

Stanley Abrams,¹ Ph.D.

Polygraph Validity and Reliability: A Review

In 1963, a subcommittee of the House Government Operations Committee, under the chairmanship of John E. Moss of California, was directed to investigate Federal use of the polygraph [1]. They drew their conclusions from discussions with both researchers and examiners in the field; a review of the literature; and an evaluation of governmental usage of the polygraph. Two years later they presented their findings, which did not favor the continued use of this instrument. Their report indicated that, "There is no lie detector, neither machine nor human." They further pointed out that this technique had not been proved to be valid in either laboratory research or actual criminal investigations. These statements, inevitably, have had an influence upon polygraph usage. The Federal Government has reduced its use of this technique, and labor unions have relied heavily upon these findings to argue against the employment of this instrument in business and industry. This is in spite of the fact that the polygraph has been demonstrated to aid in recovery of stolen money and material and to act as a deterrent to employee theft [2-4]. It can be assumed, also, that the Moss subcommittee report has influenced the jurists of this country, to the extent that polygraph evidence has generally been ruled inadmissible in the courts.

Since the Moss report, there has been a great deal of criticism of these conclusions. For this reason an attempt will be made in this paper to review the literature on the polygraph which preceded the Moss study and the research which has been reported since that time in order to present a clear picture of the validity and reliability of the polygraph.

Investigations have been carried out in two separate and distinct realms: in the laboratory and in actual life situations. In the former, the approach has been highly varied. Volunteer subjects (Ss), often college students, have attempted deception to such varied activities as denying that they had chosen specifically numbered cards, to the testing of individuals who lied about having taken part in a mock crime. Most experimenters have readily admitted that a great difference exists in the emotional response of a college student who is voluntarily participating in an experiment and an actual criminal suspect whose penalty for being detected in a lie may mean prison, personal embarrassment, or a financial loss. The obvious difference in the fear of detection is great enough in these two procedures to classify them separately. In fact, in Trovillo's view [5], "Simulated emotion in psychology classes, on the lecture platform, in drama, and in experimental laboratories has done more to clutter up and confuse honest polygraphic reporting than all the quackery of 50 years!" Berrien [6] stated, "The subject has little at stake. Within the laboratory the subject's life, liberty, reputation, or property do not depend upon his ability to concoct a satisfactory alibi." Even using such motivating factors as betting the subject (S) that

Received for publication 14 Dec. 1972; revised manuscript received 26 Feb. 1973; accepted for publication 1 March 1973.

¹ Clinical psychologist, Kaiser Foundation, Permanente Mental Health Clinic, 5055 N. Greeley Ave., Portland, Ore. 97217.

he cannot "beat" the instrument smacks of artificiality. According to Kugelmass and Lieblich [7], "Experimental studies . . . may have limited generality for actual lie detection in the field." Wolffe [8] evaluated the effectiveness of the polygraph for its use for security purposes during World War II and reported that, "Experience garnered through laboratory studies of artificial crimes is not a good guarantee of success in dealing with actual criminal investigations." It has been demonstrated by Gustafson and Orne [9] that deception was detected far more readily when motivation to lie was present in the experiment, giving further corroboration to the claims that accuracy would be greater in actual life situations where the motivation to deceive would be magnified.

In addition to the lesser fear and guilt in the laboratory situation, there are a number of factors that diminish the polygraph's effectiveness in the experimental setting. In contrast to an actual criminal investigation, due to the time factor, most researchers have had to limit the number of charts they have administered to each of their Ss. In a law enforcement setting, the suspect is examined and reexamined until the polygrapher can make a firm determination as to truth or deception. In addition, the majority of laboratory studies do not employ trained and experienced examiners, which is undoubtedly the major determinant in attaining an accurate diagnosis. In criminal investigations the examiner may employ a variety of polygraph approaches, but in contrast, most experimenters in the laboratory have used only one technique. Some laboratory studies also have been weakened considerably by the fact that they have used instruments with less than three sensors and some researchers have used only a single measure, such as the galvanic stimulus response (GSR).

The differences that exist between laboratory and field studies must inevitably result in lesser validity and reliability for the former. This, therefore, must be considered when evaluating the findings of these two research procedures.

Regardless of which of the two procedures is employed, there is some common agreement. It is assumed by both experimentalists and polygraph examiners alike that there are certain individuals who are not amenable to testing. Flock [10] indicated that psychopaths, psychopathic liars, and those with circumscribed amnesia are untestable. In the latter group, these individuals purportedly find an action so intolerable that they repress it to the extent that there is no awareness and, therefore, no emotional response to stimuli related to this action. This was demonstrated to some extent by Weinstein et al [11] when hypnosis was employed to force the memory of the performance of a mock crime into the unconscious. When this was accomplished, the polygraph examiner could not indicate with certainty that these Ss had committed the act, and he had to rate their polygrams as inconclusive. In the same study, guilt for an act which had not been committed was induced in Ss by hypnotic procedures, and the polygraph administrator was sufficiently misled by their responses to evaluate them as deceptive when they denied their part in the mock crime. In an actual investigation of a bank employee who gave an indication of deception regarding theft during the periodic testing of employees, Dearman and Smith [12], a psychiatrist and a psychologist, demonstrated that this man was innocent but so laden with guilt that, on the polygraph, he appeared to be lying.

Levitt [13] stated that false negatives could be attained by an individual who does not "regard a lie in the usual way." In this category he included the psychotic, severely neurotic, psychopathic, and retarded. In evaluating the validity of the polygraph with the latter group, Abrams and Weinstein [14] found accuracy below the chance level with those Ss who were less than borderline retarded. In 1923, Larson [15] reported no difficulty in diagnosing accurately the recidivist or "hard boiled crook." These individuals, however, are not necessarily psychopaths. In 1953, Jost [16] showed the patterns of schizophrenics

to have flat blood pressure and GSR tracings while the respiratory tracings were unstable. In contrast to this, he reported that patterns of neurotics were quite reactive. The tracings of psychopaths were found to be flat. Heckel et al [17] found that polygraph testing of individuals with neurotic or psychotic disorders showed results which in some instances might produce erroneous conclusions regarding truth or deception. In disagreement with Jost's findings, they reported that the emotionally disturbed population showed higher reactivity. In this investigation normals were compared with nondelusional and delusional patients who were under psychiatric care. Four polygraph experts evaluated the charts and accurately diagnosed all of the normal Ss as innocent. Little validity or reliability, however, was found among their ratings of the emotionally disturbed, and the greater the degree of disturbance the lower the accuracy. There were both inconclusive and inaccurate ratings of those with emotional problems, while there were completely accurate and no inconclusive ratings among the four raters on the five normals. Ruilman and Gulo [18], like Jost, found less reactivity on the records of psychopaths. There was, however, sufficient variation among the Ss so that some gave perfectly adequate responses.

