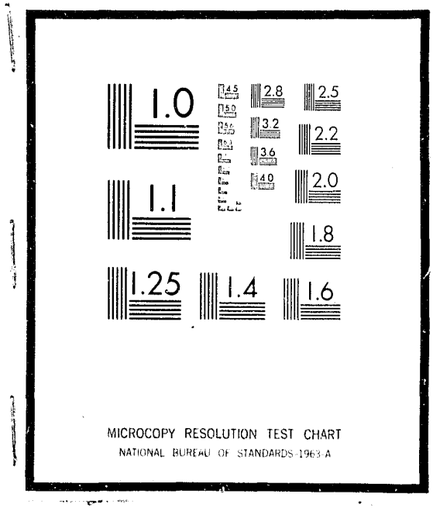

Validity and Reliability of Polygraph Examinations of Criminal Suspects

Report No. 76-1
Contract 75-NI-99-0001
National Institute of Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
U.S. Department of Justice

Gordon H. Barland, Ph.D.
Project Co-Director

David C. Raskin, Ph.D.
Project Director

Department of Psychology
University of Utah
Salt Lake City, Utah 84112
March 30, 1976

Introduction

Although the polygraph technique has been used for over half a century for testing the veracity of criminal suspects, information about its accuracy is limited. Many laboratory studies have demonstrated that the technique can be highly accurate (see Orne, Thackray, & Paskewitz, 1972 for a review of the literature). However, when trying to determine the accuracy of the technique, results of laboratory studies cannot be generalized to field situations because of many important differences between laboratory and field situations, (Abrams, 1973; Barland & Raskin, 1973). Unfortunately, in field situations it can be so difficult to determine whether the polygraph examiner's decisions were correct, that there has been only one scientifically acceptable study made of the accuracy of the polygraph in real-life situations (Bersh, 1969).

Other problems in determining the accuracy of the polygraph in field situations include the fact that the polygraph examiner has available to him much information concerning the probable guilt or innocence of the subject in addition to the biological information contained in the polygraph charts. Usually the examiner is briefed by the investigators prior to the examination, he often has access to police reports, and he closely observes the subject's appearance and behavior. Moreover, the results of the examination often influence the subsequent handling of the investigation. The polygraph examination does not occur in isolation, but is embedded within the context of a series of related events. The actual contribution of the polygraph charts to the total polygraph examination is often difficult to isolate.

Perhaps for those reasons, there has been only one scientific study of the accuracy of the polygraph technique as it is used today by field practitioners (Bersh, 1969). That study assessed the accuracy of polygraph examiner decisions with criminal suspects examined by U.S. military examiners. A random selection of 324 cases in which the examiners had made definite decisions were presented to a panel of four military attorneys after all references to the outcome of the polygraph examinations had been removed. The panelists were instructed to review each case for completeness of case information and to decide the guilt or innocence of each suspect, rejecting any case in which there was not sufficient evidence. In making their decisions they were told to ignore legal technicalities and to indicate whether or not the suspect had actually committed the act(s) of which he was accused. The decisions of the polygraph examiners were then compared against the panel in the 157 cases where all four panelists' decisions had been unanimous. The polygraph examiners agreed with the unanimous panel in 92.4% of the cases.

The Bersh study is the only field validation of the accuracy of the polygraph technique which employed a stringent external criterion of guilt or innocence. However, the design of the Bersh study has several characteristics which limit the generalizability of the results: (1) The sample of cases may have been biased by the exclusion of cases which one or more of the panelists deemed to have insufficient evidence; (2) the subject population was limited to military personnel on active duty. The distributions of age, education, intelligence, criminal background, and physical fitness of the subjects may have been different from those characteristic of nonmilitary criminal suspects; (3) military examiners usually

have much more investigative material available for study prior to the examination than do most private polygraph examiners, thereby increasing the likelihood that the military examiner's decision may utilize more non-polygraphic information; and (4) all polygraph examinations of suspects were performed for the investigative agency and with the full knowledge by the subject that the outcome could affect the investigation.

