Polygraph

Volume 11 September 1982 Number 3

Contents

Global Evaluation: An Inductive Approach to Case Resolution
Stanley M. Slowik 215

Effects of Beta Blocking Drugs on the Polygraph Detection Rate: A Pilot Study
E. Elaad, G. Bonwitt, O. Eisenberg and I. Meytes 225

Psychophysiological Detection of Deception in Japan: The Past and The Present
Junichi Fukumoto 234

The Dichotomization Theory for Differential Autonomic Responsivity Reconsidered
Gershon Ben-Shakhar and Israel Lieblich 239

The German Tatbestandsdiagnostik; A Historical Review of the Beginnings of Scientific Lie Detection in Germany
H. Herbold-Wootten 246

On the Accuracy of the Polygraph: An Evaluative Review of Lykken's Tremor in the Blood
Gordon H. Barland 258

PUBLISHED QUARTERLY

P.O. Box 74, Linthicum Heights, Maryland 21090

GLOBAL EVALUATION: AN INDUCTIVE APPROACH TO CASE RESOLUTION

By

Stanley M. Slowik*

Over the last several years, a great deal of controversy has arisen over certain polygraph examiners' use of so-called "Global Evaluation" in association with polygraph examinations.

At one extreme, allegations have been made that global evaluation consists of an examiner merely spreading out a set of polygraph records and superficially gazing over the charts in their entirety. Then, after a few moments of studious contemplation, the examiner strikes an appropriately profound pose and proclaims his opinion in absolute, unequivocable terms.

At the other extreme, critics of polygraph insist that examiners are so obsessed with destroying the innocent and coddling the guilty that they actively "pre-program" the subjects to intentionally produce false positive or false negative responses. In a modification of this complaint, the critics assert that examiners are so incompetent and weak-minded as to "unconsciously" overlook the actual responses on the charts in deference to spontaneous conclusions based solely on a desire for self-serving outcomes.

While some of these criticisms are undoubtedly motivated by a wish to destroy polygraph without regard to credence or fact, in my opinion most of the confusion over the concept of global evaluation is a result of both misunderstandings and misapplication of the term "global" and the procedures followed in global evaluations as it is practiced within the profession. (Lykken, 1981)

English & English, in their text, A Comprehensive Dictionary of Psychology and Psychoanalytic Terms define global as: "taken as a whole without attempt to distinguish separate parts or functions."

By this definition, I know of no examiner using a "global" approach to either chart interpretation or case resolution. One of the fundamental reasons why Reid & Inbau divided their original text into two separate books, one on polygraph and the other on interrogation, was to make a clear and distinct separation between the quantifiable, scientific aspects of polygraph and the art and skill of structured, professional interrogation. (Reid & Inbau, 1967 & 1977)

If, however, universal definition of "global" is used, meaning: comprehensive, entire or total, then the term "global" might be appropriate in describing the procedures followed.

---

*Stanley M. Slowik is the Chief Executive Officer of John Reid & Associates. He is a past contributor to Polygraph and is the Director of Reid & Associates Denver Office.

A more appropriate and accurate term for this procedure might be inductive case analysis in which one goes from the particular to the general, i.e., analyze separate variables and reach an overall conclusion regarding a particular relevant question or the issue under investigation. However, just as the psychological profession has caused confusion by improperly describing inductive procedures with the psychological definition of "global," I doubt seriously if they would accept the term inductive case analysis as an appropriate description of this practice.

Therefore, if the polygraph community is to be saddled with the term "global evaluation," I wish to make it eminently clear that I am hereafter referring to comprehensive case resolution and not the narrow psychological definition of global.

Global Evaluation is, therefore, not a general circumspection of a set of polygraph charts with no response comparison between relevant and control questions. In fact, one of the primary purposes of Reid's development of control questions was to provide just such an objective basis for comparing responses between relevant and control questions.

Global Evaluation does not permit an examiner to allow his investigative experience, behavioral observations or intuition to supersede or ignore the physiological recordings present on a set of polygraph charts, therein utilizing the instrument as nothing more than a psychological prop for interrogation. The American Polygraph Association has long stood in the forefront of those opposing such an application of polygraph instrumentation. (APA Constitution and By-Laws)

Global Evaluation is, in essence, a diagnostic approach to a single, paramount goal: determining the subject's truthfulness to the issue under investigation. In reaching this goal, Global Evaluation utilizes four components in a pre-planned, structured and analytical method. The four components are:

1. Fact Analysis
2. Behavioral Observation and Analysis
3. Chart Interpretation
4. Post-test Interrogation

While each of these components, in turn, have varying degrees of importance and benefit depending on the unique circumstances surrounding each individual polygraph subject and particular case under investigation, only under the most extreme circumstances should one ever allow Fact Analysis, Behavioral Observation or Post-Test Interrogation to influence on opinion based on chart analysis alone. The extreme case I speak of essentially involves the situation of false positive or false negative chart analysis. These situations most commonly involve circumstances in which the examiner's analysis of the case facts and subject behavior symptoms contradict his interpretation of the charts and persuades him to engage in a post-test interrogation. This interrogation may subsequently lead to a confession which, in turn, may be collaborated by outside physical evidence. Obviously, therefore, one of the primary benefits of Global Evaluation is to reduce or eliminate false positive and false negative errors and improve the overall accuracy of the polygraph process. In the case of the unsuccessful post-test interrogation, the examiner does not render an

opinion contrary to the charts, but may, again, only where appropriate, report the examination as inconclusive and schedule a re-examination.

Central to the process of Global Evaluation, therefore, must be a clear understanding of the four components previously mentioned, the procedures applied and the relative weight given these functions in arriving at an overall case conclusion. Let me begin with a discussion of Fact Analysis.

FACT ANALYSIS

All polygraph examinations should begin with the examiner first receiving some information regarding the issue under investigation. This information may take the form of statements made to various parties by the subject, witnesses to the event and/or the victim. It may include reports from investigators, court transcripts, laboratory reports, pictures, sketches or physical evidence. The examiner's review of this material constitutes Fact Analysis. We would, of course, prefer that a goodly portion of this material come from sources other than the polygraph subject.

Facts are analyzed for a variety of purposes, only one of which directly involves an assessment of the subject's truthfulness to the issue. Briefly, facts are reviewed to:

1. Determine if the issue to be investigated is suitable for the polygraph technique.

2. Determine if the subject is mentally, emotionally, physically or linguistically suitable for testing.

3. Assist in question formulation.

4. Assist in development of post-test interrogatory themes.

5. Determine if sufficient information exists to conduct an examination, and, finally

6. To assess the subject's truthfulness to the issue under investigation.

It is this last point that is often misinterpreted and can lead to charges that examiners "pre-determine" the outcomes of their tests and ignore chart interpretation entirely.

Those examiners with backgrounds in formal criminal investigation field procedures need not be reminded that the vast majority of criminal cases today are disposed of primarily on the basis of the investigation alone, without any assist from polygraph. While some investigations result in unclear or ambiguous findings and, yes, some with erroneous conclusions, experienced, professional investigators, such as those attached to law enforcement and governmental agencies, are correct in identifying the guilty party in most cases. The investitagors use the same data that I suggest be considered in Fact Analysis.

Fact Analysis, like chart interpretation, fingerprint comparisons, medical testing and many other endeavors associated with criminal investigation, relies on the probabilities of events occurring in relation to the suspect's innnocence or guilt. There are no absolutes. If your investigation conclusively indicates that a certain suspect was, in fact, incarcerated at the time a shooting took place, the probabilities are strongly in favor of the suspect not having pulled the trigger. Of course, it is possible that he could have escaped from his cell, committed the crime, and returned to the jail undetected, but the probabilities are clearly indicative of the suspect's innocence in the actual shooting.

The significance of any item in Fact Analysis and the degree or weight accorded it, must, therefore, be dependent on the unique circumstances of each issue under investigation. As a general principle, Fact Analysis should consider the facts in their totality. Rarely, does any single piece of information carry such meaning as to supersede the relevance of all other information.

The topic of this paper does not allow me to adequately detail the entire process of Fact Analysis, but let me offer a brief example for your consideration.

In the case of an internal embezzlement of $5,000.00 from a teller's drawer:

1. What is the suspect's position at the bank, i.e., what access did the suspect have to the place of disappearance?

2. Was the subject working the day/time of the disappearance?

3. Obviously, the greater the distance between the suspect in time or space from the disappearance, the more probable his or her non-involvement. Is there any physical or testimonial evidence linking the suspect to the scene?

The suspect's fingerprints found in an area where he is not authorized to be present, film of the suspect present at a time when he denies being there and, to a lesser degree, witness's testimony contradicting the suspect's statements, all increase (but do not necessarily prove) the probability of the subject's untruthfulness in denying the crime.

4. Does the suspect have a past history of involvement in similar crimes? While each case must rest on its own merits, the habitual offender does exist. Just as the alcoholic has a propensity to fall off the wagon, so too does the likelihood of an individual with a chronic history of theft increase the probability of involvement in the case at hand.

5. Has the suspect confessed, then retracted his statement? Although the media would sometimes have you believe that most employers and investigators, as a matter of policy, bludgeon false confessions out of innocent parties, with few exceptions, I have yet to see many innocent people admitting to crimes they did commit.

As you might surmise at this point, Fact Analysis is usually only as helpful as the quality of the investigation that precedes the polygraph

examination. If the investigation was conducted superficially, too quickly, or the case itself does not lend itself to this approach, very little weight can be accorded any determinations resulting from the analysis. Likewise, the examiner must also consider the quality and the source of the information. Most examiners have read autopsies that were later shown to be incorrect as to the cause of death. Statements made by witnesses can be deliberate fabrications or honest mistakes, but erroneous nonetheless. And certainly, facts that come only from the suspect's Defense Attorney must be viewed differently than facts submitted by an independent, professional investigative agency.

In its most simple form, Fact Analysis is practiced by every prosecutor considering charges against a suspect, by every field investigator seeking direction in his investigation, by most courts in determining the preponderence of evidence. The polygraph examiner should, therefore, consider the same data, but only as an adjunct to his final determination. In most cases, the Fact Analysis will conform to the examiner's conclusions in chart analysis. When they don't, the examiner should proceed cautiously and be aware that the case at hand demands additional scrutiny.

BEHAVIORAL ANALYSIS

The next procedure one might consider in inductive case analysis or comprehensive, global evaluations is Behavioral Analysis.

A physician may use a urine-pregnancy test as an indirect measure of fetal growth. The test will be a false negative if given too soon after conception, and a false positive if the patient is taking certain medications. The physician, therefore, must consider factors other than the test results alone in reaching a conclusion. Likewise, verbal and nonverbal subject behavior, particularly as exhibited during the pre-test interview, can provide invaluable information not only in determining the subject's physical, mental and emotional suitability for testing, but also provide useful information in determining the suspect's truthfulness to the issue under investigation. Observed Behavior, is, of course, a fundamental building block in Clinical Psychology, but in polygraph, it is much more specialized to concentrate on behavior related primarily to truth and deception.

As with Fact Analysis, Behavioral Analysis dictates a "cluster" approach to evaluation, i.e., that no single observation in and of itself carries great weight, but the totality of the observations may often provide useful clues and cross-checks in determining truth or deception.

As the Horvath study (1973) clearly indicates, even verbal statements alone can be statistically significant indicators of truth or deception, provided they are elicited in a structured technique and evaluated by properly trained and experienced examiners. One Utah study also indicates that when the examiner has little or no formal training in structured Behavioral Analysis techniques or little field experience in polygraph behavioral observation, the statistical significance of Behavioral Analysis is reduced. (Raskin, Barland, Podlesny, 1978)

As this paper is not intended to be a lengthy discourse on Behavioral Analyses, which most polygraph schools teach over an extended period, let

me briefly return to the example of the $5,000.00 missing from the teller's drawer, omitting the non-verbal and unsolicited verbal statements of the suspect. Again, let me reaffirm that I am dealing with the probabilities that a given suspect response indicates truth or deception, not the certainty that any given reply absolutely indicates truth or deception.

If, during the pre-test interview, after asking the subject a few innocuous questions such as the suspect's name or address, the examiner deliberately establishes eye contact with the suspect and clearly explains that the purpose of today's test deals with the disappearance of $5,000.00 from a certain teller's drawer, and if the suspect is responsible for this disappearance, the polygraph will clearly show this, the examiner then directly and unequivocably asks, "Did you steal that missing $5,000.00?" to which the suspect responds, "To the best of my knowledge, I don't think I did, as far as I can recall, so help me, God." or "Why do you want to know, you gravy sucking pig?" or "Do I have to answer that?", such responses are far more typical of one involved in the theft of the $5,000 than the innocent who most likely will respond with a calm, unequivocal, "No."

If the suspect is asked who he suspects of having stolen the $5,000 and responds by giving specific names and a reasonable justification for a suspicion, such an answer is more commonly heard from the truthful. The untruthful is more apt to respond by providing no direction for the investigation, e.g., "I can't say - it could be anyone," or attempts to lead the investigator away from the true suspect as in, "I'm sure it couldn't be anyone who works at this bank."

One of the most productive behavior eliciting questions deals with the punitive aspects of the crime being investigated, e.g., "Suppose someone here at the bank actually did steal that missing $5,000 -- what do you think should happen to the person who stole that money?" Truthful suspects are usually strong on punishment, "Well, they certainly should give it back and be fired." or "They should be arrested and prosecuted." whereas the guilty party answers in terms of what he'd accept happening to him if he is detected, "Well, I don't know -- maybe you should find out why he took the money." or "I can't say ... I don't want to judge anyone."

Obviously, as practioners of behavior analysis will tell you, there are many occasions wherein the subjects verbal and non-verbal responses are too ambiguous to interpet: an analogy to inconclusive polygraph charts. In such cases, behavior will be of no use in confirming a chart diagnosis, but as an aside, I can experientially relate that ambiguous behavioral analyses most frequently coincide with inconclusive polygraph records, which is probably consistent with the direct relationship between the observed behavior and the physiological tracings both originating from the same psychological stimuli. In fact, many non-verbal behavioral observations are sympathetic responses occurring in parameters different but simultaneous with those recorded by the polygraph.

CHART INTERPRETATION

After Fact and Behavioral Analyses, global evaluation next considers the true basis of polygraph, the physiological tracings on the polygraph chart.

All things being equal, the polygraph chart analysis is far and away the most reliable and valid indicator of truth and deception. If, for some unknown reason, an examiner had to choose among Fact Analysis, Behavior or Charts as the sole source of reaching a diagnosis, I would strongly recommend the charts.

Why then bother with Fact Analysis or Behavior Symptoms? Simply because chart analysis is neither infallible nor always conclusive. By this, I do not mean to imply that if the charts are inconclusive, the examiner reports the subject truthful or untruthful on the basis of Facts or Behavior. Rather, if the Facts or Behavior indicate truthfulness, while the charts are inconclusive, the examiner may consider examining other likely suspects in the case before scheduling the inconclusive suspect for a re-examination. Or, in the case of Facts and Behavior indicating untruthfulness on the part of the suspect whose charts are inconclusive, a post-test interrogation of some sort might be considered.

Facts and Behavior Analyses can't change the physical reality of the charts, but often times can explain unexpected results.

In the case of the investigator who tells the examiner before the test that the suspect became ill several times the day before the test and then the examiner observes the subject to be pale and sickly, the resulting erratic and inconsistent responses observed on the charts may, in fact, be the results of illness and not a purposeful attempt to distort the records, a phenomenon normally associated with the untruthful.

As mentioned at the onset of this paper, the greatest benefit of global evaluation, i.e., consider the Facts and Behavior along with the charts, concerns the undesirable but real possibility of error in chart interpretation.

Whether the result of some physical or emotional aberration on the part of the polygraph subject which causes the subject to fail to conform to the normal psychological and physiological processes or some error or omission on the part of the examiner, perhaps in subject preparation or question formulation, some polygraph subjects produce charts in direct contradiction to the actual truth. While such errors remain a statistical minority, it should be the desire of any profession to continuously attempt to reduce and eliminate all such errors. Global evaluation can provide the additional information an examiner may need to identify the conditions under which such errors are more likely to occur. The following, therefore, is a theoretical discussion of possible sources of error and the suggested method of using Fact and Behavioral Analyses to identify the circumstances under which such errors may be present.

1. Errors resulting from some physical or emotional aberration on the part of the subject.

A lot of attention has been given by the media and some members of the psychological community to the possibility of errors resulting from subjects who might clinically be diagnosed as psychopaths. While the Raskin-Hare (1978) study conducted in British Columbia indicates that such is not the case, at least with regard to a control question technique in an experimental situation, the charges persist that psychopaths cause false

negatives or "beat" the test. These arguments, of course, assume that the psychological basis for response is limited to fear, conditioning or guilt alone, totally ignoring the emotional weight or psychological set theories long espoused by Reid and Backster or the Attention theory Raskin teaches in his short course in Utah. (Reid, 1962; Backster, 1972; Raskin, 1981)

In addition, it is widely held that various pharmaceuticals can cause false negatives (Waid, 1981) particularily if one holds that the total absence of any deceptive responses denotes truthfulness, a contention that both practitioners of Control and R & I techniques refute as a legitimate criteria for decision-making. Fact Analysis can forewarn the examiner if the subject has a history of mental disorders, has been acting in an emotionally erratic manner or has a problem with drug or alcohol dependency. Behavioral Analyses may confirm the "red flag" identified in Fact review and further determine if such conditions are present at the time of the examination. If, in such cases the Fact and Behavior Analyses are consistent with the chart interpretation in terms of the suspect's truthfulness, the probabilities are strongly in favor of the examiner's opinion being correct. If, however, in the same circumstances the Fact and Behavior Analyses oppose the conclusion reached on chart interpretation alone, the possibility of the charts being in error increases. Depending on the degree of the contradiction, in some such cases, it may be advisable to report the subject as unsuitable for the polygraph technique, i.e., render no opinion, rather than risk an erroneous conclusion based on chart interpretation alone.

2. Errors resulting from oversights or omissions on the part of the examiner. In my experience, false positive errors in which truthful subjects are reported as being deceptive, most often occur when the examiner fails to establish proper psychological set between relevant and control questions, uses improper terminology in question formulation, fails to deal adequately with subject anger or hostility, or fails to deal properly with outside issues or the "fear of false positives" itself.