This writer's findings on schizophrenics, neurotics, and character disorders have been that their responses vary regardless of the diagnostic category. Neurotics as a whole tend to be quite reactive, with blood pressure rising and dropping, and respiration being irregular and jagged. The GSR generally seems less responsive, often dropping consistently. On others, however, it was overly responsive. The same variation was found in the other diagnostic categories, some being flat and others extreme in their reactions. While these individuals are definitely more difficult to evaluate, and it is easier to err, neurotics and character disorders can still be evaluated on the polygraph. (If, however, the psychotic's thought disturbance is so severe as to distort reality to the extent that his perception of the world is inaccurate, his judgment impaired, and his emotions so inappropriate as to make fear of detection or guilt over an action meaningless, he is obviously untestable.) It should be remembered that there are those who, in spite of being psychotic, are sufficiently in contact with reality to be fully cognizant of their actions and fearful of the punishment that might be meted out for a criminal action. Certainly, professionals in the mental health field are more than aware of the emotional responses of psychotics who are quite capable of expressing panic, rage, or profound depression, just as there are others who are completely flat. Generally, the polygraph examiner can determine which Ss are not testable by either their behavior or the pattern of their physiologic responses on the polygrams.

These findings indicate that not everyone can be accurately evaluated through the polygraph procedure. Ss with certain medical, drug, and psychiatric conditions cannot and should not be examined. The polygraph examiner is not infallible, but one should be aware that neither are those expert witnesses in pathology, radiology, psychiatry, ballistics, fingerprinting, or even radar, all of whom are admissible into the courts.

A history of lie detection and some of the early experimentation is described in detail by Trovillo [19,20], but for the purposes of this paper, it suffices to begin with Benussi [21]. In 1914, using respiration alone, he reported considerable success in detecting lies by comparing the inspiration and expiration rate of Ss after telling the truth and after lying. In 1926, Landis and Wiley [22] reexamined this procedure along with their use of a cardiopneumo technique. In an attempt at testing the lower limits of their approach, where detection would be most difficult, they had Ss choose cards on which there were geometric figures and then lie as to which they had chosen. In a second phase of their study, the Ss were presented two stories, one in which they were to lie and another in which they were to be truthful. Their accuracy rates were only at 50 percent for the first portion of their

study, and 57 percent for the second. They believed that the increase in successful judgments was due to the more realistic aspect of the latter part of their investigations. Testing Benussi's inspiration and expiration ratio, accuracy was found to be only at the 63 and 50 percent levels, respectively. Employing either the cardio-pneumo procedure or Benussi's ratio, the results were only slightly above chance expectation.

In contrast to the use of respiration, in 1917 Marston [23] evaluated 107 records using blood pressure alone. With this approach there were errors in only four cases so that 96 percent accuracy was attained. He stated that, "Blood pressure . . . constitutes a practically infallible test of the consciousness of an attitude of deception." In 1921, Langfeld [24] compared a word association technique with blood pressure as a means of differentiating truth from deception. While the word association test was not that efficient in detecting lying associated with a mock crime, "Blood pressure easily picked out the guilty subjects." Burt [25], in that same year, compared blood pressure with respiration in three separate procedures. The first related to lying or truth telling in response to letters and digits; the second to deception or truth in relating stories; and the last used the mock crime paradigm. In addition, observers were present who attempted to determine truth or deception by the subjects' responses and behavior. The observation technique was found to be far less effective than the systolic blood pressure approach which was 91 percent accurate, and respiration 73 percent accurate. In a similar investigation, Landis and Gullette [25] examined 25 Ss using only systolic blood pressure as their measure. Each S chose one of two stories in which they either lied or told the truth. Their accuracy level was only 55 percent, just slightly better than chance. They also reported that no typical patterns were found for either truth or deception among the tracings.

Ruckmick [27] employed simple, three-letter, monosyllabic words of low affective value, and even with this approach achieved 78 percent correct judgments. This was, in fact, increased to 83 percent when one of the undergraduate students was eliminated as a judge. He reported that certain individuals have a greater aptitude for interpreting results, and indicated also that the speed and number of correct judgments rapidly increased with practice. The former was reaffirmed by Marston [28] in 1921. He instructed his Ss to enter a room and, if they chose, take one or more articles of the 50 there, and do their utmost to convince the examiner of their innocence. As additional motivation, they were allowed to keep any of the articles for which they were able to deceive the examiner. Employing a blood pressure apparatus alone, 74.3 percent accuracy was attained. The author himself, however, made 34 correct judgments on 35 Ss, attaining an accuracy level of 97.1 percent. Of the other examiners, one group attained a perfect score. When they were given an additional 10 Ss to evaluate in order to retest their ability, they were equally successful. Marston concluded that the polygraph was "of considerable practical value even when operated by non-experts." He reported further that some examiners who, by reason of experience or special aptitude, can attain "practically absolute reliability." He reported further that, "In October 1917, at the request of the Psychological Committee of the National Research Council, tests of this type (systolic-diastolic blood pressure) were conducted in the Harvard Laboratory, with a view to determining their value in government service during the war, and were reported upon as having given 100 percent accuracy of judgment under very difficult conditions." Reverend Walter Summers [29] reported that in over 6,000 laboratory experiments and 50 actual cases he had obtained results in the range of 98 to 100 percent accuracy. He stated that he did not obtain perfect accuracy because of the weaknesses inherent in the laboratory situation.

In 1942, MacNitt [30] indicated, "We agree with other authorities that blood pressure changes may be a valid measure in approximately 75 percent of the cases examined."

Studying the effectiveness of the GSR in 194 cases, he attained 99 percent accuracy. Later, using simple word association tests, his results were 75 percent correct. From this he concluded that not only blood pressure measurements were effective, but also that GSR and cardiac rate and amplitude were valid and reliable measures of deception.