The purpose of this study was to extend the Bersh study using criminal suspects tested by a private polygraph examiner and referred for examination by defense attorneys or law enforcement agencies. The criminal suspects were thus drawn from a sample more comparable to that of subjects examined in private applications of the polygraph technique. The panelists used in this study each made decisions on all cases in the sample, using a 5-position scale which indicated their decision and their confidence in it. The case information provided to the examiner prior to the test sometimes included the initial police report of the incident; however, case information was often limited to that provided by a brief telephone conversation with the defense attorney. Finally, the effects of various personality and socioeconomic variables that might influence autonomic responsivity in a realistic detection of deception situation were investigated.

Method

Subjects

A total of 102 subjects were examined on the polygraph (by Gordon H. Barland) at the request of police, defense attorneys, or prosecuting attorneys in Utah and Nevada. All subjects were suspected of having committed a criminal offense, and most were under arrest at the time of

the examination. No victims, witnesses, or persons involved in civil suits were included. Ten subjects were examined on matters in which multiple suspects were given polygraph examinations. There were 92 cases in which each polygraph examination was independent from the others.

The sample consisted of 84 males and 18 females ranging in age from 15 to 54 years (M=28.3). Their level of formal education ranged from 3 to 16 years (M=11.3). Thirty-four subjects reported that they had no previous arrests, and 64 reported from 1 to 75 previous arrests (M=6.3). The personal background data were lost for 4 subjects. Twenty-seven subjects reported that they had been previously examined on the polygraph by a different polygraph examiner, either on the same issue or on an unrelated crime.

Apparatus

Seven different models of Keeler, Stoelting, and Lafayette field polygraphs were employed at various times and locations in this study. A Keeler 6317 battery operated polygraph, described previously (Barland & Raskin, 1975), was used with 34 subjects. Because of the poor operating condition of the Keeler, it was replaced with a Lafayette 4-channel field polygraph with an electronically-enhanced cardio, with which 29 subjects were examined. A 4-channel Stoelting polygraph with a wet cardio activity monitor (CAM) was used with 29 subjects. Three other Stoelting and one other Lafayette polygraphs were used on 10 subjects.

In order to standardize the data, only the commonly-used measures of respiration, skin resistance response (SRR), and cardiovascular activity were utilized. Respiration was recorded by a pneumatic tube positioned on the abdomen near the diaphragm or on the thorax so as to produce a pen

excursion between 1.2 and 2.5 cm. The SRR was recorded from two stainless steel electrodes attached to the volar surface of the index and fourth fingers of the subject's nondominant hand. No electrode paste was used. Cardiovascular activity was measured by an occlusive cuff located on the upper part of the subject's dominant arm. The cuff was inflated to a pressure sufficient to make a pen excursion of approximately 1.2-2.5 cm with the dicrotic notch in the middle of the descending limb of the pulse wave. The pressure was usually about 90 mm Hg, except with the Lafayette polygraphs which permitted pressure of approximately 50-60 mm Hg to be used.

A variety of test rooms and locations was used. In all cases the rooms were relatively small and plainly furnished in the manner described by Reid and Inbau (1966, p. 5). Testing was normally conducted at the University of Utah, but occasionally other examiners' facilities, police stations, or attorneys' offices were used. The location and the occasional outside noises had no apparent effect on autonomic responsivity. The examiner was usually alone with the suspect throughout the examinations.

Procedure

Each suspect was examined individually for at least three hours. Prior to the examination the examiner was briefed by the person requesting the examination concerning the case facts and the issues to be covered. The briefing was often conducted by telephone. When each suspect appeared for the examination, he was advised of his constitutional rights. He signed a standard form which indicated he was aware of his rights and agreed to be examined on the polygraph. The examiner then obtained information from the suspect concerning his family background, medical

history, and miscellaneous personal information such as age, educational level, and number of previous arrests and polygraph examinations. Several scales (see below) from the Minnesota Multiphasic Personality Inventory (MMPI) were orally administered to the suspect for research purposes.