Control questions should be chosen so as to be psychologically equivalent to the relevant questions in their propensity to evoke deceptive responses. The subject's actual involvement or lack thereof, to the issue under investigation, should then determine the degree of response between the two types of questions. If the control questions are too "narrow," i.e., too limited in time, scope or degree, in effect severely limiting the possibility of an innocent subject being deceptive to the controls, false positives result. If the examiner, in the formulation of the relevant questions, uses terminology that is too emotionally provocative, i.e., words such as murder or rape, a false positive may occur. If the examiner fails to dissipate unusual subject anger or hostility in the pretest, errors can occur against the innocent. If the examiner fails to select control questions and develop them with the subject or, as in some techniques, include a question to deal with outside issues, both false positive and false negative errors can happen.

False negative errors occur when the examiner fails to adequately stimulate the subject, when the subject's motivation to respond is suppressed such as with career criminals or street gang members. As mentioned earlier, false negatives can happen if the examiner fails to deal with an overriding outside issue of great emotional significance to the

subject. Finally, there is Orne's (1978) contention that the "friendly" examiner encounters false negatives whenever he deals with a case for the defense. I cannot ascribe to this contention, as I feel the "problem" if you will, is again one of pre-test stimulation and easily overcome by the procedures for such cases. (Abrams, 1977)

In all of the situations just described, Fact and Behavioral analyses can greatly assist the examiner in proceeding cautiously when interpreting charts. In some cases, global evaluation not only alerts the examiner to the possibility of an increased probability for error, but may also suggest some adjustment in procedure on a re-examination to eliminate error.

POST-TEST INTERROGATION

The fourth and final element to global case resolution, post-test interrogation, does not technically fall within the realm of polygraph per se. But since many interrogations occur in such close proximity to the polygraph examination, a brief discussion of interrogation is warranted.

When the Fact and Behavioral Analyses indicate deception, the charts are inconclusive, and the situation allows it, a post-test interrogation can often resolve what otherwise would be reported as an inconclusive case. If the interrogation results in a valid and substantiated confession, we would suggest still recording the chart interpretation on the books as inconclusive, for that, in fact, is the case, but report the case as verified untruthful, citing the interrogation as the basis for this decision.

If the Fact and Behavioral Analyses indicate truthfulness, but the charts clearly indicate deception, I suggest a post-test interrogation. If a valid and substantiated confession results, the subject is reported deceptive on both the basis of the charts and the interrogation. But, if the interrogation is unsuccessful and, to compound the situation, the subject exhibits behavior during the interrogation typical of truthful subjects, I would continue testing other subjects on the case and arrange to have the subject back for a re-examination. If a re-examination were not possible, I would carefully reassess the situation in terms of the probability of a false positive error and in an extreme case, decline to render any opinion of the truthfulness.

In cases where the Fact and Behavioral Analyses are ambiguous and the charts conclusive, the subject is reported out on the basis of the charts alone. We would, however, interrogate subjects, where appropriate, whose charts indicate deception and report the subject deceptive regardless if they confess or not.

Post-test interrogation in global evaluation, in essence, is a validating procedure to resolve both marginal and contradictory situations. When opinions are rendered, the basis of the opinion is clearly identified, be it charts, interrogation or both.

In summary, comprehensive Global Evaluation - the inductive approach to case analysis - is nothing more than plain, old common sense. Those vague, uneasy feelings that sometimes make a given case just not "feel" right, can be more than intuition if some program of structured Fact and

Behavioral Analyses is employed. These tools, in turn, serve as cross-checks to chart interpretation. Post-test interrogation, where permitted, is a validating device to confirm or refute diagnosis. This system, if used improperly, without formal training or supervised experience in application, can be a disaster, as examiners fail to identify a solid, structured basis for their conclusions. Taken as a whole, or globally, the inductive approach to case resolution serves to reduce inconclusive results and minimize errors.

References

Abrams, S. A Polygraph Handbook for Attorneys. Lexington, Massachusetts: D.C. Heath & Company, 1977. Pp. 62, 177-179.

Backster, C. "Testimony in Court Martial of Capt. Ernest Medina, re: My Lai Massacre, 1970/71." In Zimmerman, C. The Polygraph in Court. Auburndale, Mass: B.H.F. Printing, 1972.

Constitution and Bylaws, Standards and Principles of Practice, Section 12, American Polygraph Association.

Horvath, F.S. "Verbal and Non-Verbal Clues to Truth and Deception During Polygraph Examinations." Journal of Police Service and Administration. 1(2)(1973): 138-152.

Lykken, D. A Tremor in the Blood. New York: McGraw-Hill, 1981, pp. 32-33; 73-74.

Orne, M.T. "Implications of Laboratory Research for the Detection of Deception." Legal Admissibility of the Polygraph, Edited by Norman Ansley. Springfield, Illinois: C.C. Thomas, 1975, pp. 114-116.

Raskin, D.C. Lecture presented at the University of Utah, Department of Psychology, Salt Lake City, Utah, October 1981.

Raskin, D.C., Barland, G.H. and Podlesny, J.A. "Validity and Reliability of Detection of Deception." National Institute of Law Enforcement and Criminal Justice, June 1978, p. 10.

Raskin, D.C. and Hare, R.D. "Psychopathy and Detection of Deception in a Prison Population." Psychophysiology 15(1978): 126-136.

Reid, J.E. "The Emotionally Weighted Question in Lie-Detector Testing." Paper presented at the American Academy of Polygraph Examiners' Ninth Annual Meeting, Chicago, August 1962.

Reid, J.E. and Inbau, F.E. Criminal Interrogation and Confessions, 2d ed. Baltimore: Williams & Wilkins, 1967.

Reid, J.E. and Inbau, F.E. Truth and Deception, The Polygraph 'Lie Detector' Technique. Baltimore: Williams & Wilkins, 1977.

Waid, W.; Orne, E.C.; Cook, M.R., Orne, M.T. "Meprobamate Reduces Accuracy of Physiological Detection of Deception." Science 212(3 April 1981): 71-72.

EFFECTS OF BETA BLOCKING DRUGS ON THE POLYGRAPH DETECTION RATE: A PILOT STUDY

By

E. Elaad, G. Bonwitt, O. Eisenberg and I. Meytes*

Abstract

A pioneer study of the efficacy of the polygraph on subjects under the influence of Trasicor, a Beta Blocking agent, was conducted by the Scientific Interrogation Unit of the Israeli Police and the Department of Cardiology of the Ichilov Medical Center.

Twenty volunteers underwent a series of 5 POT tests each; half with Trasicor administered beforehand and half with a placebo. Detection rate was highest for the placebo group, while that of the Beta Blocking group was still greater than chance. The GSR channel was found to be the most efficient identification index, but is also most vulnerable to the effect of Trasicor. The respiration channel indicates identification at a level better than chance, which is not effected by the Beta Blocking agent. The cardio channel was not found an efficient index of identification.

The importance of investigating the influence of drugs on the polygraph efficacy relates to the growing belief, held by subjects, that using depressant drugs before the test will defeat the instrument and avoid incriminating results.

Unfortunately, there is a serious shortage of experimental data concerning the effects of drugs on the polygraph detection rate. The prevailing opinion among polygraphers is that barbiturates and other depressing drugs appear to reduce the polygraph efficacy (Berman, 1967, Hess, 1975).

The physiological responses recorded by the polygraph are somatic symptoms of autonomic activity influenced by adrenergic stimulation. Detecting changes in the autonomic activity, such as changes in skin resistance, changes in respiration and various cardiovascular preturbations, is the essence of the polygraph operation. The assumption is that those changes reflect the subjects fear of being found untruthful.

The Beta Adrenergic Blocking Drugs decrease the peripheral autonomic activity by blocking the effects of adrenaline at beta receptor sites

* E. Elaad and G. Bonwitt are with the Scientific Interrogation Unit, Criminal Identification Division, Israel Police Headquarters, Jerusalem, Israel. O. Eisenberg and I. Meytes are with the Department of Cardiology, Ichelov Hospital, Tel Aviv, Israel. Reprints are available from E. Elaad, at the Israel Police Headquarters, Criminal Identification Division, Jerusalem, Israel.

(Kelly, 1975). This interference with the autonomic activity may be detrimental to the polygraph detection rate.

The aim of the present study is, therefore, to explore the field of polygraph efficacy in detecting subjects' involvement with concealed information while being under the effect of Beta Blocking Drugs.

Method

Subjects: Twenty volunteers from the Department of Cardiology staff of Ichilov Hospital (nurses, physicians, technicians, etc.) agreed to participate in the experiment. Nine of them were males and eleven females. The mean age was 27.7 (SD=5.17). All subjects were in good health.

Examination paradigm: The Guilty Knowledge Technique (Lykken, 1974) was used in the experiment. This technique is based on the assumption that a strong response to an item reflects the relevancy of that item to the subject, thus a guilty subject will recognize the relevant item and react to it, while an innocent subject will respond in an equivalent manner to the critical as well as to the control items.

Apparatus: A Stoelting 22771 field polyscribe, operating on four channels, thoracic and abdominal respiration, Galvanic Skin Response (GSR) and Cardiovascular activity, was employed in the experiment. Thoracic and abdominal respiration recordings were obtained from pneumatic tube positioned around the upper thorax and abdomen. The GSR electrodes were placed on the second and fourth fingers of the left hand. The blood pressure cuff, which was attached to the right arm, was inflated to a pressure of 60 mm Hg.

Procedure: The experiment was conducted by three experimenters. Two were psychologists from the Scientific Interrogation Unit of the Israeli Police who had experience in polygraph examining. The third was a physician from the Department of Cardiology of Ichilov Medical Center where the experiment took place. Experimenter A (a psychologist) filled out a standard examination form for each subject including data on the subject's health (to ensure that the subject will not be endangered by the drug), five personal items such as name, surname, father's name, date of birth and place of birth were also registered on the form. For every relevant item, four similar irrelevant items were added by Experimenter A. The position of the critical items among the irrelevant ones varied at random. Experimenter B (a physician) gave each subject a yellow pill to ingest, taking care not to reveal any information about the nature of the drug. One group (5 males and 5 females) received 40 mg. Trasicor-80 (a Beta Blocking drug), the other group (4 males and 6 females) received the same dosage of placebo. About one hour later the subject went to the examination room and handed the form filled by Experimenter A to Experimenter C (the other psychologist) who seated the subject and attached him to the polygraph. The subject was then given recorded instructions telling him that he would be asked several questions. Each question presents five items, one critical item and four neutral ones. The subject was further told to deny any relation to both the critical and neutral items. The questions were asked in a fixed order for all subjects. Each set of items was repeated twice in two consecutive presentations. A 10 second interval between the presentation of different items was used. A buffer item was

placed in the beginning of each question to control the noise of the first response. Experimenter C, while conducting the polygraph test, didn't have any idea of which experimental treatment (Trasicor or placebo) the subject had received, nor had he any idea of which item was the critical one for the subject.

Results

A. Beta Blocking drugs were reported to decelerate pulse rate response (Gaind et al., 1975, Turner, 1975). In order to ensure the construct validity of the Trasicor manipulation, the subjects' pulse rate was determined at two different time periods. The rate was determined initially immediately after presenting the first item of the first question, and the second determination was immediately after presenting the first item of the fifth question. Pulse rate was measured for a .15 second period and then transformed to beats per minute. The results are presented in Table 1.

Table 1

Mean Pulse Rate (Beats per Minute) for Two Drug Conditions Varying in Time Periods

Time Periods	Drug Conditions		Marginal Mean
	Placebo	Trasicor	
First Question	74.4	63.2	68.8
Fifth Question	76.4	65.6	71.0
Marginal Mean	75.4	64.4	69.9

A two-way analysis of variance with repeated measures (Winer, 1971) carried out on heart rate data, revealed a significant main effect for the drug conditions ($F=10.9, df=1/18, p < .01$) and a significant main effect for periods ($F=5.21, df=1/18, p < .05$). No significant interaction effect.

The results indicate that the average pulse rate of the BB (Beta Blocking) group was found to be slower than that of the placebo group over all the test periods.

B. BB effect on the polygraph detection rate: Three independent polygraph examiners were asked to rank order the physiological responses to the items within each question. The responses were ranked from 1 (the strongest response) to 5 (the weakest response). The examiners scored the response ranks from the polygraph chart which presented all four response channels. The total response score was the sum of rank scores for each response, so that examiner agreement on the strongest response produced a score of 3. Agreement on the weakest response resulted in a score of 15.

Using the normal approximation to the binomial distribution, the detection rates of both the placebo and the BB groups were significantly different from chance at the .001 level. ($z=43.8$, $z=37.5$ for the placebo and BB groups respectively.)

A more recent approach claims that a signal detection paradigm is most appropriate for this kind of data analysis (Ben Shakar et al., 1970). This model has the advantage of providing an index of the polygraph detection rate which is based on the comparison of all the distributions of responses to both the relevant and neutral items (Lieblich et al., 1976). The model enables an evaluation of the price (in increased rate of false positive errors) for every marginal addition to the correct detection rate.

To create the polygraph efficacy index, all total response scores were tabulated in accumulative frequency tables, both for the relevant and neutral items. Separate tables were produced for the BB group and for the placebo group. A sequence of 13 cutoff points relating to the 13 possible total scores of the three rankers (3-15) was used. Each point defines the cumulative percent of correct detections (Hit Rate) and the cumulative percent of false positive errors (False Alarm Rate). A Receiver Operating Characteristic (ROC) curve describing the sequence of all 13 cutoff points was drawn. The area under the ROC curve defines a detection efficacy index.

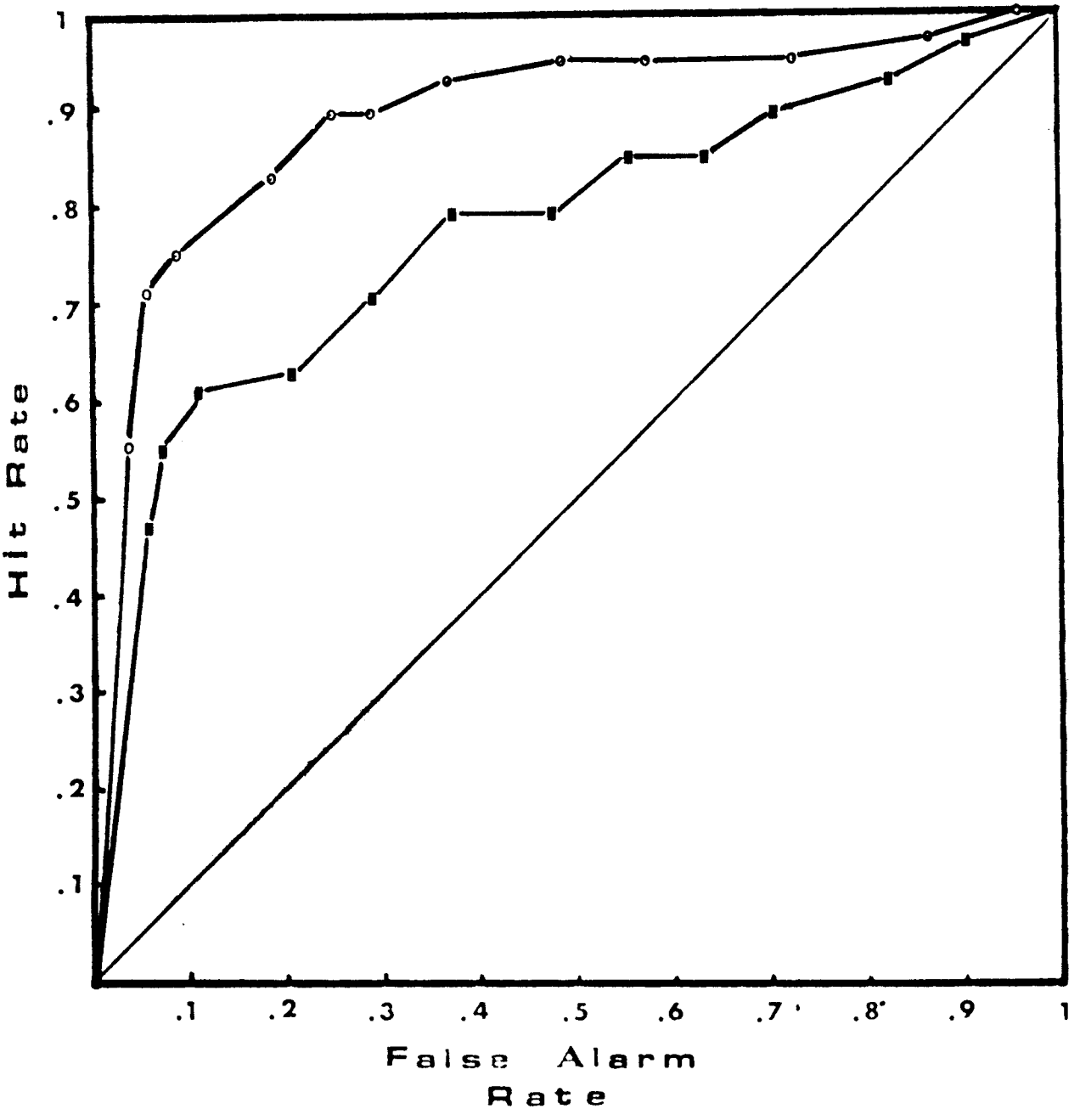
The diagonal in Figure 1 represents similar distribution of the hits and false alarms. Therefore, the area under the diagonal ($a=0.5$) represents an index for a random ranking of the responses. It is easy to see that the area under the ROC curve of the Trasicor group ($a=.785$) and the area under the ROC curve of the placebo group ($a=.899$) are considerably larger than that under the diagonal.

Bamber (1975) suggested a method with which the variance of the area under the ROC curve can be estimated and confidence intervals can be produced. Using Bamber's method, implemented to detection of information by Ben Shakhar (1977), a 99% confidence interval for the area in each of the drug conditions was computed. The confidence intervals for the placebo and BB groups were .828 - .970 and .670 - .890, respectively indicating that the difference in the detection rate between the two groups is significant ($p < .01$).