In 1959, using the mock crime paradigm, Lykken [31] examined 49 Ss with the GSR and attained 93.9 percent correct classification. He employed electric shocks with each lie that was detected in an attempt to create some degree of fear of detection so as to make it comparable to a real situation. One year later, Lykken [32] reported a second study in which he employed 20 Ss who were first given practice in both inhibiting GSR and producing false GSR responses. In addition, to motivate them, he offered them ten dollars if they could "beat" the test. In this case, 100 percent accuracy was attained. Trovillo [20], however, stated that while GSR was successful in experimental situations, in actual case work it had not been of consistent value.

Baesen et al [33], using a Keeler polygraph, attempted to differentiate between 50 pairs of college students as to those who had guilty knowledge from those who had actually participated in a mock crime. Correct judgments were attained in 86 percent of the cases and the investigators felt that additional testings might "provide an even higher percent." Van Buskirk and Marcuse [34] also used a Keeler polygraph to determine which of four cards were chosen by each of 50 Ss. Only two charts were run on each S and accuracy was attained at the 72 percent level. In a second testing, four charts were run on each S and correct judgments were increased to 84 percent, with both testings being statistically significant at the <0.01 level of confidence. Had those charts that were felt to be indefinite been eliminated, accuracy would have been increased to 92 percent. Thus, he demonstrated that additional testings and the exclusion of inconclusive records would increase successful judgments. Ellson [35], comparing the various sensors, attained a success rate of 80 percent with the GSR alone. When information was given to the S as to whether he misled the examiner, no change occurred in the accuracy. Various responses on the part of the Ss were also studied: saying yes to all questions; saying no to all questions; giving random responses; and giving no response. Lie detection was most effective when the Ss responded truthfully to all questions but the key question. Ellson also reported that the more charts administered, the greater the accuracy. On the first trial, correct judgments were made in 50 percent of the cases, on the second accuracy increased to 62 percent, and on the third success reached 79 percent. He indicated that with lying, differences occurred in pulse rate, systolic blood pressure, breathing amplitude, and heart rate, while little change was found in breathing rate and diastolic blood pressure.

When a number of receptors is employed, there is a greater success rate than when only one measure is used. This must be kept in mind when evaluating the results of research which have utilized only one or two sensors as compared to the three or more used in the typical polygraph approach. It should also be recognized that when the results of each sensor are reported separately, accuracy is improved when the findings of the various receptors are combined.

In 1963, Kubis [36] carried out a series of three studies, with mixed results. Using a mock crime situation, an attempt was made to differentiate between a thief, a lookout, and an innocent suspect. He employed five examiners with three months of training but no experience, and attained an average accuracy of 78 percent, which was statistically significant. While general consistency among the examiners was found, there was considerable difference in competency among the five testers. While the examiner with the least success was correct in 73 percent of the cases, those with an apparent aptitude for this technique reached 92 percent accuracy. This high level of success was attained by

some examiners in spite of the fact that examination time was only 20 min, and that only two charts were run by inexperienced individuals. As their experience in testing increased, their success level improved. In the second experiment, Ss were given assumedly classified information and were to lie that they had knowledge of it. The results, however, were not statistically significant. In the last phase of the investigation, three techniques were used to determine if the Ss could mislead the examiners. The first, Yoga, which was essentially a relaxation method, did not confuse the examiners and they were accurate in 80 percent of the cases. The use of muscle tension and visual imagery, however, resulted in effectively misleading the testers to the extent that their correct judgments were reduced to as low as 10 and 25 percent, respectively.

Block et al [37] attempted to reverse the truth-lie response by giving electric shocks each time a subject told the truth and no shock when he lied. They were unsuccessful in accomplishing this with the GSR and they indicated that, "This study demonstrated the relative stability of the criteria of deception and the accuracy of the identification under conditions designed to obscure the criteria and to confuse the diagnosis."

Of all the laboratory research that had been carried out before the Moss report, only Kubis' investigation would cast real doubt upon the validity of the polygraph. He, himself, however, was not that influenced by it, for he stated that, "Accuracy on a complicated lie detector experiment can reach a figure close to or beyond 90 percent for some examiners." Orlansky [38], who reviewed the literature in 1962, stated that, "While the method of lie detection has been used extensively and is regarded favorably by its practitioners, the degree of its validity is still not known. This situation is the result of a failure to collect objective data . . ." In spite of this, he pointed out, "The conventional polygraph, with its three polygraph indicators, obviously can be used to detect deception more accurately than would occur by chance alone. The reported occurrences are rarely below 75 percent and sometimes approach 100 percent."

In an attempt to clarify the validity level of the laboratory studies reported, all of the findings were averaged with the exclusion of those in which specific techniques were employed to mislead the examiner. The mean accuracy level was 81 percent. This is believed by the writer to be impressively high, when it is considered that research, which used only one sensor, and other studies which used now outdated equipment were part of the average. The degree of success was also limited by the weaknesses inherent in the laboratory approach as indicated earlier in this paper. In research carried out in actual criminal investigations previous to the Moss hearings, an even greater degree of validity has been reported, assumedly for reasons stated earlier in comparing laboratory research with studies carried out in the field.

In the majority of criminal investigations, it is impossible to obtain complete verification of the results. However, five studies have been reported in which determination of guilt was obtained, resulting in complete information as to who was deceptive or truthful during the examination.

In 1921, Larson [39], employing a combination of blood pressure and respiratory measures, examined 100 girls to determine those of them responsible for a series of thefts amounting to approximately \$600. All of the suspects but one were seen as non-deceptive. She admitted her guilt so that the 99 innocent and the one guilty girl were accurately diagnosed for a success level of 100 percent. Reporting on a similar investigation of 90 girls living in a college boarding house, Larson [15] again determined which girl was responsible for stealing in the dormitory. In a third study he was able to demonstrate those of 38 college girls who were guilty of shoplifting. Bitterman and Marcuse [40] tested 81 individuals suspected of being involved in a campus dormitory theft of \$100. They

employed a Keeler polygraph and found all of the suspects to be innocent. This was verified later when someone outside of the dormitory was found to be involved. As in Larson's studies, 100 percent accuracy was attained. In another campus theft, reported by Winter [41], 25 girls were examined using both respiratory and blood pressure measures, and through the latter, the thief was identified. Thus, 100 percent correct judgments also were achieved. In these five studies where complete verification was attained 100 percent accuracy was achieved in every instance.

