this study that warrant note. First, deceptive cases were used in this study because nondeceptive cases were not available in sufficient numbers. Therefore, the question regarding the racial differences in nondeceptive cases is not addressed here. It remains possible that Arther's contention regarding racial differences in the cardiograph may apply to those cases.

Similarly, the lack of available female cases also limits the generalizability of this study. Arther makes no distinction for sex in his contention of racial differences in physiological arousal patterns. It remains possible that females do show Arther's reported racial differences. Buckley and Senese (1991) found that the cardiovascular channel contributed significantly to differences in decision accuracy for truthful African-American females: they were detected at a lower rate than other groups. The Buckley and Senese finding runs contrary to Arther's predication inasmuch as the cardiovascular channel lowered decision accuracy instead of increasing it for truthful African-American females. Because the sample was quite small, a mere five, Buckley and Senese suggested the finding may have been anomalous. Whether there are racial differences in polygraph data for females remains unresolved.

It is also important to add a remark about our choice to use only confirmed cases rather than testing unconfirmed cases as well. It may be argued that unconfirmed cases are qualitatively different from confirmed cases, and that racial differences that might exist in the field are lost when only the subset of confirmed cases are selected for analysis. Our decision to use confirmed cases was based on a single factor: our interest in comparing physiological reactions with ground truth. To answer Arther's speculation that "cardio reactions of blacks are generally not only much greater but also more valid and reliable than those of whites," it was necessary to use cases where validity could be checked. As such, it confined our choice to using cases where ground truth was known. Future researchers may include unconfirmed cases also, to eliminate the possibility that the putative racial effect was not overlooked because of our selection criteria.

In summary, our data did not support Arther's assertion of racial differences in cardiovascular responsiveness, at least with deceptive African-American and Caucasian males. Work remains to determine whether a racial effect is to be found among truthful and untruthful females.

References

Abrams, S. (1989). *The complete polygraph handbook*. Lexington Books: Lexington, MA.

Arther, D. (1998). Arther's Golden Rules: Update. *The Journal of Polygraph Science, 33*(2).

Buckley, J. P., & Senese, L. C. (1991). The influence of race and gender on blind polygraph chart analyses. *Polygraph*, *20*(4), 247-258.

Capps, M. H., & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph*, *21*(1), 39-43.

Dutton, D. W. (2000). Guide for performing the objective scoring system. *Polygraph*, *29*(2), 177-184.

Harwell, E. M. (2000). A comparison of 3- and 7-position scoring scales with field examinations. *Polygraph*, *29*(2), 195-197.

Johnson, L. C. & Landon, M. M. (1965). Eccrine sweat gland activity and racial differences in resting skin conductance. *Psychophysiology*, *1*, 322-329.

Juniper, K. & Dykman, R. A. (1967). Skin resistance, sweat-gland counts, salivary flow, and gastric secretion: Age, race, and sex differences, and intercorrelations. *Psychophysiology*, *4*, 216-222.

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. (1998). A comparison of 3- and 7- position scoring scales with laboratory data. *Polygraph*, *27*(3), 210-218.

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

Kugelmass, S., & Lieblich, I. (1968). The relation between ethnic origin and GSR reactivity in psychophysiological detection. *Journal of Applied Psychology*, *52*(2), 158-162.

Lazarus, R. S., Tomita, M., Opton, E., & Kodama, M. (1966). A cross-cultural study of stress-reaction patterns in Japan. *Journal of Personality and Social Psychology*, *4*(6), 622-633.

Lieblich, I., Kugelmass, S. & Ben-Shakhar, G. (1973). Psychophysiological baselines as a function of race and origin. *Psychophysiology*, *10,* 426-430.

Murphy, J. K., Alpert, B. S., Walker, S. S. & Willey, E. S. (1988). Race and cardiovascular reactivity: A replication. *Hypertension*, *11,* 308-311.

National Research Council (2002). The polygraph and lie detection. Available at http://www.nap.edu/catalog/10420.html.

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation – Technical Memo*. Report No.: OTA-TM-H-15. Washington, DC.

Reed, S. (1993). *Subcultural report -- Effects of examiner's and examinee's race on psychophysiological detection of deception outcome accuracy*. Report DoDPI94-R-0012. Department of Defense Polygraph Institute. Ft. McClellan, AL.

Reid, J. E., & Inbau, F. E. (1977). *Truth and deception: The polygraph ("lie detector") technique (2nd ed.).* Williams & Wilkins: Baltimore, MD.

Sternbach, R. A., & Tursky, B. (1965). Ethnic differences among housewives in psychophysical and skin potential responses to electric shock. *Psychophysiology*, *1,* 241-246.

Van Herk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. *The Newsletter of the Canadian Association of Police Polygraphists*, *1*(1), 28-47.