C. The distinctive efficacy of the different physiological channels: After the multiple channel recording was judged, the recordings of each physiological channel were separated (the respiration channels were united) and were given, separately, to three independent polygraph examiners for evaluation. The examiners rank ordered the responses in the same manner described earlier. To control for ranked effect, the area under the ROC curves of each examiner was computed. There were significant differences between rankers in evaluating each of the three channels.

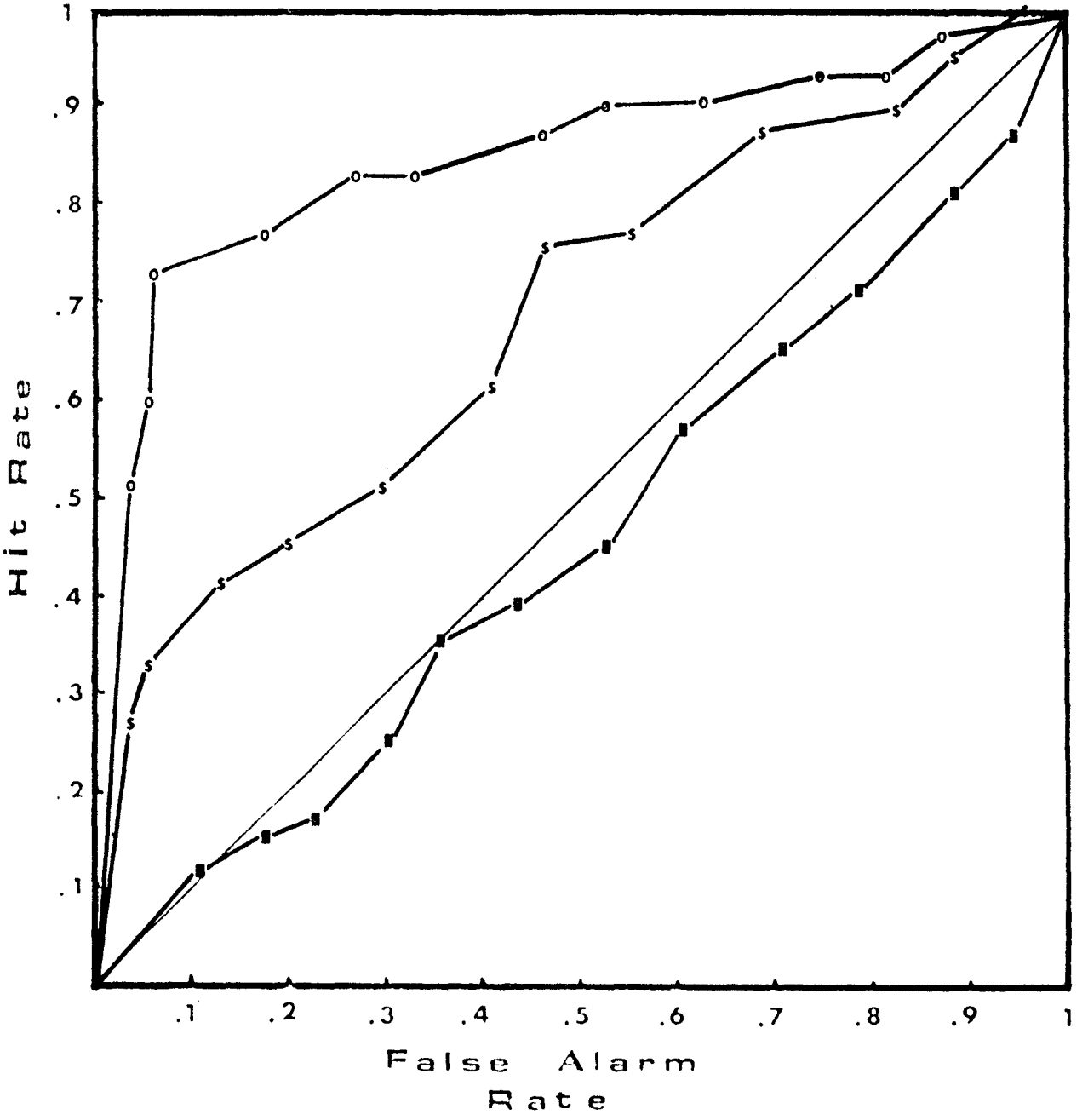
In order to determine the value of each physiological channel in detecting concealed information, a 99% confidence interval for each channel (only for the placebo group) was computed. The intervals were .763 - .945 for GSR, .567 - .804 for respiration and .341 - .585 for cardiovascular activity. The results indicate that the detection rate of the GSR and

Figure 1
Receiver Operating Characteristic Curves for Trasicor and Placebo samples across all response channels. (Circles=Placebo sample, Squares=Trasicor sample.)



respiration are significantly better than chance while the detection rate of the cardio is not. However, the GSR detection rate is significantly better ($p < .01$) than that of respiration. The results are well demonstrated in Figure 2.

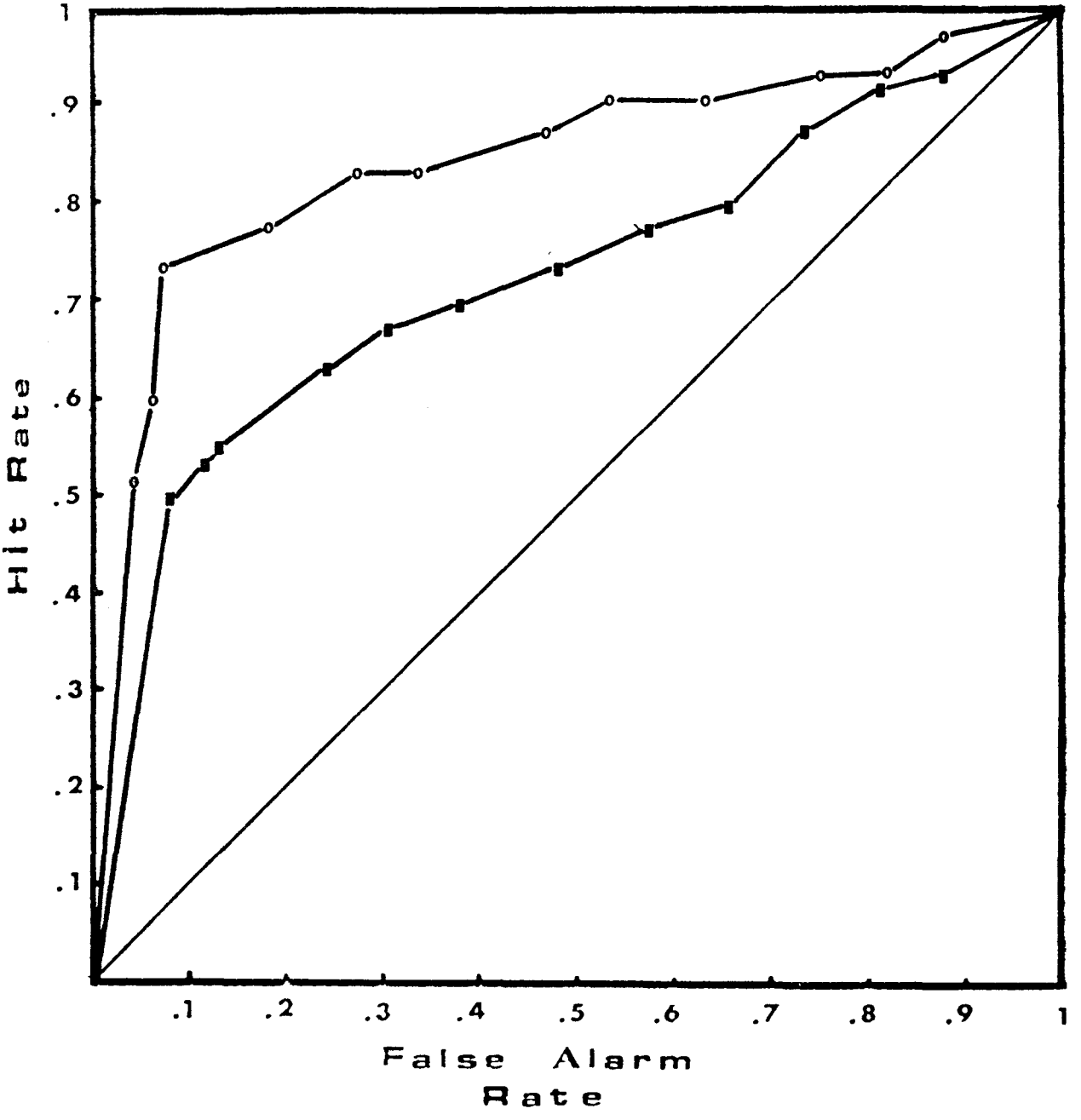
Figure 2
Receiver Operating Characteristic Curves for Each Response Channel-Placebo sample. (circles=GSR, dollars=Respiration, squares=Cardio.)



D. Trasicor effect on the detection rate of the physiological channels:

GSR: To determine BB effect on the GSR detection rate, the area under the ROC curves for both the trasicor and placebo groups were computed. The outcomes were .732 for the Trasicor group and .854 for the placebo group. A 99% confidence interval for each drug group was computed and was found to be .615 - .849 and .763 - .945 for the BB and placebo groups respectively. The results indicate that the difference is significant ($p < .01$). (See also Figure 3.)

Figure 3
Receiver Operating Characteristic Curves for Trasicor and Placebo samples while only the GSR was exposed to the rankers. (circles=Placebo sample, squares=Trasicor sample.)



Respiration: The area under the ROC curves for both the BB and placebo groups reflecting the separate evaluation of respiration, was found to be .746 and .690 respectively. The 99% confidence intervals were .643 - .849 and .576 - .804 respectively. The results indicate that the difference is not significant.

Cardio: The area under the ROC curves for each drug group in the evaluation of the cardio channel was found to be .485 (Trasicor) and .463 (placebo). The 99% confidence intervals were .371 - .600 and .341 - .585 respectively. The difference is not significant.

Discussion

The purpose of the present study was to determine the effects of Beta Blocking drugs on the polygraph detection rate. It was found that the dosage of 'Trasicor 80' given to the subjects was enough to reduce their pulse rate during the entirety of the test session. This result enables the attribution of other effects found to Trasicor influence.

It was further found that Trasicor decreases the polygraph detection rate performed in the guilty knowledge technique. A closer inspection of the results reveals that Figure 1 and Figure 3 present most similar patterns of ROC curves, indicating that the GSR had the greatest influence on the rankers decision when all three response channels were exposed to them. Furthermore, all three rankers agreed with that inference. It is argued, therefore, that the reduction of the polygraph detection rate was caused mainly by the strong influence of the BB on the skin resistence responses which proved to be the most efficient channel when a detection technique, based mainly on orienting responses, was employed.

It is interesting to point out that Hess (1975), on the basis of his experience as a field polygraph examiner, claimed that most drugs have a more dramatic effect on the skin resistance tracing than on other channels. The present study, even though it is limited to Trasicor, lends some support to this argument.

The results of the separate evaluation of respiration recordings revealed that its detection exceeds chance level and must not be neglected in the guilty knowledge technique. Furthermore, respiration seemed to improve the overall detection rate especially because skin resistance responses have the quality of rapid habituation. The negative results concerning the detection rate of the cardiovascular activity can be explained by the fact that, in general, they show a longer latency to stimulation.

All in all it can be concluded that skin resistance responses are most vulnerable to BB influences while respiration recordings are not effected by the drug. Following that conclusion it can be claimed that using Beta Adrenergic Blocking drugs may be detrimental to polygraph techniques which emphasizes the GSR responses (guilty knowledge technique) while the control question methods, which put more attention to changes in respiration, might be less effected.

Finally, two limitations must be considered before the results can be applied to the field situation.

a. The subjects who participated in the study were not in a strong emotional state which is typical of the field situation.

b. The sample of subjects in the present study is quite different in SES factors from the typical population in the field.

For these two reasons the present study should be considered as a pilot study which must be replicated using a larger sample of subjects, more closely approximating the field population and examined under more stressful conditions.

Acknowledgement

The authors wish to thank Dr. G. Ben Shakhar for his assistance in the methodological aspects of the study, Ms. Yael Deshe, M. Kleiner and M. Amit for their assistance in evaluating the polygraph recordings.

References

Bamber, D., 1975. "The Area Above the Ordinal Dominance Graph and the Area Below the Receiving Operating Characteristic Graph." Journal of Mathematical Psychology, 12, 387-415.

Ben Shakhar, G., Lieblich, I., and Kugelmass, S., 1970. "Guilty Knowledge Technique: Application of Signal Detection Measures." Journal of Applied Psychology, 54, 409-418.

Ben Shakhar, G., 1977. "A Further Study of the Dichotomization Theory in Detection of Information." Psychophysiology, 14, 408-413.

Berman, M.A., 1967. "Drugs vs. the Polygraph." Journal of Polygraph Studies, 1(4): 1-3.

Gaind, R., Suri, A.K., and Thompson, J., 1975. "Use of Beta Blockers as an Adjunct in Behavioral Techniques." Scott, Medical Journal, 20, 284-286.

Hess, C., 1975. "Observations Regarding the Effects of Specific Drugs on Polygraph Tracing." Polygraph 4, 314-328.

Kelly, D., 1975. "Report of a Symposium on Beta Adrenergic Receptor Blockade in Psychiatry (Leading Article)." Scott, Medical Journal, 20, 279-280.

Lieblich, I., Ben Shakhar, G., and Kugelmass, S., 1976. "Validity of the Guilty Knowledge Technique in a Prisoner's Sample." Journal of Applied Psychology, 61, 89-93.

Lykken, D.T., 1974. "Psychology and the Lie Detection Industry." American Psychologist, 29, 725-739.

Turner, P., 1975. "Some Studies of Beta Receptor Blocking Drugs on the Central Nervous System in Man." Scott, Medical Journal, 20, 290-291.

Winer, B.J., 1971. Statistical Principles in Experimental Design, 2d ed., New York: McGraw-Hill.

Psychophysiological Detection of Deception in Japan:

The Past and the Present

By

Junichi Fukumoto

Japan is one of the most advanced users of the polygraph in the world, along with the United States, Canada and Israel. The polygraph has played an important role in scientific identification, and the investigator has used the polygraph to solve many brutal crimes. When one leafs through a book on police science, one will certainly find several pages devoted to the polygraph. In Japan, the results of polygraph testing have frequently been admitted as evidence in court. However, in Japan its application is not as widespread as described by Lykken (1974), for it is used solely in criminal investigations.

It is little known that Japanese psychologists devised an original method of using electrodermal activity (EDA) in the 1930's. Following World War II, Japanese police authorities made a great effort to make practical use of the polygraph for the detection of deception, as used in America. However, because of significant legal and social differences between Japan and America, the Japanese use of the polygraph is remarkably different. Nonetheless, the authorities concerned with the polygraph had to be sensitive to the trends in the United States. In an effort to further exchange information on the use of the polygraph, I shall discuss the history and the circumstances surrounding the use of the polygraph technique in Japan.

Pioneers in Detection of Deception Research

The search for psychophysiological methods of detection of deception was begun in Japan by psychologists at universities. During the 1920's, several psychologists were keenly interested in EDA as an indicator of emotions. Akamatsu, Uchida & Togawa (1933) suggested the use of EDA for the detection of deception, because the skin conductance level (SCL) was found to decrease during emotional changes. The first report of psychophysiological detection of deception in Japan was by Akamatsu et al., (1937), when they reported good success using SCL. In 1938 and 1939 they reported additional experiments using SCL for the detection of deception. Togawa later applied this method for testing a spy, in Japan's first practical application of psychophysiological detection of deception.

As it did with other fields of psychology, World War II interrupted lie detection activities. When the war was over the Japanese police became interested in lie detection as a means of scientific investigation. Following a suggestion by Togawa the National Institute of Police Science (NIPS) endeavored to establish lie detection methodology. The case reports

---

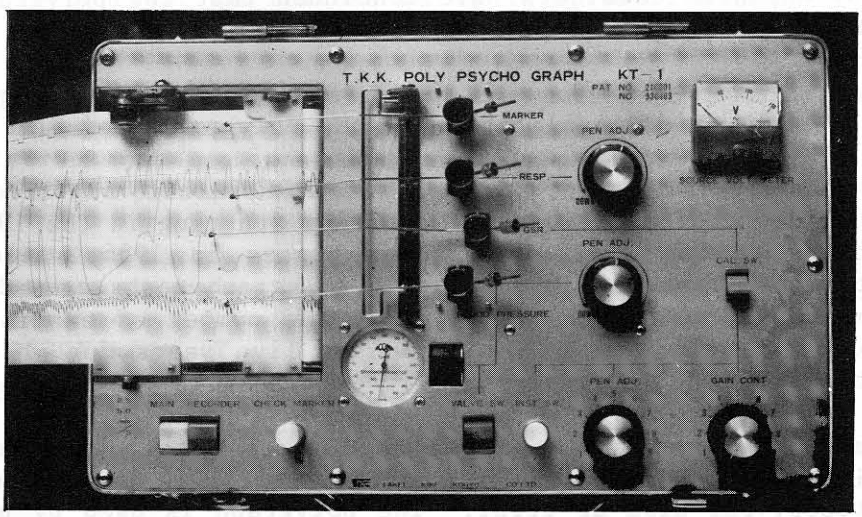
Requests for reprints should be addressed to the author at: Scientific Investigation Research Laboratory, Yamaguchi Prefectural Police Headquarters, 1-2 Takimachi, Yamaguchi-shi, Yamaguchi, Japan.

were published in the reports of the National Police Agency (NPA). Iijimi (1950, 1951) published general reviews of the use of the polygraph in America, from the point of view of its use as an interrogative technique. In 1951 Imamura was the first Japanese to be trained in the Keeler-type polygraph technique at the Far East Criminal Investigation Laboratory under the auspices of the U.S. occupation forces. The next year he introduced the use of the polygraph in lie detection. Immamura and his colleagues made an effort to promote the spread of the polygraph in Japan, based upon the methods used in America. Some of their more significant contributions include the comparison of EDA measurement versus a polygraphic methods (Yamashita, 1955), the comparison of various physiological measures (Imamura, Yamashita, Suzuki & Yamaoka, 1960), and the construction of a new question method. These systematic studies by the NIPS researchers made an important contribution for the conversion from EDA measurement to the polygraphic method. There were also some studies reported by university researchers in the early post-war years (e.g., Togawa, Soma & Mochizuki, 1950; Uruno & Narai, 1956), but university professors lost interest in the detection of deception. Consequently, in Japan the study of lie detection has been entrusted to the police, centering on the NIPS.