Among those studies in which only partial verification was possible, Marston [28] reported that 20 examinations were given to actual defendants for the Psychological Committee of the National Research Council in 1917, and that all of those which could be verified were accurate. Using blood pressure alone, he indicated that he could detect truth or falsehood of the entire story as well as the parts. In 1942, MacNitt [30] tested 59 embezzlement cases and obtained 97 percent accuracy with the GSR alone. He indicated that the cardiac rate and amplitude were also valid and reliable measures of deception.

In studying larger samples, Inbau and Reid [42] were able to verify only three errors in 4,093 individuals who were tested. Of those found to be deceptive, 97 percent of those which could be verified were found to be accurate. Trovillo [5] reported the findings of nine police departments, totaling 7,622 examinations. Of these, 8.2 percent were classified as indefinite. In the case of those found deceptive, 64 percent confessed. The Seattle and the Chicago Police Departments reported 3.3 and 2 percent known error, respectively. In a second investigation, Trovillo [43] reported that between the years of 1930 and 1941, the Chicago Scientific Crime Detection Laboratory examined 1,127 Ss, 40 percent of whom were diagnosed as guilty, 40 percent were believed to be innocent, and 20 percent of the tests were inconclusive. Inaccuracies were found in 2 percent while 85 percent of those labeled guilty were verified as such. In a study by Lyon [44], 100 juveniles were tested, with verification being accomplished in 40 cases. All of these were shown to be accurately diagnosed. McLaughlin [45] indicated that 65 percent of those suspects found deceptive by the Texas Department of Public Safety were found to be guilty. Of 4,141 suspects tested by the Michigan State Police, Langley [45] reported that guilty knowledge was diagnosed in 1,706 Ss and verified in 51 percent of these. Of the total, 6 percent of the tests were classified as inconclusive, 52 percent innocent, and the remaining 42 percent guilty. Smallwood [47] reported on 4,000 examinations administered over a 3-year period by the Wichita Police Department. Of these tests where deception was indicated, 55.1 percent confessed and 41.2 percent of the individuals who did not admit guilt were found so by the courts. He reported that, "Trial judges the country over, convinced of its merits, have admitted the results of lie detector tests in evidence in unappealed and therefore unreported cases." In another large scale investigation, Trovillo [19] indicated that 2,131 Ss were examined on the polygraph at the Scientific Crime Detection Laboratory of Northwestern University between the years of 1935 and 1938. Of these, only 12 errors in diagnosis of guilt or innocence have been verified. Chatham [48] reported that, "In our own experience more than 100,000 polygraph examinations (90 percent of which were personnel examinations) had recently been evaluated and revealed that in personnel work the proved margin of error was less than 1 percent and uninterpretable records did not exceed 2 percent." Maudet [49], in a French journal, reported that at the Centre de Poitiers of the American Army, 137 suspects were tested. Of these, 57 were found to be lying, 75 truthful, and 5 indefinite. A total of 88 percent of those found deceptive admitted their guilt. Orlansky [38] summarized the examinations performed by the United States Army Military Police and of the 1,302 Ss who demonstrated deceptive responses, 52 percent admitted their guilt, 11.8 percent were shown to be guilty by further investigation,

and in 27.5 percent guilt could not be substantiated. He also visited five government organizations and although he was not able to obtain any data for investigation, he was informed that in preemployment procedures where verification was attained, accuracy was between 95 and 97 percent, and failures ranged from 0.1 to 3 percent.

Hardman [50] stated that, "Though it is admittedly not an infallible device for ascertaining the truth, its great usefulness is seldom questioned in the field." He reported that over 10,000 extrajudicial examinations had been conducted and of those that had been found deceptive, 75 percent had confessed their guilt upon completion of the second test.

McEvoy [2], in describing the polygraph's use in business, stated that some of the world's largest department stores had employed this instrument for a 6-year period. They had apprehended 90 percent of the guilty and had never convicted one who was innocent. He reported that an official of a large detective agency indicated that 50 percent of the cases can be solved by direct investigation, but in others great difficulty is met. Here they had used the polygraph and found it to be 90 percent accurate. Lloyd's of London is sufficiently impressed with this instrument, McEvoy indicated, that they reduce the premium on fidelity bonds to banks which administer periodic polygraph examinations to their employees.

Wolfe [8], at the request of the Emergency Committee in Psychology of the National Research Council for World War II, surveyed the literature, and communicated with those with extensive experience, including research psychologists and those associated private laboratories. He concluded that, "With highly competent and well-trained operators, a record of approximately 90 percent correct can be predicted." Cureton [51] polled members of the American Psychological Association and related disciplines and polygraph examiners in 1953. His findings indicated that of the examiners, 70 percent felt the polygraph to be highly valid, and 28 percent believed it to be moderately so. Of the professional groups, 41 percent found it highly valid and 57 percent felt it to be moderately valid. Thus, 98 percent of both groups view this technique as being a reasonably valid approach for the detection of lying.

There are no claims in the literature that the polygraph is infallible, but many have indicated that it is accurate in 95 percent of the cases when competent examiners are employed. In summary, what has been reported thus far in the field research appears to corroborate the above statement. Of all those verified, guilty and innocent alike, no error was greater than 3 percent, with the majority being 2 percent or less. There were only five studies in which complete verification was possible, and in each, 100 percent correct diagnoses were made. In all of the other investigations an average of only 65 percent verification could be attained, but of these, 98 percent accuracy was reported. If it can be assumed that in the unverified 35 percent the validity is just as high, then the polygraph will have been demonstrated to be a highly effective tool. Its validity would be greater than that of most psychological or psychiatric diagnostic techniques, the results of which may be reported in the courts.

As was expected, the reported accuracy of field research was at a much higher level than the findings of validity studies in the laboratory. While it is almost meaningless to total and average these findings because of the great discrepancy in experimental paradigms and the instruments employed, it does present an approximation of overall correctness of diagnoses. Of all the laboratory experimentation reported here, regardless of whether just one sensor or more were employed, or whether it was a simple design or a complicated mock crime, the mean accuracy level was 81 percent. It is believed that laboratory research will continue to demonstrate consistently lower accuracy until such time as a research paradigm can be developed that will create a situation that is comparable to an actual

criminal investigation. In some manner, the fear of detection, the tension, the conflict, and the possible feelings of guilt must be all incorporated into the experimental situation. In addition, trained and experienced examiners employing acceptable polygraph instruments must be used as well as attempts made to deal with all of the other weaknesses described earlier. Until that time, laboratory and field research are not comparable.

With statistics at 81 percent accuracy for laboratory studies and 98 percent for field investigations, one would wonder why the Moss committee responded so negatively to the use of the polygraph. Perhaps it was related to Orlandy's conclusion that there was not sufficient controlled research to make a determination at that time. Since the Moss decision, further research has been carried out, but investigation into this technique remains minimal.

In laboratory studies, Kugelmass et al [52] tested the effectiveness of the GSR under two conditions, that of responding yes to all questions and comparing this with saying no. Using 27 Ss, his results were statistically significant of the <0.001 level for detecting deception. To the yes answers he attained 70 percent accuracy and 60 percent for the no responses. He found no statistically significant difference in detecting deception regardless of whether the Ss had responded truthfully or lied to the key item. In another investigation, Davidson [53] divided 48 Ss into 12 groups, each involved with a different mock crime. Of the Ss, three were highly motivated by a reward of from \$25. to \$50. if they could mislead the examiner, while the other five Ss were to receive only \$1.00. In contrast to Gustafson and Orne's [9] findings, no significant difference was found in being able to detect guilt between the highly motivated and the poorly motivated groups. Of the four Ss in each group, the procedure was developed so that one S was successful in committing the mock crime, one failed, one did not try, and one was innocent. Employing only the GSR, 98 percent overall accuracy was obtained. All 36 innocent Ss were correctly diagnosed while 11 of the 13 guilty were detected. Blum [54] tested the effectiveness of the polygraph in diagnosing true from false items of information presented to law enforcement agencies by police informants. A total of 20 informants was instructed as to which story to tell. Some of these stories were completely true and others were completely false. Still others had both truthful and deceptive information contained in them. A Stoelting polygraph was employed with the total examination limited to 2 h. Of those nine Ss who were completely truthful, all were accurately diagnosed and all 11 Ss who lied completely or whose stories were partly deceptive were correctly interpreted for an accuracy level of 100 percent. Some errors were made on portions of stories where the stories were made up of a mixture of truth and falsehood.

Employing a mock crime paradigm, Barland and Raskin [55] tested 72 Ss suspected of taking \$10. To increase motivation, they were allowed to keep the money if they could mislead the examiner. Excluding the inconclusive ratings, 81 percent of the Ss were correctly classified after three charts were administered. This was statistically significant at the <0.001 level. Another six examiners evaluated the charts without having seen the examinations and were successful in 79.3 percent of the cases which was also significant at the <0.001 level.

Horvath and Reid [56], in an interesting combination of laboratory and field research, employed ten polygraph examiners, seven of whom had been testing for over one year and three who had been in the field for only four to six months. Evaluating the records without seeing the examinations, they each studied 40 Ss, 20 of whom were verified as deceptive and 20 as innocent. The accuracy level for all 10 examiners was 87.8 percent, but the more experienced attained success in 91.4 percent of the cases as in contrast to

the 79.1 percent for the inexperienced. It was assumed that greater validity would have been achieved if the examiners had been able to observe the actual testing.

The accuracy level for laboratory research carried out after the Moss study was 83 percent. This was essentially the same as the 81 percent average success level attained in laboratory research prior to 1963. A second but unfortunate similarity lies in the fact that relatively little research has been carried out in the last nine years.

Results from investigations carried out in the field also are lacking in number of studies done. Also notable is the almost complete lack of research from other countries. Almost all of the writing from France, Germany, England, and other European countries is only a review of what has been carried out in the United States. It is quite probable that Russia has done extensive work in this area but none of it has reached the available literature in the United States. The two exceptions to this are Israel and Japan. In the case of the latter, considerable study has been carried out. This author is in the process of accumulating their findings to make them available in this country. One paper, published in the *Interpol (International Criminal Police Review) Journal* [57] in 1966, indicated that considerable use of the polygraph is made by Japanese law enforcement agencies and the results at that time were being accepted into the lower courts as evidence. There was no indication of their validity findings except, "Many subjects diagnosed as 'positive' confess their crime."

In communication with Dr. Makoto Shimizo, Director of the Polygraph program, Dr. Shoei Iseki, Director of the National Institute of Police Science, and Mr. Akihiro Suzuki, a teacher and researcher of the Polygraph Institute in Tokyo, they reported the same difficulties in determining validity because of the inability to verify results in criminal investigations. They also believed accuracy varied greatly, depending upon the expertise of the individual examiner. Further reports of their investigations must, however, await translation.

Judge Pfaff [58], of the Superior Court of Los Angeles County, reported on a controlled study on the use of the polygraph in domestic relations cases. He indicated that this instrument is an invaluable aid in the administration of justice when it is employed by a competent examiner. Michelson [59], in a Canadian journal, presented a case of arson which is usually difficult to solve. He stated that the polygraph was particularly valuable in arson cases because there are few, if any, direct clues that have not been destroyed by the fire. The suspect, in this instance, was found to be deceptive on the polygraph in spite of severe depression and symptoms which at times reached paranoid proportions. There are, however, many reports of individual cases in which both the innocent and the guilty were diagnosed accurately through the polygraph approach. Individual reports of this nature, unfortunately, have little value in determining the actual validity of this approach.

Reid and Inbau [60] reported that John E. Reid and Associates, a private polygraph laboratory in Chicago, has a verified error of less than 1 percent. They indicated that in comparing the polygraph with other judicially accepted techniques, this instrument fares well. They quote Pasamanick, Dinitz, and Lefton [61], who stated that, "Any number of studies have indicated that psychiatric diagnosis at present is so unreliable as to merit serious question when classifying, treating, and studying patient behavior and outcome." In spite of this, psychiatrists are admitted into court as expert witnesses.