The Development of the Apparatus

The original apparatus used in the 1930's was a galvanometer which measured SCL changes. In 1943 the Yokokawa Denki Company put a psychogalvanometer on the market which was adopted in 1947 as a lie detector by the Metropolitan Police. This device spread throughout the country's police departments, for the police had just been required to apply more scientific, advanced investigative methods. However, it did not work out well, because the operation of the apparatus was too complicated. It was therefore replaced by a polygraph. The Keeler model 302 polygraph was first introduced to Japan in 1953, and served as a pattern for two domestically produced polygraphs introduced in 1955: the YKK Polygraph, manufactured by the Yamakoshi Seisakusho Company, and the TKK polygraph, made by the Takei Kikikogyo Company. These Japanese polygraphs were gradually adopted at all prefectural police headquarters.

Figure 1
The Polygraph Instrument Used in Japan



The KT-1 Polypsychograph shown in Figure 1 is manufactured by the Takei Kikikogyo Company and consists of a single pneumograph, skin resistance response with self center, a cardiograph, and an event marker. The Japanese polygraphs are unified and standardized, but some Japanese psychophysiologists have criticized the polygraphs used by the police, claiming that this standardization has caused stagnation and stifled improvements and the study of other physiological measures. We should be impressed by the observation, "... polygraph operators are using the equivalent of trains in a rocket age." (Sternbach, Gustafson & Colier, 1962).

Qualifications of the Examiner

Since 1953 police officers and technical officers have been trained as polygraph examiners at the NIPS, which is the only polygraph training course in Japan. In 1972 the Criminal Investigation Bureau of the NPA published the "Polygraph Examination Guideline" in an attempt to set suitable standards. Polygraph examiner trainees are now selected from those university graduates majoring in psychology or other related technical fields. About 100 polygraph examiners have been assigned to all identification sections and crime laboratories in Japan. This has been described in more detail by Takahashi (1976). Some psychologists have advocated higher standards for qualifying examiners, as there is no public license system in the field of psychology in Japan.

Types of Tests Used

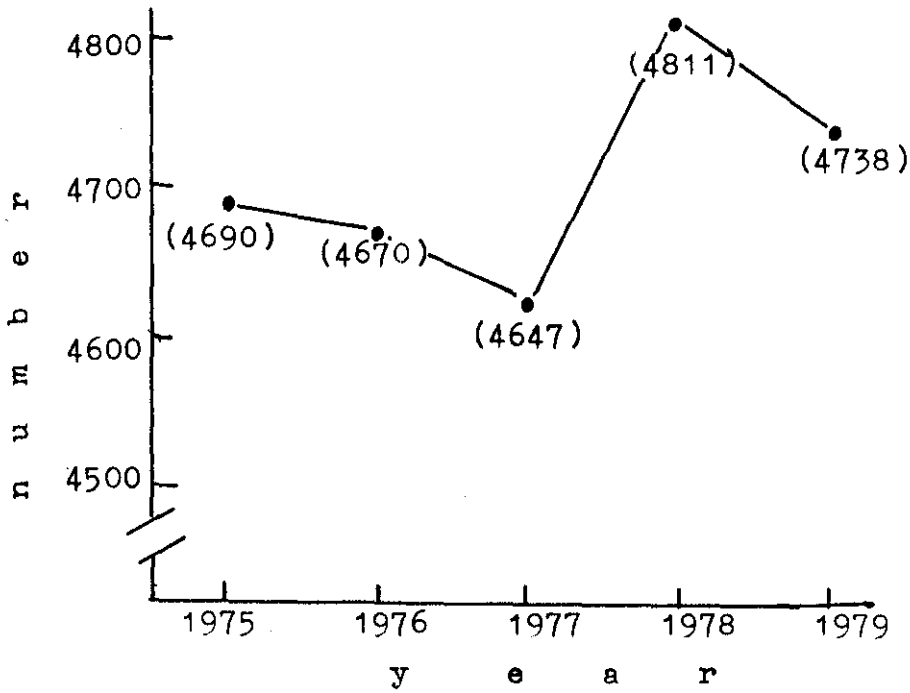
In 1937 Akamatsu and his colleagues recommended a type of Hidden Key question technique. When the polygraphic method was introduced to Japan, the control question test (CQT), peak of tension test (POT), and zone comparison test (ZCT) were also introduced. The NIPS strived to establish that method which best fit the polygraph situation in Japan. Through examiners' experiences, the CQT and POT tests were retained, based on the validity of the test construction. At the present time, the POT test is generally considered the test of preference, with the CQT test serving as a supplement. This is based upon the more reasonable theoretical foundation of the POT test. I am one of the polygraphers who have doubts about the theoretical basis of the CQT (Suzuki, 1968; Onishi, Ota & Tanaka, 1967). Many Japanese examiners have concluded that the polygraph technique is not "lie detection" per se, but is guilty knowledge detection (Lykken, 1960).

Field Application of the Polygraph

As previously mentioned, the polygraphic method superceded single channel measurement of electrodermal activity starting in the 1950's. The polygraphic method was applied to actual cases in 1956. According to Identification News, about 4,600 to 4,800 polygraph tests are administered in Japan annually. As shown in Figure 2, the number of examinations has remained comparatively stable over the last five years for which figures are available.

The use of the polygraph has made a great contribution not only in the investigation of such crimes as thefts, burglaries, arson, murder, and traffic matters, but also at trial. The polygraph is used by police, the

Figure 2
The Number of Examinees



Inspection Bureau of the Postal Service, and the Self Defense Force. However, no details of its use by nonpolice organs have been published. Although there are no regulations restricting commercial use of the polygraph, the application of the polygraph has been limited to official investigative organs. This phenomenon is conspicuously different from the situation in the United States.

Conclusion

In Japan, the detection of deception originally developed on the basis of electrodermal activity recorded as the sole channel. After World War II, the polygraphic method was adopted, modeled upon that of the U.S. The polygraph test is considered as a psychological test or psychological procedure, in contrast to its use as an interrogative technique in America.

In Japan, we should be aware that the polygraph examination has some important associated topics. These include the legal use of the results of the polygraph examination, the way of thinking peculiar to the Japanese, the moral implications, etc. Details of these will be reported in the near future.

References

Akamatsu, P., Uchida, Y., & Togawa, Y. "The Measurement of PGR(1)." Phylosophia, 3, 230-241, 1933. (In Japanese)

Akamatsu, P., Uchida, Y., & Togawa, Y. "The Measurement of PGR(2)." Phylosophia, 7, 171-204, 1937. (In Japanese)

Akamatsu, P. & Togawa, Y. "The Application of PGR." Journal of Educational Psychology, 13-4, 311-314, 1938. (In Japanese)

Akamatsu, P., Uchida, Y., Togawa, Y. & Miyata, Y. "The Measurement of PGR(5)." Phylosophia, 9, 195-211, 1939. (In Japnaese)

Iijima, E. "A Criminal Investigation Technique." Reports of NIPS, 13, 40- 58, 1950. (In Japanese)

Iijima, E. "On Recent Various Scientific Lie Detection Methods and Their Evidence Problems." Reports of NIPS, 16, 132-154, 1951. (In Japanese)

Imamura, Y. "The Investigation Techniques Using the Psychogalvanometer." Reports of NIPS, 19, 124-132, 1952. (In Japanese)

Imamura, Y., Yamashita, S., Suzuki, S. & Yamaoka, K. "A Study of Physiological Responses in Lie Detection." Reports of NIPS, 51, 355- 364, 1960. (In Japanese)

Lykken, D.T. "The Validity of Guilty Knowledge Technique: The Effects of Faking." Journal of Applied Psychology, 44, 258-262, 1960.

Lykken, D.T. "Psychology and the Lie Detection Industry." American Psychologist, 725-738, 1974.

Onishi, K., Ota, J. & Tanaka, Y. "The Desirable State on the Polygraph Examination." Polygraph Reports, 45, 104-107, 1967. (In Japanese)

Psychology Room, Criminology Section. (Imamura, Y.) "Lie Detection Test and Peak of Tension Technique." Reports of NIPS, 32, 64-76, 1956. (In Japanese)

Sternbach, R.A., Gustafson, L.A. & Colier, R.L. "Don't Trust the Lie Detector." Harvard Business Review, 40-6, 127-134, 1962.

Suzuki, S. "A Theoretical Consideration to the Polygraph Examination." Polygraph Reports, 49, 24-26, 1968. (In Japanese)

Takahashi, Y. "The Police Operation of Polygraph Detection and Its Assessment From Judicial Standpoint in Japan." Polygraph, 5-3, 223-234, 1976.

Togawa, Y., Soma, H. & Mochizuki, K. "A Study on the Lie Detection by PGR." Journal of Japanese Psychology(an extra number), 1950. (In Japanese)

Uruno, F. "The Detection of Deception on the Statement Using PGR Method." Applied Psychology 16th Annual Meeting, 68, 1953. (In Japanese)

Uruno, F. & Narai, J. "An Experimental Study on Lie Detection." Reports of NIPS, 33, 117-129, 1956. (In Japanese)

Yamashita, S. "Lie Detection Technique and Actual Test Case by Keeler-Polygraph." Reports of NIPS, 28, 392-398, 1955. (In Japanese)

THE DICHOTOMIZATION THEORY FOR DIFFERENTIAL AUTONOMIC RESPONSIVITY RECONSIDERED

By

Gershon Ben-Shakhar and Israel Lieblich

Abstract

The effect of the serial position of a relevant (chosen) stimulus on the differential skin conductance response evoked by it was investigated. One hundred and three subjects were tested in four experimental conditions in which they were presented with a verbal sequence of 8 numbers, one of which they had previously chosen. It was found that differential responsivity decreased as a function of the serial position of the chosen number. This finding contrasts with a strict version of the dichotomization theory. A modified version of this theory was suggested to account for the present finding, as well as for previously reported results. It focuses on the different attributes of the stimuli. It states that habituation to a stimulus is dependent on the frequency with which attributes of the stimulus were presented either as components of the stimulus or as components of other stimuli.

A set of studies which compared autonomic responsivity to relevant and neutral stimuli established an empirical rule that differential autonomic responsivity to relevant stimuli is negatively related to their relative frequency (Ben-Shakhar, 1977; Ben-Shakhar, Lieblich & Kugelmass, 1982; Lieblich, Kugelmass & Ben Shakhar, 1970). The relevant stimuli in these studies were objects (e.g., numbered cards) chosen by the subjects. We proposed the dichotomization theory to account for this empirical rule. According to the theory, persons who choose certain (relevant) stimuli will manifest independent habituation processes to the two types of stimuli (relevant and neutral). This requires habituation to generalize within each stimulus category (i.e., relevant and neutral), but not across categories (i.e., from the relevant to the neutral, or vice versa). Hence, increasing the relative frequency of one category should result in more habituation of the stimuli comprising this category, and thus less relative responsivity to them. The generalization of habituation within categories was demonstrated by Ben-Shakhar (1977).

While in general the dichotomization theory described the data better than alternative explanations, certain anomalies were observed. For example, a strict dichotomization theory predicts that in a situation

---

This present study was supported by a grant from the Eshkol Institute, Faculty of Social Sciences, The Hebrew University of Jerusalem. We wish to thank Judy Lustig and Shlomo Bruk for their assistance.

Address requests for reprints to: Gershon Ben-Shakhar, Department of Psychology, The Hebrew University of Jerusalem, Jerusalem 91000, Israel.

where the relevant and neutral categories are equiprobable (e.g., .50), it would not be possible to differentiate between them using a psychophysiological measure. However, in most studies using such base rate conditions, the skin conductance response (SCR) evoked by a relevant stimulus was greater than the SCR to a neutral stimulus. In addition, a straight forward prediction from the dichotomization theory is that the differential autonomic responsivity will be a function of the relative frequencies of the two categories, but will not be dependent on which category is the more frequent. For example, if the relevant stimulus is presented with a relative frequency of 1/8 and the neutral stimulus with a relative frequency of 7/8, the theory predicts the same autonomic responsivity as observed in a situation in which the relative frequencies are reversed (i.e., neutral stimulus - 1/8, relevant stimulus - 7/8). However, both Ben-Shakhar et al. (1982) and Ben-Shakhar (1977) found that a rare relevant stimulus resulted in greater SCRs than a rare neutral stimulus.

In order to understand these anomalies while keeping within the framework of the theory, it would be useful to explore further the boundaries of the theory. For example, a strong test of the theory would be to examine the prediction that differential autonomic responsivity to a relevant stimulus is independent of its serial position. Let us assume several sequences of stimuli, each of which is comprised of one relevant and seven neutral stimuli. The serial position of the relevant stimulus can vary from 1 to 8. If the two types of stimuli habituate independently, then the autonomic responses should not be a function of the serial position. This prediction contrasts with that derived from an approach that views the habituation of the whole sequence as a unitary process. The present experiment was designed to explore the influence of the serial position of the relevant stimulus on the relative SCR evoked by it.

Method

Subjects

Thirty-five male and 68 female Hebrew University undergraduate students participated in the experiment. They were recruited through an ad and were paid for their participation.

Apparatus

A constant current (18μA) SCR apparatus (direct coupled mode) coupled to a Brush II recorder was used. The recording system was placed in a control room separated by a one-way mirror from the room in which the subject was tested. Beckman electrodes were used for skin resistance measurement. The current density was 31.75μA/cm^2. A Revox tape recorder was used to present questions to the subjects. A PDP 8/E computer was used to transform the highest change in resistance, within 5 sec from the end of each question, into a conductance change score.

Procedure

The subject was seated at a table facing a blank wall. Beckman electrodes were attached to the volar side of the index and fourth finger of the subject's left hand, using masking tape with pressure such that the subject felt comfortable. Beckman electrode paste which contained sodium

chloride, glycerin, tragacanth and benzyl alcohol .5% was used to ensure proper contact. All testing was carried out in an air-conditioned laboratory maintained at 26° C. At the end of a 2-min baseline recording, the subjects were instructed to choose a card from a set of four numbered cards placed before them, and to record their choice on a supplied form. Unknown to the subject, all four cards were identical. Subjects were then told that they would hear a series of questions about the card numbers. They were requested to sit quietly and listen to the questions, and not to answer any of the questions, even if they referred to the chosen card. The subjects then heard a series of 8 questions of the form: "Did you choose card number _____?" The series was comprised of the numbers 2 through 9, in random order. The whole procedure was repeated four times: each time the subject made a choice from a different set of four identical cards, and each time a different recorded question series was played. The interquestion intervals within each series were randomly set in the range of 11-19 sec. with a mean of 15 sec. A 1-min interval was allowed between the end of a series and the instructions for the next choice. All questions were prerecorded and transmitted to the subject through a loudspeaker. The position of the chosen number in the series was manipulated so that it could occupy the second, fourth, sixth, or eighth ordinal position. The four possible serial positions of the chosen card define four experimental conditions in the present experiment. The subjects were randomly assigned to the four groups under a constraint of approximately equal quotas. For group A, the chosen numbers were positioned in the second, fourth, sixth and eighth ordinal positions in the four sequences respectively. For groups B, C, and D the orders were rotated (e.g., for group B the positions of the chosen numbers were fourth, sixth, eighth and second in the four sequences respectively, etc.)

Results[1]

Responses from the subjects were transmitted to the PDP/8E system. Using an A/D converter and a sampling rate of 20 per second, the maximal conductance change obtained from the subject within the 5 sec following each question was computed. In parallel to the computer, the conductance changes were monitored on a Brush II recorder. For each subject the conductance changes to all 4x8 stimuli were transformed into standard scores relative to the subject's mean and standard deviation. The purpose of the transformation was to eliminate individual differences in responsivity and to enable meaningful summation of responses of different subjects by putting them on the same scale. For reasons of convenience the standard scores had a mean of 10 and a standard deviation of 3. The results were analyzed using several methods, to ensure that conclusions would not depend upon one arbitrary measure. In all the analyses, the responses to the first question in each sequence were omitted, to minimize the effects of initial startle.

For each of the four conditions, an ROC curve was generated using the standardized responses to all the chosen card numbers as SN events and to all the other numbers as N events. The construction of the ROC curves was described in Lieblich et al. (1970). The ROC is a measure of the distance between two random variables. If x and y are two random variables with

---

[1] A rejection region of $p \leqslant .05$ was used in all statistical tests.

distributions F and G rspectively, the ROC curve is F(c) as a function of G(c) where c assumes values from $-\infty$ to $+\infty$. In our case, the ROC curve describes the degree of separation between the distribution of the responses to the chosen cards and the distribution of responses to the other cards. The area under the ROC curve is a statistic that assumes values between 0 and 1. An area of 0.5 would mean that the two distributions are identical and therefore it would not be possible to use the subject's response to identify which card had been picked. An area of 1 means that there is no overlap between the two distributions, and therefore the chosen card could be detected without any errors.

Bamber (1975) showed that the area under an ROC curve has an asymptotic normal distribution. Bamber described a method for estimating the variance of the area statistic and for computing confidence intervals for the true area when fairly large samples are being used. Using this method a 95% confidence interval was computed for the area in each of the four conditions. The areas under the ROC curves as well as the 95% confidence intervals are presented in Table 1. Areas under ROC curves were also computed for the four experimental conditions using, for each subject, the first sequence only. These areas and their respective confidence intervals also appear in Table 1.

Table 1

Areas under ROC curves, 95% confidence intervals for the areas, correct identification rates, estimates for strength of effects, and mean standardized responses for chosen and nonchosen numbers in the four experimental conditions, using all 4 sequences for each subject and using only the first sequence for each subject

Experimental Conditions	Areas	95% Confidence Intervals	Correct Identification Rates	Strength of Effects	Means (SDs) Chosen Numbers[a]	Means (SDs) Nonchosen Numbers[a]
Data Based on all 4 Sequences For Each Subject						
2	0.77	0.72–0.83	49.02%	1.14	12.40 (3.41)	9.33 (2.54)
4	0.71	0.66–0.76	34.31%	0.80	11.64 (3.14)	9.43 (2.68)
6	0.60	0.54–0.66	19.61%	0.37	10.61 (3.03)	9.57 (2.80)
8	0.61	0.55–0.67	25.49%	0.40	10.83 (3.38)	9.68 (2.75)
Data Based on the First Sequence For Each Subject						
2	0.89	0.82–0.96	68.00%	2.01	13.06 (2.59)	8.87 (1.99)
4	0.87	0.81–0.93	53.85%	1.50	12.73 (2.61)	9.12 (2.38)
6	0.69	0.57–0.80	30.77%	0.54	10.89 (2.63)	9.42 (2.73)
8	0.69	0.57–0.81	38.46%	0.76	11.82 (3.58)	9.63 (2.75)

[a]Mean standardized responses and in parentheses the standard deviations of the standardized responses.

The difference between the mean standardized responses to the chosen numbers and the mean standardized responses to the nonchosen numbers divided by the common standard deviation (strength of effect) was computed for each condition and these data are presented in Table 1. The number of subjects who displayed the largest SCR to the chosen number was counted in each condition, defining the number of correct identifications. Table 1 shows the proportion of correct identifications in each condition. In addition, the mean standardized responses for the chosen numbers and for the nonchosen numbers in each condition were computed, and these means are also presented in Table 1. A Chi square analysis was performed to test for dependence between experimental conditions and detectability. Independence was rejected ($x^2(3) = 8.41$). This analysis used only the first sequence for each subject. A Cochran test (Siegel, 1956) was performed, using the data of all sequences, to test for differences in detectability between the four experimental conditions. The null hypothesis was rejected ($x^2(3) - 24.10$). In addition differences in the distribution of the standardized responses to the chosen numbers, among the four experimental conditions, were tested using the Friedman analysis of variance by ranks (Siegel, 1956). Again the null hypothesis was rejected ($x^2(3) = 13.91$).

All the above analyses were also performed separately for each of the 4 sequences, to assess the possibility of habituation of differential responsivity. These results appear in Table 2. Differences in detectability between the 3 sequences were tested using the Cochran test for related samples. Differences in the distribution of the standardized responses to the chosen numbers among the 4 sequences were tested by the Friedman test, revealing a significant order effect ($x^2(3) = 15.13$).

Table 2

Areas under ROC curves, 95% confidence intervals for the areas, correct identification rates, estimates for strength of effects, and mean standardized responses for chosen and nonchosen numbers as a function of the 4 sequences

Sequences	Areas	95% Confidence Intervals	Correct Identification Rates	Strength of Effects	Means (SDs) Chosen Numbers[a]	Means (SDs) Nonchosen Numbers[a]
1	0.78	0.74–0.83	47.57%	1.11	12.12 (3.01)	9.26 (2.50)
2	0.68	0.62–0.74	31.13%	0.75	11.71 (3.49)	9.55 (2.76)
3	0.63	0.57–0.69	26.47%	0.51	11.03 (3.21)	9.59 (2.76)
4	0.59	0.52–0.65	22.68%	0.34	10.57 (3.34)	9.61 (2.34)

Note.—The data for each sequence are based on all subjects.

[a]Mean standardized responses and in parentheses the standard deviations of the standardized responses.

It is clear from the results that the serial position of the relevant stimulus does have an effect on the relative magnitude of the SCR evoked by it. This result is in contrast to a strict interpretation of the dichotomization theory and requires, therefore, a modification of that theory. One possibility is to give a more detailed characterization of the stimuli comprising the sequence. For example, in our case, each stimulus can be defined by several attributes, such as its physical auditory features, and its semantic meaning as a number. Some stimuli in the set have an additional attribute--signal value derived from the fact that they were chosen by the subject. Each attribute when presented induces a process of habituation. The autonomic response to each presentation of a stimulus is an additive function of the state of habituation of its different attributes. To illustrate, let us represent all the attributes of the nonchosen stimulus as N, and all the attributes of the chosen stimulus as S + N. Let us describe two of the experimental conditions used in the present experiment:

(1) N; S + N; N; N; N; N; N; N
(2) N; N; N; N; N; N; N; S + N

In sequence (1) the chosen number appears at serial position 2, while in sequence (2) it appears at serial postion 8. According to the present formulation, the response to S + N in the first sequence should be greater than the response to S + N in the second, since attribute N is only in its second habituation trial in sequence (1), whereas it is in its eighth habituation trial in sequence (2). In both cases attributes S is in its first habituation trial.

This revision is consistent with the previously reported set of findings relating differential autonomic responsivity to the relevant frequencies of the chosen and nonchosen stimuli in a given sequence. The smaller the relative frequency of the chosen stimulus, the smaller the relative frequency of the S element. This element would, therefore, habituate less, and would evoke greater responsivity relative to the S element of the chosen account for the anomalies mentioned in our introduction. According to this modified version of the dichotomization theory, one would indeed expect enhanced responsivity to a chosen number even in a situation where its relative frequency is 0.5. In such a situation the N attribute undergoes exactly twice as many habituation trials as does the S attribute. The second anomaly mentioned in the introduction could also be accounted for by the present formulation. According to this formulation, the differential autonomic responsivity is a function not only of the relative frequencies of the different stimuli, but also of the number of attributes and their frequencies. Thus, when the rate stimulus is the nonchosen number, the response evoked by it is smaller than the response evoked by a rate chosen stimulus, because the latter has an S attribute which the former lacks. Since this new formulation was derived in a post-hoc manner from the present results, it will require an independent set of tests. For example, the modified theory might be examined under conditions where all attributes of the stimuli are neutral. One way to achieve this is to examine habituation to stimuli varying in the number of neutral attributes but having some attributes in common. Then, responses to stimuli with attributes common to all other stimuli should habituate more rapidly than responses to stimuli with fewer common attributes. Such an examination could provide a generalization of the theory, and would

clarify whether the fact that some attributes had signal value is responsible for the results.

Inspection of Table 2 reveals a reduction in differential responsivity with repetitions of the detection task. This effect was observed both in the area statistics and in the correct identification rates. The phenomenon may imply that differential autonomic responsivity could undergo an habituation process. This effect is in accordance with results reported by Lieblich, Naftali, Shumeli, and Kugelmass (1974). Further research is needed to understand the mechanism of habituation of differential autonomic responsivity.

References

Bamber, D. "The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Characteristic Graph." Journal of Mathematical Psychology, 1975, 12, 378-415.

Ben-Shakhar, G. "A Further Study of the Dichotomization Theory in Detection of Information." Psychophysiology, 1977, 14, 408-413.

Ben-Shakhar, G., Lieblich, I., & Kugelmass, S. "Interactive Effects of Stimulus Probability and Significance on the Skin Conductance Response." Psychophysiology, 1982, 19, 112-114.

Lieblich, I., Kugelmass, S., & Ben-Shakhar, G. "Efficiency of GSR Detection of Information as a Function of Stimulus Set Size." Psychophysiology, 1970, 6, 601-608.

Lieblich, I., Naftali, G., Shmueli, J., & Kugelmass, S. "Efficiency of GSR Detection of Information With Repeated Presentation of Series of Stimuli in Two Motivational States." Journal of Applied Psychology, 1974, 59, 113-115.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

* * * * * *

The German Tatbestandsdiagnostik

A Historical Review of the Beginnings of Scientific Lie Detection in Germany

By

H. Herbold-Wootten

Abstract

Scientific lie detection is anchored in the German Psychology in the beginning of this century. In fact, lie detection-then called "Tatbestandsdiagnostik" was the first practical application of the developing scientific Psychology. The originator of the "Tatbestandsdiagnostik" for diagnosis of a crime was Max Wertheimer who is known to most Psychologists only as a gestaltist.

The polygraph technique is usually assumed to be developed in the first decades of this century in the United States. This assumption is clearly reflected in Trovillo's sentence (1939): "Lie detection came about through the application of a method or methods to a specific end; it was never a first act in the growth of an idea."(p. 858).

There was in fact a "first act", It took place in the oldest German University in Prague.

In 1900 Max Wertheimer, then twenty years old, under the influence of contemporary interests represented by W. Stern and H. Gross, began his work at the Physiological Institute of the University of Prague. In 1904 he published a paper together with Julius Klein entitled: "Psychologische Tatbestandsdiagnostik". This word is difficult to translate and it entered American literature in a misleading form. The subtitle of this paper was "Ideas about experimental methods for the purpose of revealing the involvement of a person in a criminal act (Tatbestand)". The article begins with a question: "Isn't it possible to diagnose in a perpetrator the concealed knowledge of his criminal action independent of his statements?" (P. 172)[1] Two years later (1906) in his dissertation he wrote: "Isn't it possible to find experimental methods that allow us to discriminate between persons that know about a crime and those that do not?" (p. 59) In 1935 he formulated it this way:

> The Tatbestandsdiagnostik tries to reveal the objectively present symptoms that the experience of a criminal action (Tatbestand) creates in the reactions of a person independent or even against his verbal statements. For the exact identification of critical symptoms two things are necessary:
>
> 1. A comparison with the reaction pattern of a non-involved person by identical experimental setting and

Reprints of this article are available from the author at 2228 Windward Shore Drive, Virginia Beach, Virginia 23451.

2. A comparison of reaction patterns in the same person to critical and irrelevant stimuli. (p. 1105)

Wertheimer's first article received great attention from the scientific world and created a lively discussion. (Freud, 1907; Grabowsky, 1905; A. Gross, 1907; H. Gross, 1905; Hoegel, 1907; v.d. Hoeven, 1908; Jung, 1905; Kramer & Stern, 1906; Lederer, 1906; Loffler, 1906; Rittershaus, 1909; Weygandt, 1905; Zurcher, 1906) There was, at this time, a discussion between Wertheimer and Jung who claimed that he first had the idea of "Tatbestandsdiagnostik", Jung withdrew this contention in 1908.

The train of thought is the following: If a person had a special, unusual, extraordinary experience this should leave behind traces in his mind or consciousness. Those persons that did not have the same experience should lack those traces. These traces were called "complexes". Therefore, "Tatbestandsdiagnostik" was later also called "Komplexforschung", or "complex search". In other words, it can be expected that these complexes are present in the perpetrator, but not in a person that did not commit the crime in question.

In this first paper Wertheimer and Klein (1904) developed a complete program designed to solve the problem of detection of these complexes. It was self-evident to Wertheimer that the experimental setting in which these complexes could be detected was the association method which he indeed adapted in the following years to the special demands of the "diagnosis of criminal actions". However, he also stated that, in addition, physiological changes in the body would be a reliable source to detect those complexes. In this context he mentioned all recording systems that were at that time available: the plethysmograph which was known through the writings of Muller and Mossi, the pneumograph, the spymograph and hydrosphygmograph known through the work of Lombroso, the psychograph which was a device that recorded involuntary trembling of the finger and was used by Bekhterev and later by Luria and a combination of these devices. Wertheimer and Klein also mentioned altered states of consciousness by narcotica, toxica, and hypnosis to detect these complexes. It was a complete program.

Wertheimer knew the work of Lombroso and mentioned him in his first paper (1904) in a very critical footnote which is quite interesting and shows his strict scientific approach to the problem:

> A single case of similar intention is to be found in a paper by Lombroso entitled 'New Advances in the Study of the Criminal' under the title 'A Crime Cleared by the Hydrosphygmograph'. When a suspect was told about a theft the trace of the hydrosphygmograph dropped 14 mm and by this way his guilt was proven. Any detail about such a capability of the tracing is missing. The report of the results consists of fourteen lines (p. 98).

Although Wertheimer mentioned every physiological recording system that had, at that time, been proven in experimental settings as a reliable indicator of mental or emotional activity, he did not include the GSR. He probably simply forgot it. Tarchanoff in 1890 recorded potential differences in the skin and Fere in 1888 recorded skin conductance; both had published papers about changes in the body as recorded by their methods

under emotional stress. Other experimenters replicated these results under various experimental settings. It is difficult to believe that Wertheimer did not know of these publications.

The work of Wertheimer and Klein was a systematic approach to a practical problem - to detect the perpetrator by use of all physiological and psychological methods that were known at that time through the results of experimental research. Wertheimer was also aware of the special problem that those persons would create, those who know of a crime but are not the perpetrator themselves.

To Wertheimer it was self-evident that the main source for the detection of the perpetrator was the association reaction method which was quite in accordance with the trends of the time.

The first report of an experiment with associations came from Galton (1879) who, in an experiment with himself during a period of one month, found that associations are not freefloating. They arise by certain rules. One of these rules is that "out of every one hundred words twenty-three would give rise to exactly the same association." He also stated, "that the mind is perpetually travelling over familiar ways" (p. 155). Today we would, of course, not call this an experiment. In the same year (1879) Wundt founded the first Psychological Laboratory in Leipzig. He developed the association method to an almost standardized form. The design is as follows: The subject would be confronted with words, either acoustically by spoken word or visually - a card with the word written on it would be sequentially displayed to the subject. The subject was required to respond to this acoustical or visual stimulus with the next word that came into his mind. Usually the list consisted of 100 words of three different types. Sixty words were irrelevant, 20 words were critical or relevant and were mixed with the irrelevant words, and 20 words were considered as postcritical, those words that follow a critical word, but are not related to the critical or complex words. The reaction word and the reaction time were recorded. Symptoms that there is an underlying complex present in the subject are the following (complex reactions):

The clear disclosure of knowledge which only the perpetrator could have, [2]

the translation of the stimulus word to another language,

mere repetitions of the stimulus words,

slips of the tongue,

absence of a reaction, failure to reply,

an excuse for the absence of a reaction.

In addition to the content of the reaction word, a complex is also demonstrated by a prolonged reaction time. The usual reaction time - there are individual differences - is roughly between 1 and 2 seconds. The postcritical words sometimes revealed the complex by perseverance either in the prolonged reaction time or in the content of the reaction word.

The association method existed in several versions of which only one is of interest here - the reproduction method. In this method the subject was required to react twice to the same list of words and was instructed to react in exactly the same way on the second trial as he did on the first. "Guilty" subjects had the tendency to change the reaction word on the second trial to correspond to the relevant stimulus word.

When Wertheimer developed his program of "diagnosis of criminal knowledge" (Tatbestandsdiagnostik) the association method was a well-known test by all psychologists at that time. What was unique, in Wertheimer's article, was his proposal to use this method in a practical application: to identify those persons who had committed a crime. It was, in fact, the first practical application of theoretical psychology.

Wertheimer's article was to stimulate many experiments by others designed to detect "complexes" and the technique appeared to work quite well, even in weak issues, such as determining which subject had viewed a certain picture that the other subjects had not seen (Wertheimer's dissertation). Munsterberg, a Wundt student, and the head of the Psychological Laboratory at Harvard for ten years following 1892, together with other Americans who had studied at the Wundt Laboratory introduced the method in this country. In 1909 Robert M. Yerkes and Charles S. Berry published a paper about an experiment that they conducted during the course in Applied Psychology given in the Harvard University Summer Term in 1908. The title of the paper was: "The Association Reaction Method of Mental Diagnosis (Tatbestandsdiagnostik)". This translation was in the following years used for "Tatbestandsdiagnostik" by all authors who worked in this field, like Henke & Eddy (1909), Leach & Washburn (1910), Goldstein (1923), and many others. It is completely misleading; it is the purely descriptive translation of the one method that was primarily used at that time. Therefore, one method of performing "Tatbestandsdiagnostik" was used as the label for the entire concept. This translation lead to much confusion and misunderstanding. As one example of many: Goldstein (1923) wrote: "Both Jung and Wertheimer claim the discovery of the association reaction method of studying mental processes" (p. 562). Actually, neither of them claimed this discovery.

The Harvard experiment by Yerkes and Berry (1909) was conducted as follows: Two students were selected as subjects. The subjects were instructed to look under one of two covered stacks. Under one stack was a deck of cards, under the other stack was a "dancing mouse". The purpose of the experiment was to determine which of the students had looked under which stack. The report makes it obvious that it is remarkably easy to identify which student had looked under which stack simply by comparing their reaction times to the critical stimulus words.

Even more interesting than how nicely and entertaining this technique worked in experimental settings, is how well it worked in actual cases. The technique was, however, rarely applied and reports of the application to real cases are unfortunately rare. Jung (1905) reported an actual case involving a thief. The case was never verified nor was the thief reported and convicted. Heilbronner (1907) in the Netherlands applied the method in one case and reported very clear results; but the defendant had previously admitted the crime and there was, therefore, nothing to "detect".

Stein (1909) in Zurich used the method with criminal inmates of the mental hospital in which he worked. His subjects were: One insane female, four normal convicted criminals and five suspects that had not confessed - altogether a very heterogenous group. Stein used 10 tests with 100 words each. Fifty words were irrelevant, 25 were complex words and 25 were considered as postcritical. The most obvious indication of a "complex" was the reaction time. The results of this application were, however, ambiguous and Stein concluded that the method should not be used alone and that psychoanalysis is necessary to discriminate between true perpetrators and those persons with guilt complexes.

Montet (1910), a psychiatrist in Switzerland, described a murder case. The defendant was his patient and stated amnesia for the crime. Montet performed several batteries of association tests and came to the conclusion that the defendant did not show any signs of complexes at all, howver, the defendant surprisingly confessed the crime after one of the test sessions. Montet considered the lack of symptoms in the tests as a result of lack of affect and repentance in the defendant, whom he called a neuropath.

Zeiler (1944) reported an interesting case. A father and son were suspected of murdering a customer that had visited their house with 45ØØØ. The customer disappeared and both father and son were suspected of having murdered him. There was evidence that the father initiated and helped to commit the crime but that the son was the actual murderer. Zeiler, a lawyer, tested the son and came to the conclusion that it was overwhelmingly clear that the son committed the murder. Among other signs for an underlying complex, the son required three seconds to respond with a reaction word to the word "father" and then produced the unusual response "bald headed". The usual reaction word to "father" is "son". The defendant later confessed and was sentenced to death.