The literature discussing the use of the polygraph in business continued to present information that indicates not only the great value of this approach in apprehending those who steal, but also in reducing losses when the polygraph is administered regularly over the years. Menocal [3] points out the savings to employers through the reduction in the cost of fidelity insurance premiums when the polygraph is used to screen new em-

ployees. *Business Week* [4] described the use of this device by a supermarket when it was discovered that employees were taking home one to two dollars in cash or merchandise each week. This small amount, however, totaled one and one-half million dollars per year. Reexamination of these employees six months later demonstrated that only 3 percent of those who stole continued this practice when they knew further testing would be carried out. In discussing the value of the polygraph in business, Menocal stated that "Irrespective of the lie detector's scientific accuracy, it works—it reduces theft losses."

The finding that the polygraph technique is a valid instrument for measuring deception is meaningless unless there is also high reliability. While validity is the degree to which the polygraph measures what it purports to, reliability is an indication of the degree of consistency among polygraph examinations or examiners. It may be the extent of agreement between different examiners in rating the same polygraph record or the consistency of findings among repeated testings of the same S. Unfortunately, there has been even less investigation of the polygraph's reliability than its validity.

All of the investigations of reliability also have evaluated the validity of the polygraph and, therefore, have already been presented in the earlier portions of this paper. Thus, the descriptions of the reliability studies will be kept to a minimum.

Heckel et al [17] reported perfect consistency of judges in rating five normal subjects. They were not so accurate, however, when they rated psychiatric patients. Here, their reliability was quite low. In another investigation, Barland and Raskin [55] tested 72 Ss after committing a mock crime for which they could keep the \$10. stolen if they could mislead the examiner. A total of three charts were administered to each S. The Ss were evaluated by six examiners who were not present during the administration of the tests. There was "very close congruence between the results obtained by the examiner who conducted the examinations and the results obtained by examiners who had not seen the subjects." Weinstein et al [11], in studying the responsiveness of retardates to the polygraph, reported complete agreement between two examiners' judgments in all but one of the cases. Bitterman and Marcuse [40] attained a high degree of agreement among three judges in evaluating 51 Ss involved in a case of theft of \$100. from a campus dormitory. It is of note that a polygraph was not employed in this investigation, only the sensor measuring cardiovascular response. A study aimed specifically at reliability was carried out by Horvath and Reid [56]. They used 40 polygraph charts, 20 of which were verified as guilty and 20 as innocent. A significant difference was found between the seven experienced and the three inexperienced examiners. The experienced polygraphers were accurate in their evaluation on an average of 91.4 percent in contrast to 79.1 percent for the inexperienced. A high degree of consistency was present for those polygraphers who had been in the field for over a year as in contrast to those who were inexperienced interns and had only been in the field from four to six months. One must be impressed, however, with their degree of consistency, considering their newness to the field. Van Buskirk and Marcuse [34] used 50 Ss, requesting that each draw a card on two occasions and lie on one and tell the truth on the second. Then two examiners made a judgment as to where the lie occurred, using a Keeler polygraph. A second judgment was made one month later and the results indicated 84 percent agreement on the cards and 94 percent on the records between these two judgments. Reliability in this investigation was felt to be satisfactory. In another study of the reliability of the polygraph, Kubis [36] employed a mock crime paradigm. Five examiners, in evaluating the 336 Ss, and differentiating thief from lookout, from innocent suspect, he stated, "One is first impressed by the uniformity in the agreement scores." The percentage of agreement ranged from 72 to 87 percent.

The reliability studies, even more than the research on polygraph validity, have been relatively few and unquestionably more experimentation should be carried out. Investigations should emphasize polygraph studies in the field rather than in the laboratory because there is consistent agreement that laboratory findings cannot be safely generalized to actual criminal investigations.

The findings of this review indicate that the polygraph approach is a valid and reliable method for detecting deception. While this is in disagreement with the Moss report, it is believed that the findings of the Moss committee cannot be accepted as being an accurate interpretation of the research that had been carried out at that time. Later experimentation has added further corroboration to the effectiveness of the polygraph. It must be recognized, however, that its consistency and accuracy are dependent entirely upon the competence and experience of the examiner. The equipment itself is secondary to the polygrapher who must use his expertise in the interrogation of the subject, the development of the test questions, the administration of the examination, and the interpretation of the polygrams. If for any reason the examiner lacks the aptitude for any of these functions, errors can occur.

In 1944, the Forensic Section of the American Psychiatric Association adopted a number of resolutions, one of which was that "the machine can give valuable results only in the hands of thoroughly trained physicians and psychologists who will evaluate the data derived by applying other available methods and making use of all independently obtainable evidence" [62]. While this would be a desirable situation, it is not feasible. It behooves the American Polygraph Association to increase its efforts toward setting standards for examiners and polygraph schools. Emphasis should be placed on courses in psychology and physiology as well as other aspects of polygraph techniques. In addition, methods must be devised to evaluate the competence of polygraphers and to make the names of those who meet the standards of the Polygraph Association available to the courts. From this list of names the courts can select qualified examiners and can judge more adequately their testimony.

When the capability of the examiner can be assured, the polygraph technique has greater potential to be of value in law enforcement and in the courts. Limiting the use of this approach would make the truth more difficult, if not impossible to obtain, thus weakening the judicial system at a time when the increasing crime rate is overloading many court dockets. A realistic need exists and it is believed that the polygraph can be a definite asset to the courts. It is time, therefore, for jurists to reconsider admissibility of polygraph findings into the courts under judicial notice.

Many have felt that the effectiveness of the polygraph already has been sufficiently demonstrated and should be admitted into the courts as evidence. Merker [63] has stated that, "As far back as 1952, there was general scientific recognition that the polygraph possesses efficacy, and that reasonable certainty can follow from polygraph tests." In his opinion, "The time has come for the courts to admit polygraph tests into evidence on behalf of a defendant in a criminal case." In agreement with this, Wicker [64] stated that, "Polygraph interrogation is now the best available method of detecting deception. The time has come for the courts to reappraise this type of evidence and perhaps to admit it on the issue of the credibility to be given to the testimony of a key witness." He indicated further that, "There is today in our courtrooms entirely too much intentional perjury . . . usually difficult, and often impossible for even an experienced trial lawyer to expose on cross examination many of the lies of false swearing witnesses." Hardman [50] believed, "The possibility of error inherent in the present day use of lie detectors seems virtually outweighed by the opposing probability of closing the door to truth." Chatham [65] reported that the Committee on State Legislation of the New York State Bar Association

indicated that, "The science of lie detection has reached such stature that its aid should be made available in the process of judicial fact finding." In agreement with these statements, Crane [66] indicated that, "In our present state of congestion of court dockets, any system which would enable us to quickly diagnose 80 percent of the accused certainly ought to be worthy of consideration and adoption."