Here in the United States it was Munsterberg (1914) who used the method in several actual cases. In one reported case - Orchard vs. Idaho - he wrote:

> Orchard had confessed to having committed 14 murders in relationship to an anarchistic mine worker's union and was convicted and sentenced to death. In this case he was a witness against the leaders of the group who stated that Orchard's evidence was a lie. I used the association reaction method to find out whether Orchard was indeed the sorrowful, repentant, confessor who did not want to conceal anything or whether he was a cunning hypocrite who hoped to get away from death by the defamation of the radical leaders. The experiments showed such a complete equanimity, such a true lack of any prolonged reaction time even when the most dangerous points were touched that I became convinced that Orchard did not conceal anything (p. 513).

Crosland (1929) obviously had a great deal of experience with the association reaction method and reported his results in actual cases. He cleared seven criminal cases, five of which were thefts and in six of these cases he obtained confessions. He had a total of fifteen suspects and forty control persons (the inhabitants of an institution, fifteen of which possessed information about the crime). He used the usual list of one hundred words, twenty of which were relevant. Crosland used forty different criteria and lumped them to an index number by statistical methods.

One of the reasons that the verbal association method did not get into the field was because it was not practical. It required much time to develop a list of words for each separate case, perform the experiments and later interpret the results. It was also soon discovered that the method had serious flaws. The general difficulty was that the method could not be standardized. This is especially true for the list of complex words. Even more difficult was the decision regarding which reaction word should be considered as a complex reaction. This was mostly a very subjective decision. Some experimenters, therefore, did not consider the reaction word at all and used only the reaction time as an indication for an underlying complex. This was neither safe nor satisfying. At the end of the century, by the experiments of Cattell it was known that it takes less time to react to concrete nouns than to abstract nouns: also less time is required to respond to concrete nouns than to verbs. Crane (1915) replicated these finding. Lucille Dooley (1916) found that certain words are complex words for everyone - words such as "marriage", "dream", "death", "sin", "coffin", "bride", "friendship". An attempt to clarify this dilemma was the construction of "association dictionaries" in which words were listed with their common reaction words and the percentages of their appearance. Swartz and Crosland (1933) demonstrated experimentally that subjects could, by practice, learn to react faster to complex words than to irrelevant words. They used words of sexual content and trained their subjects in five trials, all in one session, to react faster to complex words than to irrelevant words. This effect had already been observed by Leach and Washburn (1910). Marston was also aware of this fact. Marston discriminated between "good liars" and "bad liars"; the "good liars" were the fast ones.

It became obvious that each case would need to be completely standardized, like a psychological test, before a suspect could be reliable tested. This was, of course, troublesome and impractical. Because of this, the association reaction method, as a practical means of "lie detection", fell into general disuse. Subsequently, the development of lie detection in the United States turned away from the association reaction method. P.V. Trovillo (1939) documented well the historical development. Generally the trend was to use unigraphical and duographical instruments and direct questioning (Marston, MacNitt), until Keeler combined three recording systems in one instrument.

One word about Benussi: It was not Benussi's idea to use the pneumograph to detect lying as many polygraphists seem to think. Ten years before Benussi's experiment Wertheimer had already proposed using breathing symptoms for the diagnosis of crimes. It was Mosso (1878) who first reported experiments in which he observed that the breathing pattern changed under certain stimuli. That a person's breathing is affected by mental and emotional activity was known at the end of the last century. A. Lehmann (1884) stated in a paper about the relationship of breathing and attention that "the breathing flattens under concentrated thinking" (p. 93). Paul Mentz (1895) conducted experiments concerning "the effect to acoustical stimulation on pulse and breathing". Storring (1906) introduced the inspiration/expiration ratio. Benussi (1914) was the first who conducted an experiment to detect lying by the inspiration/expiration ratio. He found that after lying the I/E after the answer compared to the I/E before the answer is greater than 1. When telling the truth this ratio is smaller than 1. Benussi found in his experiment that the inhalation/exhalation ratio in relation to lying or telling the truth is remarkably stable.

He could detect deception even if he instructed his subjects to voluntarily influence their breathing patterns. Except for one repetition by Benscher in 1931 Benussi's results were never replicated. Perhaps the successful detection of deception in his experiments was caused by his special experimental design: He left it up to the subjects to determine at which point during the experiment they wanted to lie (a card test).

Burtt (1921) repeated experiments with the inhalation/exhalation ratio but never obtained Benussi's clear results. There is no indication that this method was ever used in actual cases in the United States. It was used, however, in Germany.

Schutz (1924), a psychiatrist in Leipzig attempted to use this method with prisoners in a jail. His subjects were obviously not cooperative; most of them suffered from a condition called "jail neurosis" and exhibited abnormal reaction patterns. It is not unlikely, however, that many cases of "jail neurosis" were simulated. Schutz concluded that the method was not suitable to actual criminal cases, mainly because he could not obtain clear tracings.

Klemm (1931) attempted this method again with six prisoners in a jail. All were alleged to have committed different crimes. Some of his subjects played "insane", others were actually insane in the legal sense. Many of his tracings were distorted by the subject's talking. The actual questioning had to be conducted in the presence of a judge in jail. Klemm had no standardized questioning procedure. He mixed relevant with irrelevant questions, combined several different relevant questions successively, asked each subject the questions in a different sequence; in some cases he asked the relevant questions more often than in others. It is little wonder that he obtained inconclusive results in most of these cases.

A final word about the first Polygraph. In 1920, Lowenstein in Bonn, introduced an apparatus that could record simultaneouly the movement of each foot, each hand, the head in all three dimensions and had two pneumographs. The subject was placed in a huge chair with his feet hanging free in the air, the head attached to a kind of helmet. The instrument recorded ten traces. Lowenstein, a psychiatrist, used this apparatus experimentally, and also in actual cases. He detected whether or not the inmates in his hospital were suffering from true or false amnesia. The instrument was huge, impractical, and was difficult to keep in working order. This monster never had any practical importance, but it was a true Polygraph.

From the very beginning of the "Tatbestandsdiagnostik" in Germany there was an emotionally charged controversy regarding whether or not the results should be legally admissible in court - the clear tendency was that it should not be. In 1954 the Federal Supreme Court of the Federal Republic of Germany ruled that the result of a Polygrpah test is not admissible in court even if the defendant requests to undergo a polygraph examination to prove his innocence. Since this decision German scientists no longer conduct research in this field.

Footnotes

[1] This and all other citations of German articles are translated by the author.

[2] Assumed a person is suspected of having stolen a red leather wallet with money, the reaction word "red" to the stimulus word "wallet" would have been considerred as a complex reaction.

References

Bechterew, v. W. Die Anwendung der Methode der motorischen Assoziationsreflexe zur Aufdeckung der Simulation. Zeitschrift fur die gesamte Neurologie und Psychiatrie, 19192, 13, 183-191.

Benscher, I. Zur Psychologie des Gedankenlesens. Neue Psychologische Studien, 1931, 5, 67-110.

Benussi, V. Die Atmungssymptome der Luge. Archiv fur die gesamte Psychologie, 1914, 31, 224-273.

Burtt, H. The inspiration-expiration ratio during truth and falsehood. Journal of Experimental Psychology, 1921, 4, 1-23.

Burtt, H. Further techniques for inspiration-expiration ratios. Journal of Experimental Psychology, 1921, 4, 106-110.

Cattell, J.M. Experiments on the association of ideas. Mind, 1887, 12, 68-74.

Cattell, J.M. Mental association investigation by experiment. Mind, 1889, 14, 230-250.

Crane, H.W. A study in association reaction and reaction time, with an attempted application of results in determining the presence of guilty knowledge. Psychological Monographs, 1915, 80.

Crosland, H.R. The psychological methods of word-association and reaction-time as tests of deception. University of Oregon Publication, 1929, 1 (Psychology Series No. 1).

Crosland, H.R. The relative efficacy of visual and two auditory methods of presenting the word-stimuli in an experiment which combines the association-time technique with the psychogalvanic technique. University of Oregon Publication, 1931, 1 (Psychology Series No. 3, 153-178).

Crosland, H.R. The correlation existing between reaction-time and P-G-R's in a combination experiment involving the word-association and galvanic methods. University of Oregon Publication, 1931, 1 (Psychology Series No. 3, 179-202).

Dooley, L. A study in correlation of normal complexes by means of the association method. American Journal of Psychology, 1916, 27, 119-151.

Fere, C. Note sur les modifications de la resistance electrique sous l'influence des excitations sensorielles et des emotions. Comptes Rendus des Seances et Memoires de la Societe de Biologique, 1888, 40, 28-33; 217-219.

Freud, S. Tatbestandsdiagnostik und Psychoanalyse. Archiv fur Kriminal-Anthropologie und Kriminalistik, 1907, 26, 1-10.

Galton, F. Psychometric experiments. Brain, 1879, 2, 149-162.

Goldstein, E. Reaction times and the consciousness of deception. American Journal of Psychology, 1923, 34, 562-581.

Grabowsky, A. Psychologische Tatbestandsdiagnostik. Beitrage zur Psychologie der Aussage, 1906, 2, 126.

Gross, A. Die Assoziationsmethode im Strafprozess. Zeitschrift fur Kriminal-Anthropologie und Kriminalistik, 1905, 19, 49-59.

Heilbronner, K. Die Grundlagen der "psychologischen Tatbestandsdiagnostik". Nebst einem praktischen Fall. Zeitschrift fur Strafrechtswissenschaft, 1907, 27, 601-656.

Henke, F.G. & Eddy, M.W. Mental diagnosis by the association reaction method. Psychological Review, 1909, 16, 399-409.

Hoegel. Die "Tatbestandsdiagnostik" im Strafverfahren. Monatschrift fur Kriminalpsychologie und Strafrechtsreform, 1907, 4, 26-31.

van der Hoeven, H. De invloed der affectieve meerwaarde van vorstellingen in het woordreaktie-experiment. Met een anhangsel over de bruikbaarkeit van het experiment in het straf-proces. Medical Dissertation, Leiden 1908.

Jung, C.G. Die psychologische Diagnose des Tatbestandes. Schweizerische Zeitschrift fur Strafrecht, 1905, 18, 368-408.

Jung, C.G. Zur Tatbestandsdiagnostik. Zeitschrift fur angewandte Psychologie, 1908, 1, 163.

Jung, C.G. Die psychologische Diagnose des Tatbestandes. Zurich: Rascher, 1941.

Kaganiec, H.J. Lie-detector tests and "freedom of will" in Germany. Northwestern University Law Review, 1956/57, 51, 446-460.

Klemm, O. Uber die Atmungssymptomatik bei Untersuchungsgefangenen. Neue Psychologische Studien, 1931, 5, 111-132.

Kramer, F. & Stern, W. Selbstverrat durch Assoziation. Experimentelle Untersuchungen. Beitrage zur Psychologie der Aussage, 1906, 2, 457-488.

Leach, H.M. & Washburn, M.F. Some tests by the association reaction method of mental diagnosis. American Journal of Psychology, 1910, 21, 162-167.

Lederer, M. Zur Frage der pathologischen Tatbestandsdiagnostik. Zeitschrift fur die gesamte Strafrechtswissenschaft, 1906, 26, 488-506.

Lederer, M. Die Verwendung der psychologischen Tatbestandsdiagnostik in der Strafrechtspraxis. Monatschrift fur Kriminal-Psychologie und Strafrechtsreform, 1906, 3, 163-172.

Lehmann, A. Uber die Beziehungen zwischen Athmung unde Aufmerksamkeit. Zeitschrift fur angewandte Psychologie und psychologische Sammelforschung, Beiheft 51, Leipzig, 1911.

Lipmann, O. & Wertheimer, M. Tatbestandsdiagnostische Kombinationsversuche, Zeitschrift fur angewandte Psychologie, 1908, 1, 119-128.

Loffler, A. Zur psychologischen Tatbestandsdiagnostik. Monatschrift fur Kriminal-Psychologie und Strafrechtsreform, 1906, 3, 449-466.

Lowenstein, O. Uber den Nachweis psychischer Vorgange und die Suggestibilitat fur Gefuhlszustande im Stupor. Zeitschrift fur die gesamte Neurologie und Psychiatrie, 1920, 61, 304-350.

Lowenstein, O. Uber subjektive Tatbestandmassigkeit und Zurechnungsfahigkeit nebst kritischen Bemerkungen zur psychologischen Tatbestandsdiagnostik. Archiv fur Psychiatrie, 1922, 65, 411-458.

Luria, A.R. Die Methode der abbildenden Motorik in der Tatbestandsdiagnostik. Zeitschrift fur angewandte Psychologie, 1930, 35, 139-183.

MacNitt, R.D. In defense of the electrodermal response and cardiac amplitude as measures of deception. Journal of Criminal Law, Criminology, and Police Science, 1942, 33, 266-275.

Marston, W.M. Negative type reaction-time symptoms of deception. Psychological Review, 1925, 32, 241-247.

Mentz, P. Die Wirkung akustischer Sinnesreize auf Puls und Athmung. Inaugural Dissertation, Leipzig, 1895. Philosophische Studien, 1985, 11, 61-124; 371-393.

Montet, d. Ch. Assoziationsexperimente an einem kriminellen Fall. Monatschrift fur Kriminalpsychologie und Stragrechtsreform, 1919, 6, 37-47.

Mosso, A. Uber die gegenseitige Beziehung der Bauch- und Brustatmung. Archiv fur die gesamte Physiologie, 1878, 18, 441+.

Mosso, A. Uber den Kreislauf des Blutes im menschlichen Gehirn. Leipzig: S. Hirzel, 1889.

Muller, R. Zur Kritik der Verwendbarkeit der plethysmographischen Curve fur psychologische Fragen. Zeitschrift fur Psychologie und Physiologie der Sinnesorgane, 1902, 30, 340-390.

Munsterberg, H. Grundzuge der Psychotechnik, Leipzig: J.A. Barth, 1914.

Rittershaus, E. Die Komplexforschung ("Tatbestandsdiagnostik"). Journal fur Psychologie und Neurologie, 1909, 15, 61-83; 184-220.

Rittershaus, E. Die Komplexforschung ("Tatbestandsdiagnostik"). Journal fur Psychologie und Neurologie, 1910, 16, 1-43.

Schult, O. Psychologische Tatbestandsaufnahmen an Untersuchungsgefangenen. Archiv fur Kriminologie, 1924, 76, 150-151.

Schnitzler, J.G. Experimentelle Beitrage zur Tatbestandsdiagnostik. Zeitschrift fur angewandte Psychologie und psychologische Sammelforschung, 1908, 2, 51-91.

Stein, P. Tatbestandsdiagnostische Versuche bei Untersuchungsgefangenan. Zeitschrift fur Psychologie, 1909, 52, 161-237.

Stern, W. Zur Psychologie der Aussage: Experimentelle Untersuchungen uber Erinnerungstreue. Zeitschrift fur die gesamte Strafrechtswissenschaft, 1902, 22, 315-370.

Storring, G. Experimentelle Beitrage zur Lehre vom Gefuhl. Archiv fur die gesamte Psychologie, 1906, 6, 316-356.

Swartz, B.K. & Crosland, H.R. The effect on reaction-time of various orders and distributions of relevant stimuli in the word-associations experiment. Journal of Genetic Psychology, 1933, 8, 217-269.

Tarchanoff, J. Ueber die galvanischen Erscheinungen in der Haut des Menschen bei Reizungen der Sinnesorgane und bei verschiedenen Formen der psychischen Thatigkeit. Pfluger's Archiv fur die gesamte Physiologie, 1890, 46, 46-55.

Tent, L. Psychologische Tatbestandsdiagnostik. In U. Undeutsch (Ed.) Handbuch der Psychologie, Bd. 11, Forensische Psychologie. Gottingen: Verlag fur Psychologie, Hogrefe, 1967.

Trovillo, P.V. A history of lie detection. Journal of Criminal Law, Criminology and Police Science, 1939, 29, 848-881 and 30, 104-119.

Wertheimer, M. Uber die Assoziationsmethoden. Archiv fur Kriminal-Anthropologie und Kriminalistik, 1906, 22, 293-319.

Wertheimer, M. Experimentelle Untersuchungen zur Tatbestandsdiagnostik. Inaugural Dissertation, Wurzburg, 1905. Archiv fur die gesamte Psychologie, 1906, 6, 59-131.

Wertheimer, M. Tatbestandsdiagnostik. In E. Abderhalden (Ed.) Handbuch der biologischen Arbeitsmethoden, Abt. VI, Teil, C., II, 2, Berlin, Wien, 1935.

Wertheimer, M. & Klein, J. Psychologische Tatbestandsdiagnostik. Archiv fur Kriminal-Anthropologie und Kriminalistik, 1904, 15, 78-113.

Weygandt, W. Zur psychologischen Tatbestandsdiagnostik. Monatschrift fur Kriminalpsychologie und Strafrechtsreform, 1905, 2, 435-438.

Yerkes, R.M. & Berry, C.S. The association reaction method of mental diagnosis (Tatbestandsdiagnostik). American Journal of Psychology, 1909, 20, 22-37.

Zeiler, A. Ein Versuch mit der "Assoziationsmethode". Archiv fur die gesamte Psychologie, 1944, 112, 77-82.

Zurcher, E. Zur psychologischen Diagnose des Tatbestandes. Monatschrift fur Kriminal-Psychologie und Strafrechtsreform, 1906, 3, 173-175.

* * * * * *

On the Accuracy of the Polygraph; An Evaluative Review of Lykken's Tremor in the Blood

By

Gordon H. Barland*

A Tremor in the Blood: Uses and Abuses of the Lie Detector, by David T. Lykken (New York: McGraw-Hill, 1981. 317 pages. Indexed. $14.95) is Lykken's magnum opus. In 22 chapters divided into a prologue and four parts, Lykken surveys and assesses the broad panorama of the polygraph field, charting new ground and putting each aspect into a new perspective, critically evaluating the strengths and weaknesses of the polygraph technique in a variety of applications. Unfortunately, Lykken's effort seems to have been handicapped by a lack of a clearly defined audience. On the one hand, Lykken reviewed and analyzed some of the more recent research in the detection of deception, presenting new views and thoughts of interest primarily to the scientific and legal communities. On the other hand, Lykken appears to have aimed much of the book toward the general public. This inherent conflict in the goal of the book may be responsible for what appeared to me to be some major defects in the book: a failure to document and support many of his assertions with references, a highly selective review of the literature that seemed to be polemical, a superficial approach toward a number of complex issues that in all fairness deserved a more profound analysis, and a number of misconceptions regarding the polygraph technique. Indeed, it was difficult to find pages in the book that did not contain what I consider to be errors of one sort or another, some trivial, but others substantive. These are serious charges, the more so because Lykken is regarded by many within the scientific community as an authority on the detection of deception, and because his book has been favorably reviewed by psychophysiologists familiar with the research literature (Stern, 1981; Waid, 1981). Therefore, before I mention the positive aspects of the book and the genuine contributions Lykken makes with it, it would seem appropriate to document some of the more important errors and shortcomings.

Lykken's concept of the polygraph technique is quite distorted. I would certainly agree with him that a psychologist need not have attended a polygraph school in order to understand the fundamentals of the polygraph technique (pp 4-5). Orne and his colleagues, for example, have an excellent appreciation of many aspects of the polygraph technique,(Orne, Thackray & Paskewitz, 1972), although they, like Lykken, have never attended a polygraph school. However, a degree in psychology does not necessarily confer an accurate understanding of polygraph procedures and techniques, or that it make one immune to bias when interpreting the literature. Lykken's concept of how a polygraph examination is conducted is chillingly portrayed in Chapter 2. Portions are quite accurate, such as the description of how control questions are developed. However, the duplicity of the examiner, his refusal to interpret the polygraph charts, his propensity to interrogate all suspects regardless of what the charts show, betray a profound ignorance on Lykken's part as to what constitutes

*I should like to thank the following people for their helpful comments concerning an earlier draft of this paper; Norman Ansley, Benjamin Kleinmuntz, David Lykken, Robert Stern, Udo Undeutsch, and William Waid.

the essence of the polygraph examination. This misconception colors much of Lykken's discussion of the polygraph throughout the book, such as when he evaluates the Bersh validity study. The fact that Lykken, himself, has conducted such an inquisitorial examination, which he describes on pages 87-89, speaks for the need to have polygraph examiners licensed, so that the public can be protected from untrained practitioners who might be inclined to use the polygraph as a psychological bludgeon to extort confessions. More accurate descriptions of the polygraph technique are to be found in Barland and Raskin (1973), Matte (1980), and Reid and Inbau (1977).

ACCURACY OF THE POLYGRAPH TECHNIQUE

One of the major goals of the book is to critically review the research literature to determine the accuracy of various testing methodologies. Lykken quite properly draws a distinction between the examiner who makes clinical decisions based upon analysis of both the polygraph charts and nonpolygraphic information such as the subject's demeanor and behavior, versus the examiner who bases his decisions solely upon a careful analysis of the biological information contained in the polygraph charts. While this distinction is an important one, it could be argued that the degree to which observations of the subject's demeanor affects the examiner's decision is more a continuum than a dichotomy. Consequently, a more practical approach to studying the accuracy of the polygraph might be to look at the accuracy of the decisions made by the original examiner versus the decisions of a reviewing examiner making blind analyses of the polygraph charts in isolation from all sources of extrapolygraphic information. Of course, the more the original examiner emphasizes a rigorous evaluation of the polygraph charts as the sole basis for a decision, the more his decisions approximate those made by the reviewing examiner.

Lykken contends that the polygraph is, on average, about 70% accurate, and that it cannot detect the truthfulness of innocent suspects at better than chance levels. He arrives at this conclusion after reviewing several studies, notably those by Bersh (1969), Horvath (1977), and Barland and Raskin (1976). Because Lykken's conclusion differs markedly from previous estimates of the polygraph technique's accuracy, it is important to examine the basis for Lykken's assessment.

THE BERSH STUDY

One of the main studies which attempts to determine the accuracy of the clinical diagnoses of polygraph examiners is that reported by Bersh (1969). That study, conducted by the Department of Defense, reported a 92.4% rate of agreement between the original examiners' decisions and the unanimous decisions of a panel of four military lawyers who were given access to all available case facts except for the polygraph test. Lykken analyzes this study in Chapter 6 and finds that it suffered from contamination. He argues that it is biased in favor of high validity because the examiners based their diagnoses in part on some portion of the same case facts that the four panel judges used in reaching their criterion decision.

In pleading the case that this lack of independence significantly raised the apparent accuracy of the examiners' decisions, Lykken pointed

out that (a) the examiner could not have derived much information from the polygraph charts, because a reliability study conducted by the DOD on those same charts produced a low reliability figure, which limits the validity of the test; and (b) when the evidence was presumably strong (the panel was unanimous), polygraph results agreed with the panel more often than when the evidence was presumed to be weaker (the panel was split 3:1). Lykken concludes that the clinical examiners' decisions are about 75% accurate because that is the presumed accuracy of behavioral observations.

Because the Bersh study is a pivotal study, it is important that it be analyzed carefully to determine if Lykken's rejection of it is sound. The main criticism is the lack of independence between the examiner's decisions and the panel decisions. It must be kept in mind that the purpose of the Bersh study was to estimate the accuracy of the original examiners' decisions. It did not attempt to determine to what extent the decisions were based upon the physiological information in the polygraph charts, as opposed to nonpolygraphic sources of information. Consequently, because the original examiner must know the case facts in order to conduct a proper examination, it is perhaps impossible to design a field validation study in which there is complete independence between the information available to the original examiner and to the criterion panel. Bersh was aware of the difficulties of obtaining a criterion, for he discussed the choice of possible criteria against which the examiners' decisions could be measured, and found that none was ideal. Confessions, prima facie evidence of guilt or innocence, and court-martial decisions each has inherent limitations and sources of bias. Bersh felt that the best approximation of ground truth would be a careful sifting of all available investigative information bearing upon the issue, less the polygraph results.

Lykken is quite correct in pointing out that low reliability in the blind interpretation of the polygraph charts implies that the examiners could not have derived much information from the polygraph charts. It would not necessarily mean that the original examiners were influenced by investigative information. Probably the single most likely source of nonpolygraphic information influencing an examiner's decision would be his observation of the subject's patterns of verbal and nonverbal behavior.

Unfortunately, Lykken does not mention what the reliability was, nor does he cite any reference to which the reader can refer. A 1968 report by the Department of Defense made public by the U.S. House of Representatives in 1974 discussed that validity study and a separate reliability study. It is not clear from the report whether the polygraph charts used in the reliability study were also used in the validity study reported by Bersh. The Bersh (1969) study had 157 cases in which the panel decision was unanimous. The reliability study consisted of 90 sets of charts which were analyzed by 30 examiners. Thirty of the 90 sets of charts were peak-of-tension tests. The Bersh study does not report the use of any peak-of-tension tests. If the reliability of peak-of-tension tests differs significantly from the reliability of other test formats, then the results would not necessarily apply to the Bersh study.

Brisentine, who was responsible for data collection for both the Bersh study and the reliability study, has discussed both studies (Brisentine, 1974). He pointed out that the reliability of individual analysis

ranged from 75% to 85%, the reliability of the group decisions ranged from 75% to 94%, and the percent agreement between the group decision and the original examiner was in the high nineties. Unfortunately, he did not explain what was meant by the "reliability of the group decisions." He also pointed out that the examiners involved in the study had not all been trained in the same polygraph school, and were not all familiar with the three types of test formats used in the study. That may have decreased the reliability. These deficiencies have since been remedied among federal examiners by increasing the training standards and by having the examiners trained at the same polygraph school (Brisentine, 1974). Nonetheless, the results would seem to be generally consistent with other reliability studies, and suggest no new cause for concern.

Lykken has pointed out that when the Bersh study evidence presumably was strong (the panel was unanimous), the polygraph examiner agreed with the panel 92% of the time. When the evidence was presumed to be weaker (the panel was split 3:1), the polygraph results agreed with the panel majority 75% of the time. Lykken interpreted this as indicating that the polygraph examiners may have relied largely upon the investigative information when making their decisions. Bersh, however, did not believe this to be a major factor.

> It is very unlikely, however, that the examiner's judgment is determined solely by the file, or even by the file in combination with information provided by the criminal investigator. The polygraph examination is ordinarily not given to suspects whose guilt or innocence has already been substantially or finally established. Only where real doubt exists about the guilt status of the suspect is he permitted or asked to volunteer for an examination. The fact is that the case file at the time of the polygraph examination was less complete, and often far less complete, than when it was delivered to the JAG panel. (p. 402)

Because the Bersh study involved real life criminal investigations, ground truth could not be definitely established. The panel's review of the investigative information represented the closest possible approximation of ground truth. When there was a disagreement between the panel's decision and the polygraph outcome, it is conceivable that the panel could have been wrong, even when the panel's decision was unanimous. The possibility of an erroneous panel decision would be expected to be greater in a situation where one of the panelists, after reviewing the information, arrived at a decision opposite to the majority. Of course, no one has any idea how often the minority opinion is likely to be correct, but it is not inconceivable that it might be correct perhaps 25% of the time. Finally, when the panel was split 2:2, a randomly selected decision by one pair of panelists would be expected to be correct 50% of the time. Therefore, even if the polygraph examiners' decisions were 100% correct (for the sake of discussion), it would not be unreasonable to find rates of agreement between the panel and the polygraph examiners of 92% and 75%, as the Bersh study found. Because of this alternative hypothesis to explain the results of the Bersh study, Lykken's contention that the results indicate the examiners were unduly biased by investigative information is not compelling.

One final point needs to be made concerning Lykken's assessment of the Bersh study. Lykken stated:

> An extreme example of this contamination involves the fact that an unspecified number of guilty suspects confessed at the time of the examination. Because the exams were clinically evaluated, we can be sure that every test which led to a confession was scored as Deceptive. Since confessions were reported to the panel, we can be sure also that the criterion judgement was always Guilty in these same cases. Thus, every lie test that produced a confession was inevitably counted as an accurate test although, of course, such cases do not predict at all whether the polygraph would have been correct absent the confession. (p. 99).

This is an example of how Lykken's misconceptions about how often examiners interrogate suspects seems to have biased his interpretation of the literature. Examiners do not routinely interrogate regardless of what the polygraph charts show. They generally do not interrogate unless the polygraph charts indicate the suspect is not telling the truth. Depending upon the circumstances, some examiners may interrogate when the charts are inconclusive, if the suspect's pattern of behavior indicates deception. Bersh does not mention whether the panel was always unanimous in those cases where a confession was included in the information provided them, but Lykken's assertion that a unanimous decision of guilty is inevitable is overstated. In my replication of the Bersh study (Barland, 1975), one of the panelists who was provided with a suspect's confession concluded that the suspect was innocent!

There is no doubt but that the examiners in the Bersh study had access to some of the same investigative information later provided the panel, and this lack of independence may have contributed to the rate of agreement between the panel's decisions and those of the original examiners. That rate, 92%, should therefore be regarded as the upper limit of the accuracy of the military examiners involved in criminal investigations in the early 1960's. The actual rate was probably somewhat less, although it is impossible to determine precisely how much less. Considering the issue that Bersh was studying--the accuracy of clinical judgments made by the original polygraph examiners involved in criminal investigations--the methodology seems to be the best available, and his findings should be given appropriate weight.

THE HORVATH STUDY

A second major study which Lykken relies upon in his attempt to assess the accuracy of the polygraph technique is the study by Horvath (1974, 1977), who randomly selected 56 verified polygraph files from the archives of a major police department. Half were verified innocent and half were verified guilty, each by the confession of the guilty person. Horvath had 10 examiners evaluate each chart blindly. He found that the blind evaluations were 77% correct with guilty suspects and 51% correct with truthful persons. Lykken (pp. 124-125) interpreted these results to mean that the control question test cannot detect truthfulness in real-life situations. Because of the profound importance of such a conclusion, if it be true, it is necessary to assess this study carefully.

In analyzing the Horvath study, it is immediately evident that it was never designed to estimate the validity of the original examiner's

decisions. It was intended to study the validity and reliability of blind evaluations of polygraph charts. As such, it is methodologically related to four previous studies conducted by the Chicago firm of John E. Reid & Associates, with which Horvath was earlier associated. These are discussed by Lykken (pp. 122-124), who characterized them as validity studies, although they were not designed to estimate the accuracy of the original examiners' decisions. All five studies were designed to estimate the reliability and validity of blind evaluations of polygraph charts, which is a separate, though related, issue. It is important to keep the distinction in mind when attempting to estimate the accuracy of the polygraph technique, for there are many different aspects of validity, e.g., validity of the original examiner's decisions; validity of blind chart analysis; validity with various populations such as truthful subjects, deceptive subjects, sociopaths, victims and witnesses versus suspects; validity when base rates for truthfulness or deception are appreciably different from 50%, etc. Unless the type of validity is specified, confusion can occur in discussions among the experts due to a mismatch of definitions. Throughout his book, Lykken generally failed to distinguish between the validity of the original examiners' decisions versus the decisions made by reviewing examiners.

In the 56 verified cases in Horvath's study, the original examiners were correct 100% of the time (Horvath, personal communication, 1982), a fact not mentioned by Horvath in his dissertation or article. This does not suggest that the polygraph is anywhere near 100% accurate, but it does suggest that the accuracy of the original examiners' decisions is likely to be well above chance, even with verified innocent suspects, for Horvath found no errors when he was selecting his sample from the polygraph archives of the police department involved (Horvath, personal communication, 1982). Although such studies by themselves cannot determine the accuracy of the original decisions made by polygraph examiners, they can provide very useful information about the accuracy of blind evaluations of polygraph charts (which is quite a different issue), and factors which might affect that accuracy.

The key question raised by the Horvath study is why its results were so different from the previous studies, despite the same general research design. Several alternative hypotheses arise. Lykken suggests that it is because the control question polygraph technique cannot detect truthfulness beyond chance levels in a real-life situation. This is unlikely because (a) the original examiners were right 100% of the time, and Horvath was unable to find any verified errors, (b) other studies using control question methodology in mock crime situations have found high accuracies for both deceptive and truthful subjects, and (c) Bersh (1969) study found no difference in the false negative (FN) and false positive (FP) error ratio.

While several studies have found higher FP error rates than FN error rates (e.g., Barland & Raskin, 1975), the only other published study--either a laboratory study involving the control question test in a mock crime paradigm, or in real-life situations--which suggests that the polygraph technique control question test cannot detect truthfulness beyond chance levels is my doctoral research (Barland, 1975), discussed below.

One alternative explanation of Horvath's results is that the polygraph examiners who made the blind evaluations did not know how to interpret charts very well (Raskin, 1978). That is unlikely to be the major cause, in view of the high interevaluator reliability coefficient of .89 found by Horvath. Another hypothesis is that the accuracy of the blind evaluations was reduced because the reviewing examiners did not have access to all of the polygraph charts that were obtained by the original examiners. Thirty-two percent of the examinations contained a special chart administered primarily to subjects the original examiner considered deceptive. Those charts were removed from the files before being reviewed by the blind examiners (Horvath, 1974). Another difference between the Horvath study and the preceeding ones is that the earlier studies used subjects tested by private examiners whereas the subjects in Horvath's study were tested by police examiners.

Yet another hypothesis is that the type of subjects sampled by Horvath may have been different from previous studies. For example, Horvath's study included charts from a number of truthful victims and witnesses. One of the conditioning theories of the detection of deception predicts that innocent victims or witnesses might be expected to react emotionally on the polygraph when asked about the events they experienced or witnessed precisely because they are telling the truth. If this be true, it would suggest that until appropriate modifications to the polygraph technique can be made, the clinical judgment of the examiner can be an important safeguard against false positive errors which could occur if the charts were relied upon in isolation from other factors. A review of the cases used in the Horvath study should be undertaken to investigate this hypothesis. Until additional research can determine why Horvath's results are so different from the preceeding, methodologically related studies, it would be premature to conclude that the polygraph technique cannot detect truthfulness, particularly in view of studies to the contrary (e.g., Bersh, 1969; Bitterman & Marcuse, 1947; Blum & Osterloh, 1968).

THE BARLAND STUDY

My doctoral dissertation explicitly attempted to estimate the accuracy of the polygraph technique with criminal suspects (Barland, 1975; Barland & Raskin, 1976). Lykken discusses this study on pp. 125-126. Because he feels it dovetails with Horvath's study, he considers it additional evidence that the control question technique cannot detect truthfulness.

This study was an extension of the Bersh study. A panel of four lawyers and a judge were given investigative dossiers on each of 92 independent criminal suspects to whom control question polygraph examinations had been administered. Any reference to the outcome of the polygraph tests was deleted from the dossiers. Their decisions regarding the guilt or innocence of each suspect then served as a criterion against which the decision of the polygraph examiner could be compared. The charts were also interpreted blindly by Dr. David C. Raskin, whose decisions could also be compared against the panel's decisions. As shown in Table 1, there were 51 cases in which both the panel majority and the blind examiner made definite decisions. They agreed 86% of the time. Of the 51 cases, 40 were criterion guilty, 11 were criterion innocent. The examiner's decisions were 98% accurate with the criterion guilty suspects, but

only 45% accurate with the criterion innocent suspects, for an average accuracy of 72% when the number of criterion guilty and criterion innocent suspects are equalized. The results certainly suggest that there is a significant problem in detecting the truthfulness of innocent suspects!

Table 1

Decision by panel majority

		Guilty	Innocent	Total
Decision by blind examiner	Deceptive	39	6	45
	Truthful	1	5	6
	Total	40	11	51

However, before that conclusion can be adopted, all equally plausible alternative hypotheses must be explored and rejected. One of the differences between this study and the Bersh study appears to be the amount of investigative information provided to the panel. The DOD Joint Working Group which conducted the Bersh study was able to obtain all pertinent investigative files, whereas my study relied upon university students to obtain the investigative information from police departments, prosecutors, and defense attorneys. Unfortunately, the amount of documentation they obtained was often less, sometimes far less, than desired. This may have tended to bias the panel towards calling suspects innocent in two ways. First, not every file contained all of the police's evidence against the suspect, but every file always contained the suspect's alibi or other statements supporting the suspect's innocence. Secondly, the incompleteness of the case files could have permitted the expression of any latent bias towards a "not guilty" decision by the panelists, caused by their legal training and experience. The American judicial philosophy seeks to minimize false positive errors (convicting an innocent person) at the risk of increasing false negative errors. Upon being debriefed at the end of the study, one of the two prosecutors who served as panelists stated that he had been dismayed at the incompleteness of some of the cases, and would have refused to prosecute a number of them on that basis. Although the panelists were instructed to render "no decision" rather than "probably innocent" in such cases, that prosecutor further stated that, in an effort to help out the study by making as many decisions as possible, he felt that he tended to call suspects "probably innocent" in a number of cases in which the evidence was insufficient to support a decision. Another panelist, a defense attorney, tended to call most suspects innocent even though, in one case, the suspect had confessed. He cleared suspects more often than the other panelists at a rate which was statistically different from the four other panelists. This, in turn, tended to bias the composite panel results, but it was felt that he could not be excluded from the panel post hoc.