References

- [1] U. S. Congress, House Committee on Government Operations: "Use of Polygraphs as 'Lie Detectors' by the Federal Government," (89th Congress, 1st Session) House Report No. 198, 22 March 1965.
- [2] McEvoy, J. P., "The Lie Detector Goes into Business," *The Reader's Digest*, 15 Jan. 1941, pp. 69-72.
- [3] Menocal, A. M., III, "Lie Detectors in Private Employment: A Proposal for Balancing Interests," *George Washington Law Review*, Vol. 33, 1965, pp. 932-954.
- [4] "Lie Detectors for Employees," *Business Week Magazine*, 16 Sept. 1939.
- [5] Trovillo, P. V., "Scientific Proof of Credibility," *Tennessee Law Review*, Vol. 22, 1953, pp. 743-766.
- [6] Berrien, F. K., "A Note on Laboratory Studies of Deception," *Journal of Experimental Psychology*, Vol. 24, 1939, pp. 542-546.
- [7] Kugelmass, S. and Lieblich, I., "Effects of Realistic Stress and Procedural Interference in Experimental Lie Detection," *Journal of Applied Psychology*, Vol. 50, 1966, pp. 211-216.
- [8] Wolfe, D., "The Lie Detector: Methods for the Detection of Deception," Confidential Report to Emergency Committee, revised ed., 8 Oct. 1941.
- [9] Gustafson, L. W. and Orne, M. R., "Effects of Heightened Motivation on the Detection of Deception," *Journal of Applied Psychology*, Vol. 47, 1963, pp. 408-411.
- [10] Floch, M., "Limitations of the Lie Detector," *Journal of Criminal Law and Criminology*, Vol. 40, 1950, pp. 651-652.
- [11] Weinstein, E. J., Abrams, S., and Gibbons, D., "The Validity of the Polygraph with Hypnotically Induced Repression and Guilt," *American Journal of Psychiatry*, Vol. 126, 1970, pp. 143-146.
- [12] Dearman, H. B. and Smith, B. M., "Unconscious Motivation and the Polygraph Test," *American Journal of Psychiatry*, Vol. 8, 1963, pp. 1017-1020.
- [13] Levitt, E. E., "Scientific Evaluation of the 'Lie Detector,'" *Iowa Law Review*, Vol. 40, 1955, pp. 440-458.
- [14] Abrams, S. and Weinstein, E. J., "The Validity of the Polygraph with Retardates," *Journal of Criminal Law and Criminology*, in press.
- [15] Larson, J. A., *Lying and Its Detection*, Patterson Smith, Montclair, New Jersey, 1969.
- [16] Jost, H., "The Use of Polygraph Techniques in Psychophysiological Research and Clinical Psychology," Associated Research, Inc., Chicago, 1953.
- [17] Heckel, R. V., Brokaw, H. C., Salzberg, H. D., and Wiggins, S. L., "Polygraphic Variations in Reactivity between Delusional, Non-delusional, and Control Groups in a 'Crime' Situation," *Journal of Criminal Law and Criminology*, Vol. 53, 1962, pp. 380-383.
- [18] Ruitman, C. J. and Gulo, M. J., "Investigation of Autonomic Responses in Psychopathic Personalities," *Southern Medical Journal*, Vol. 43, 1950, pp. 953-956.
- [19] Trovillo, P. V., "A History of Lie Detection," Part I, *Journal of Criminal Law and Criminology*, Vol. 29, 1939, pp. 848-881.
- [20] Trovillo, P. V., "A History of Lie Detection," Part II, *Journal of Criminal Law and Criminology*, Vol. 30, 1939, pp. 104-119.
- [21] Benussi, V., "Die Atmungssymptome der Lüge," *Archiv fuer die gesampfte Psychologie*, Vol. 31, 1914, pp. 244-273.
- [22] Landis, C. and Wiley, L. E., "Changes of Blood Pressure and Respiration During Deception," *Journal of Comparative Psychology*, Vol. 6, 1926, pp. 1-19.
- [23] Marston, W. M., "Systolic Blood Pressure Symptoms of Deception," *Journal of Experimental Psychology*, Vol. 2, 1917, pp. 117-163.
- [24] Langfeld, H. S., "Psychophysical Symptoms of Deception," *Journal of Abnormal and Social Psychology*, Vol. 15, 1920-21, pp. 319-328.
- [25] Burt, H. E., "The Inspiration-Expiration Ratio During Truth and Falsehood," *Journal of Experimental Psychology*, Vol. 4, 1921, pp. 1-23.
- [26] Landis, C. and Gullette, R., "Studies of Emotional Reaction: III. Systolic Blood Pressure and Inspiration-Expiration Ratios," *Journal of Comparative Psychology*, Vol. 5, 1925, pp. 221-253.
- [27] Ruckmick, C. A., "The Truth About the Lie Detector," *Journal of Applied Psychology*, Vol. 22, 1938, pp. 50-58.
- [28] Marston, W. M., "Psychological Possibilities in the Deception Tests," *Journal of Criminal Law and Criminology*, Vol. 11, 1921, pp. 551-570.
- [29] Summers, W. G., "Guilt Distinguished from Complicity," *Psychology Bulletin*, Vol. 33, 1936, p. 787.
- [30] MacNitt, R. D., "In Defense of the Electrodermal Response and Cardiac Amplitude as Measures of Deception," *Journal of Criminal Law and Criminology*, Vol. 33, 1942, pp. 266-275.
- [31] Lykken, D. T., "The GSR in the Detection of Guilt," *Journal of Applied Psychology*, Vol. 43, 1959, pp. 385-388.