Normally, when the variable under investigation disagrees with the criterion, the criterion is assumed to be correct. In this study, however, because of the possible sources of bias, additional research is required

to investigate those cases in which the polygraph outcome disagreed with the panel, to determine which was more likely to be correct (together with a suitable control group of instances in which they agreed). This is especially important because there were very few criterion innocent subjects in the study. Consequently, the 95% confidence interval for the false positive error rate ranges from 11% to 59% (Raskin & Podlesny, 1979).

Lykken seems to feel that the polygraph is much more accurate at detecting deception than it is in detecting truthfulness; the controversy thus focuses primarily on the issue of how accurate the polygraph is in identifying the truthful person. Additional research is required to resolve that controversy.

ASSUMPTIONS OF POLYGRAPH TESTS

One of the genuine contributions which Lykken makes with this book is his attempt to explicitly state the psychological assumptions underlying the various testing techniques which he describes in chapters 6 through 12. Lykken tries to be fair in this task:

> And we must be caseful not to infer unnecessary assumptions or to set up a straw-man theory that is easy to topple but unfair to polygraphers who have faith in this method. Let us, therefore, proceed cautiously.
>
> ASSUMPTION 1. A given subject will respond more strongly to a relevant question if he answers it deceptively than if his denial is truthful. That is, if his response would be R_I if he is innocent and R_G if he is guilty, then R_G will be larger than R_I ($R_G > R_I$).
>
> This clearly is an assumption of <u>any</u> form of lie test and it seems generally plausible. But ... it is important to emphasize that 'plausible' does not imply 'inevitable.' (p. 113)

Unfortunately, within four pages he is stating assumptions in such an extreme form that no intelligent examiner could possibly agree with them:

> ASSUMPTION 1-B. All guilty subjects will regard the relevant questions as more threatening than the control questions and therefore R_G will be larger than R_C .
>
> ASSUMPTION 2-B. All innocent subjects will regard the relevant questions as less threatening than the control questions and therefore R_C will be larger than R_G . (p. 117)

How much better it would have been if Lykken had maintained the same objective stance he had started with, by stating the assumptions, as he had initially, "A given subject who is guilty will regard the relevant questions..." That notwithstanding, Lykken has rendered a valuable service by stating explicit assumptions, for which he should be given proper credit. One could argue that it is necessary to state all assumptions in an extreme form, for any deviation from the assumption would then be expected to cause errors; whether departures from the assumption cause only

a small number of errors or whether they cause a large number of errors is a legitimate object of research. If this is what Lykken had in mind, then he should have been consistent from one assumption to another, and should have explained his rationale to the reader, most of whom are not sufficiently versed in psychometrics to know why he was being unrealistically extreme. Lykken compounded this oversight by making it appear that examiners actually believe such palpably false extremes:

> Again, we cannot help but be impressed by the simplistic, robot-like conception of human nature, the blithe disregard of the subtleties and idiosyncracies of real human minds. Infallibly, the relevant question is 'most threatening' to the guilty subject. ... Universally, we are told, the relevant question is not 'most threatening' to the innocent subject because the polygrapher will always make him believe that the control questions pose a greater threat. (p. 117)

CONDITIONAL PROBABILITY

Chapter 18, entitled "The Lie Detector and the Courts," is for many readers the most important chapter in the book. Here Lykken uses the statistical concept of conditional probability to argue that polygraph results should not be admitted as evidence in court, on the premise that when the base rate for either innocence or guilt is substantially different from 50%, certain types of polygraph errors become relatively more frequent. Thus, even if the accuracy of the polygraph technique were normally considered acceptable, under these special conditions either the FP or FN error rate would increase disproportionately to the point where no reasonable person could advocate its use as evidence.

To illustrate his reasoning, Lykken assumes that when the evidence is quite weak, 70% of the suspects taking stipulated polygraph examinations are in fact innocent, and that the polygraph is 90% accurate with guilty people and 50% with innocent people. He then concludes that:

> Out of each 100 stipulated tests, 27 out of 30 guilty suspects will be classified Deceptive and so will 35 out of 70 innocent suspects. Since only the failed lie tests will be presented to the court, 35 of the 62 lie tests offered into evidence on this principle will be erroneous, a rate of accuracy actually worse than could be obtained by flipping coins! (p. 221)

The remark about the polygraph being less accurate than the flip of a coin is factually wrong, as it assumes that chance is 50%. In the situation which Lykken describes in which the base rate for guilt is 30%, one would have only a 30% chance of detecting the guilty subjects by flipping a coin. Given the assumptions Lykken has made, use of the polygraph would actually help the court--even if it couldn't detect truthfulness better than chance--by concentrating the number of guilty subjects in the deceptive group. Table 2 summarizes the effect of the polygraph examinations given Lykken's assumptions:

Table 2

		Ground Truth		
		Guilty	Innocent	Total
Polygraph Outcome	Deceptive	27	35	62
	Truthful	3	35	38
	Total	30	70	100

Without the polygraph, Lykken has assumed that 70% of the subjects facing trial under such circumstances are innocent. If the polygraph were employed, and the results relied upon by the prosecution when deciding whether to bring the case to trial (we will assume that he dismisses charges against those who are called truthful by the examiner, and prosecutes those who are called deceptive), then the proportions of innocent suspects who are brought to trial is reduced from 70% to 56% because 35(92%) of the 38 subjects that the examiner called truthful in this situation were innocent. Is this benefit obtained at the expense of the guilty suspects? No, for the proportion of guilty suspects facing trial is increased from 30% to 44%. Thus, even if the polygraph were only 50% accurate in determining truthfulness, it would nonetheless be helpful to the judicial system by screening out 50% of the innocent suspects and concentrating the proportion of guilty persons among those who are still subject to trial. If the polygraph is significantly more than 50% accurate in detecting truthfulness, as the bulk of the evidence indicates, then the results of the polygraph testing are even more useful.

Lykken next (p. 222) presents a different example to support his contention that the polygraph is so erroneous that it should not be used as evidence. Looking at the situation in which the evidence of guilt is strong, Lykken estimates that some 80% of the suspects brought to trial are guilty. In such a situation, if the polygraph is 90% accurate in detecting deception and 50% accurate in detecting truthfulness, then it follows that polygraph evidence will be wrong 44% of the time when it is introduced as evidence. Here, again, Lykken's treatment of such an important topic is superficial. As in the previous example, he fails to ask some important questions: What percentage of such cases result in conviction? What percentage of innocent suspects are convicted? What percentage of guilty suspects are acquitted? The key issue is really this: Does the use of the polygraph help or hinder the administration of justice? With these questions in mind, let us look at the same situation, with the same assumptions that Lykken suggested.

Table 3 confirms Lykken's assertion that 8 of 18 suspects called truthful by the examiner are in fact guilty, for an error rate of 44%. If the prosecutor dismisses charges against all persons called truthful by the polygraph examiner (or if the jury acquits all such people at the trial), then the use of the polygraph appreciably aids justice, for fully 50% (using Lykken's assumption) of the innocent suspects, who might otherwise have been convicted, go free (at the cost of freeing 10% of the guilty suspects). Of those who are called deceptive by the examiner, the

Table 3

Ground Truth

		Guilty	Innocent	Total
Polygraph Outcome	Deceptive	72	10	82
	Truthful	8	10	18
	Total	80	20	100

incidence of innocent suspects has dropped from 20% down to 12%, and the incidence of guilty suspects has increased from 80% to 88%. Lykken's repeated assertion on page 223 that a polygraph with an average accuracy of 70% results in an effective accuracy at or below chance levels when conditional probability enters into the picture, is wrong, because it fails to appreciate that in such a situation, chance is not 50%, it is whatever the base rates are. The general rule for determining the statistical significance of the accuracy of the polygraph outcomes is this: As soon as the overall accuracy rate (consisting of the accuracy rate with deceptive subjects plus the accuracy rate with truthful subjects) exceeds 100, then the utilization of the polygraph would be statistically useful. Thus, if the polygraph were 80% accurate with deceptive subjects, an accuracy rate with truthful subjects of anything over 20% would be useful in the long run. In view of the more detailed analysis presented here, it would appear that the use of the polygraph by the prosecution in deciding whether to prosecute or not aids the cause of justice.

Lykken is far too intelligent to have overlooked these analyses. One is left with the uncomfortable impression that he appears to have carefully selected which statistics he published, in order to put the polygraph in the worst possible light. This is unfortunate, for it limits the contribution that his book could have made if he had treated the various topics in more detail. Many of the chapters are disappointingly short and superficial. A number of them are only six pages long, even though they deal with topics, such as the relevant-irrelevant test, which do not deserve such abbreviated treatment. In Lykken's treatment of the psychopath, for example, he cited only one study, and presented absolutely no hard data to support his attack on it, although two of the four or five other studies which looked at the effect of psychopathy upon the accuracy of detection of deception tend to support his position.

Chapter 21, a detective story, is by far the longest chapter in the book. It is an entertaining detective story which illustrates what Lykken considers to be the ideal use of the guilty knowledge test which he developed. He introduces a number of useful ideas here, such as making a tape recording of the test questions and trying them out on a naive subject in order to verify that the critical items cannot be recognized by a subject who has no guilty knowledge. But another concept he introduces here is enough to raise the hair on the back of any civil libertarian's neck. He has the police investigators deliberately rearrange the evidence--such as the position of the corpse--and photograph the false evidence. For example, the corpse is photographed in a variety of positions and locations,

to determine if the person taking the polygraph test can recognize which was the true position and location. If the police were to take Lykken's suggestion to heart, it would expose them to a devastating cross examination which could jeopardize their credibility regarding the evidence they were submitting as exhibits at the trial.

Lykken also makes a number of factual errors throughout his book, but compared to the major errors, some of which were criticized above, they pale into insignificance and will not be detailed here. I feel compelled to mention one, however, as it misstates one of my experimental results. When discussing the accuracy of the PSE voice analyzer Lykken states:

> Barland, in 1975, made simultaneous PSE and polygraph recordings in LCT lie test examinations of 66 criminal suspects ... The PSE decision as to Truthful or Deceptive agreed with the polygraph outcome 53% of the time and with independent judicial outcome 47% of the time, where 50% was chance expectancy in both cases. (An earlier report by Barland in 1973, based on a subset of these cases, had suggested positive findings, but this conclusion turned out to have resulted from a statistical error.) (p. 158)

My earlier finding was not due to a statistical error. I had found that there was no correlation between the numerical scores obtained from the PSE charts and those obtained from the polygraph charts, yet there was a significant rate of agreement between dichotomous decisions made on the PSE and polygraph charts, respectively. Because I was disturbed by this result, I constructed a scatter plot of the scores, whereupon the apparently conflicting results were resolved. There was in fact no correlation between the scores--the scatter plot was circular--but the agreement in the dichotomous decisions was caused by the fact that the center of the scatter plot was in quadrant III of the matrix formed by the dichotomous decisions (i.e., deceptive on PSE and deceptive on polygraph), combined with the fact that most of the dichotomous decision pairings also fell in quadrant III. I had mentioned the paradoxical finding and its resolution to Lykken one evening over supper some years ago, and he apparently misconstrued what had been said.

What, then, are the genuine contributions which Lykken makes with this book? There are a number of them, some of which have been alluded to already.

Perhaps the prime contribution Lykken makes is to systematically and explicitly state the assumptions underlying the various forms of polygraph testing. This will serve as a starting point for serious research to confirm, disconfirm, modify and enlarge upon the concepts he puts forth here. The value of this must not be underestimated. Another major contribution is the introduction of further ideas concerning the guilty knowledge test which he originally introduced in 1959. This test offers a number of advantages over previous test procedures in that it is the first testing technique ever developed in which one can calculate the precise probability of a false negative error. In this book, Lykken extends this concept by explaining a practical procedure for estimating false negative errors with this test.

Another major contribution which Lykken makes is to forcefully bring home to the polygraph field that it must engage in a serious research

program to support its contention that the polygraph is highly accurate, and to explore those variables which can affect accuracy. His book is presently viewed by many polygraph examiners, with some justification, as giving a misleading, distorted view of the polygraph technique, a view which is biased and superficial. I strongly disagree with those critics who feel that there is nothing good about it. At the time that the Moss subcommittee attacked the polygraph field in the mid-1960's, the prevailing view within the polygraph community was that it posed a mortal threat to the polygraph. Time has shown that those hearings were one of the best things that happened to the polygraph field. It forced the government to adopt higher standards for the selection of examiners, their training, and procedures they follow. It caused the introduction of quality control procedures within government polygraph operations. Much of the status which the polygraph field enjoys today is a result of the changes introduced following the congressional hearings. So also with Lykken's book. If it causes the polygraph community to critically assess its test procedures and methodology, to engage in serious research in an effort to refute some of Lykken's claims and assertions, the book will have served a most useful purpose. Although its contents may upset and even anger many polygraphists, this book is nevertheless one which deserves to be closely read and studied by the professional examiner.

REFERENCES

1. Barland, G.H. Detection of deception in criminal suspects: A field validation study. (Doctoral dissertation, University of Utah, 1975). Dissertation Abstracts International, 1975, 36. (University Microfilms No. 75-28,874).

2. Barland, G.H. & Raskin, D.C. Detection of deception. In W.F. Prokasy and D.C. Raskin (Eds.) Electrodermal activity in psychological research. New York: Academic Press, 1973, pp. 417-477.

3. Barland, G.H. & Raskin, D.C. An evaluation of field techniques in detection of deception. Psychophysiology, 1975, 12, 321-330.

4. Barland, G.H. & Raskin, D.C. Validity and reliability of polygraph examinations of criminal suspects. (Report No. 76-1, Contract 75-NI-99-0001, U.S. Department of Justice). Salt Lake City, Utah: University of Utah, Department of Psychology, March 1976.

5. Bersh, P.J. A validation study of polygraph examiner judgments. Journal of Applied Psychology, 1969, 53, 339-403.

6. Bitterman, M.E. & Marcuse, F.L. Cardiovascular responses of innocent persons to criminal interrogation. American Journal of Psychology, 1947, 60, 407-412.

7. Blum, R.H. & Osterloh, W. The polygraph examination as a means for detecting truth and falsehood in stories presented by police informants. Journal of Criminal Law, Criminology and Police Science, 1968, 59, 133-137.

8. Brisentine, R.A. Polygraph research in the Army. Polygraph, 1974, 3, 66-80.

9. Department of Defense Joint Services Group. Present status of DoD research on the polygraph. Report dated August 28, 1968. Published in The Use of Polygraphs and Similar Devices by Federal Agencies, Hearings before a subcommittee of the Committee on Government Operations, U.S. House of Representatives, June 4 & 5, 1974, pp. 479-502.

10. Horvath, F.S. The accuracy and reliability of police polygraphic ("lie detector") examiners' judgments of truth and deception: The effect of selected variables. Doctoral dissertation, Michigan State University, 1974.

11. Horvath, F.S. The effect of selected variables on interpretation of polygraph records. Journal of Applied Psychology, 1977, 62, 127-136.

12. Matte, J.A. The art and science of the polygraph technique. Springfield, Illinois: Charles C Thomas, 1980.

13. Orne, M.T., Thackray, R.I. & Paskewitz, D.A. On the detection of deception. In N. Greenfield and R. Sternbach (Eds.) Handbook of Psychophysiology. New York: Holt, Rinehart & Winston, 1972, pp. 743-785.

14. Raskin, D.C. Scientific assessment of the accuracy of detection of deception: A reply to Lykken. Psychophysiology, 1978, 15, 143-147.

15. Raskin, D.C. & Podlesny, J.A. Truth and deception: A reply to Lykken. Psychological Bulletin, 1979, 86, 54-59.

16. Reid, J.E. & Inbau, F.E. Truth and deception. Springfield, Illinois: Charles C Thomas, 1977.

17. Stern, R.M. Did God give us the polygraph? Contemporary Psychology, 1981, 26, 674-675.

18. Waid, W.M. A tremor in the blood(book review). Psychophysiology, 1981, 18, 355-357.

* * * * * *

reduction, control groups, session effects, and much more. Placebo effects are analyzed and placed in their proper context. ("One man's placebo is another man's treatment.")

For examiners who are interested in, and concerned about, biofeedback training as a countermeasure; and who want to become familiar with the biofeedback literature, this book offers a valuable one source, sophisticated discussion of the issues. It is current, relevant, and will clearly serve a valuable heuristic function for a wide variety of disciplines. To knowledgeable and experienced biofeedback practitioners, it offers a state of the art summary of the field and a glimpse into the future. It is a book which offers a wealth of information concerning a wide range of issues.

* * * * * *

Errata

The following errors occurred in the article by Gordon H. Barland in the September 1982 issue of Polygraph. Readers are urged to make the following corrections in their copy of the journal.

1. In the second paragraph on page 258, the sentence "However, a degree in psychology does not necessarily confer an accurate understanding of polygraph procedures and techniques, or that it make one immune ..." is incorrect.

The words "or that" should be replaced by the words "nor does".

2. In the last complete paragraph on page 270, the sentence "This test offers a number of advantages over previous test procedures in that it is the first testing technique ever developed in which one can calculate the precise probability of a false negative error." is incorrect. The word "negative" should be replaced by the word "positive".

3. The second paragraph under the headings ASSUMPTIONS OF POLYGRAPH TESTS on page 266 quotes from Dr. Lykken's book. The first sentence starts "And we must be caseful ..." The sentence should read "And we must be careful ..."

* * * * * *