- [32] Lykken, D. T., "The Validity of the Guilty Knowledge Technique: The Effects of Faking," *Journal of Applied Psychology*, Vol. 44, 1960, pp. 258-262.
- [33] Baesen, H. V., Chung, C.-M., and Yang, C.-Y., "A Lie-Detector Experiment," *Journal of Criminal Law and Criminology*, Vol. 39, 1948, pp. 532-537.
- [34] Van Buskirk, D. and Marcuse, F. L., "The Nature of Errors in Experimental Lie Detection," *Journal of Experimental Psychology*, Vol. 47, 1954, pp. 187-190.
- [35] Ellson, D. G., "A Report of Research on Detection of Deception," Indiana University Contract No. N6 ONR-18011, Office of Naval Research, 15 Sept. 1952.
- [36] Kubis, J. F., "Studies in Lie Detection Computer Feasibility Considerations," Fordham University RADC-TR 62-205, June 1962.
- [37] Block, J. D., Rourke, F. L., Salpeter, M. M., Tobach, E., Kubis, J. F., and Walch, L., "An Attempt at Reversal of the Truth-Lie Relationship as Measured by the Psychogalvanic Response," *Journal of Psychology*, Vol. 34, 1952, pp. 55-66.
- [38] Orlansky, J., "An Assessment of Lie Detection Capability," (Declassified Version), Institute for Definitive Analysis, Research and Engineering Support Division, Technical Report 62-16, July 1964.
- [39] Larson, J. A., "Modification of the Marston Deception Test," *Journal of Criminal Law and Criminology*, Vol. 12, 1921, pp. 390-399.
- [40] Bitterman, M. E. and Marcuse, F. L., "Minor Studies from the Psychological Laboratory of Cornell University: Cardiovascular Responses of Innocent Persons to Criminal Interrogation," *American Journal of Psychology*, Vol. 60, 1947, pp. 407-412.
- [41] Winter, J. E., "Comparison of the Cardio-pneumo-psychograph and Association Methods in the Detection of Lying in Cases of Theft Among College Students," *Journal of Applied Psychology*, Vol. 20, 1936, pp. 243-248.
- [42] Inbau, F. E. and Reid, J. E., *Lie Detection and Criminal Interrogation*, Williams & Wilkins, Baltimore, Maryland, 1942.
- [43] Trovillo, P. V., "Deception Test Criteria: How One Can Determine Truth and Falsehood from Polygraphic Records," *Journal of Criminal Law and Criminology*, Vol. 33, 1942-43, pp. 338-358.
- [44] Lyon, V. W., "New Deception Tests," *Journal of Genetic Psychology*, Vol. 48, 1936, pp. 494-497.
- [45] McLaughlin, G. H., "The Lie Detector as an Aid in Arson and Criminal Investigation," *Journal of Criminal Law and Criminology*, Vol. 43, 1953, pp. 690-694.
- [46] Langley, L. L., "The Polygraph Lie Detector: Its Physiological Basis, Reliability and Admissibility," *The Alabama Lawyer*, Vol. 16, 1955, pp. 209-224.
- [47] Smallwood, C. L., "Evidence: Lie-Detectors: Discussion and Proposals," *Cornell Law Quarterly*, Vol. 29, 1934, pp. 535-545.
- [48] Chatham, R. B., "The Polygraph Truth Test: Further Developments," *Tennessee Law Review*, Vol. 22, 1953, pp. 916-918.
- [49] Maudet, G., "Le Polygraphe et Son Utilisation en Justice, La Rédaction," 1969, pp. 298-317.
- [50] Hardman, T. P., "Lie Detectors: Extrajudicial Investigations and the Courts," *West Virginia Law Quarterly*, Vol. 45, 1941, pp. 37-46.
- [51] Cureton, E. E., "A Consensus as to the Validity of Polygraph Procedures," *Tennessee Law Review*, Vol. 22, 1953, pp. 728-742.
- [52] Kugelmass, S., Lieblich, I., and Bergman, Z., "The Role of 'Lying' in Psychophysiological Detection," *Psychophysiology*, Vol. 3, 1967, pp. 312-315.
- [53] Davidson, P. O., "Validity of the Guilty-Knowledge Technique: The Effects of Motivation," *Journal of Applied Psychology*, Vol. 52, 1968, pp. 62-65.
- [54] Blum, R. H. and 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*, Vol. 59, 1968, pp. 133-137.
- [55] Bariland, G. H. and Raskin, D. C., "An Experimental Study of Field Techniques in 'Lie Detection'," *Journal of the American Polygraph Association*, Vol. 1, 1972, pp. 22-26.
- [56] Horvath, F. S. and Reid, J. E., "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception," *Journal of Criminal Law, Criminology, and Police Science*, Vol. 62, 1971, pp. 276-281.
- [57] "The Use of the Lie Detector (Polygraph Test) by the Japanese Police," *International Criminal Police Review (Interpol)*, Vol. 96, 1966, pp. 62-66.
- [58] Pfaff, R. S., "The Polygraph: An Invaluable Investigation Aid," *American Bar Association Journal*, Vol. 50, 1964, pp. 1130-1133.
- [59] Michelson, R. D., "Arson and the Polygraph," *Royal Canadian Mounted Police Gazette*, 1958, pp. 19-20.
- [60] Reid, J. E. and Inbau, F. E., *Truth and Deception*, Williams & Wilkins, Baltimore, Maryland, 1966.
- [61] Pasamanick, B., Dinitz, S., and Lefton, M., "Psychiatric Orientation and Its Relation to Diagnosis and Treatment in a Mental Hospital," *American Journal of Psychiatry*, Vol. 116, 1959, pp. 127-132.
- [62] Elinsberg, W., "Forensic Psychology," *Southern California Law Review*, Vol. 19, 1946, pp. 349-409.
- [63] Merker, M. M., "Polygraph for the Defense," *New York State Bar Journal*, Vol. 40, 1968, pp. 569-572.
- [64] Wicker, W., "The Polygraph Truth Test and the Law of Evidence," *Tennessee Law Review*, Vol. 22, 1953, pp. 711-727.
- [65] Chatham, R. B., "The Polygraph Truth Test: Further Developments," *Tennessee Law Review*, Vol. 22, 1953, pp. 916-918.
- [66] Crane, G. W., "The Polygraph Truth Test," *Tennessee Law Review*, Vol. 22, 1953, pp. 771-773.