The polygraph technique was then explained to each suspect in terms of the autonomic control of homeostasis of the human body, the "fight or flight" response to threatening stimuli, and an explanation of the physiological measures recorded by the polygraph. After being cautioned to tell only the truth, the suspect was asked to explain what he knew of the crime of which he was suspected. After hearing his version of the situation, the questions to be asked during the examination were formulated with the help of the suspect and were reviewed with him. The federal modification of the Backster zone-comparison control-question test was used in all but one case in which the 1966 version of the Reid technique was used. For a more detailed description of the federal modification of the Backster zone comparison test, see Barland and Raskin (1973, p. 431, 1975).

The suspect was attached to the polygraph, and the test questions were asked a minimum of three times as required by Utah law. If the examiner was unable to make a decision after the third chart, the questions were repeated until the examiner either felt he had enough biological information to make a definite decision or that further testing would be of no use. The maximum number of charts obtained was six. Each chart consisted of 10 to 13 questions and lasted approximately about 4-5 minutes. The questions were asked in a monotone to prevent autonomic responses due to voice inflections. The questions, which were worded so as to be

completely answered with a "yes" or "no," were separated by intervals of 15 to 20 seconds to allow for recovery of any responses to the preceding question. Between charts the suspect was encouraged to make any changes he wished concerning the wording of the questions.

The polygraph charts were then numerically evaluated as described below, usually with the suspect observing. The suspect was informed of the examiner's decision and released or, if found deceptive, questioned. A more detailed description of the sequence and purpose of the examination procedure is found in Barland and Raskin (1973).

Quantification of the polygraph data

The charts were scored numerically at the time of the polygraph examination using field scoring techniques (Barland & Raskin, 1975). The magnitude and duration of responses in each physiological measure were compared for each relevant question and its adjacent control question. A typical test consisted of a series of 10 questions asked on each of three charts. Of the 10 questions, there were three pairs of relevant and control questions. Each physiological measure was scored on a 7-point scale which ranged from +3 to -3 for each pair of control and relevant questions on each chart. If the response to the control question was greater than that to the relevant question, the pair of responses was scored in the plus direction, indicating probable truthfulness to the relevant question. If the response to the relevant question was greater than to the control question, it was scored in the minus direction, indicating probable deception to the relevant question. The magnitude of the difference between the control question and relevant question responses determined whether the difference was scored as a 0, 1, 2, or 3. The

assignment of a score to the difference was done on the basis of criteria used by the federal government (USAMPS, 1970). The total score was obtained by summing the scores over all question pairs, charts, and measures, assigning equal weight to each score. If the total score was +6 or higher, the suspect was judged to be truthul; if it was -6 or lower, the suspect was generally judged deceptive; and a score between ±5 was generally considered inconclusive. It should be noted that throughout this report the terms "truthful" and "deceptive" refer to the conclusions reached by the polygraph examiner in assessing the subjects' veracity in responding to the relevant questions. The terms "innocent" and "guilty" refer to the resolution of the ultimate issue and must be based on external criteria such as judicial outcomes, confessions, or expert opinions.

Panel criterion. Each case was investigated by one of several university students employed as investigators. Where possible, each defense and prosecuting attorney was contacted, police reports obtained, and as many facts as possible bearing upon each case were collected. A summary of the evidence in each case was prepared by an assistant who had no other knowledge of the cases. Appended to the summary were copies of any available documentation collected by the investigators. In every case the true names and places were altered in order to protect the privacy of the suspect and witnesses. Only two types of information were omitted from the summary or appendices: (1) the outcome of the polygraph examination, and any reference to polygraph examinations administered to any suspect, victim, or witness; and (2) the judicial outcome.

If the suspect confessed as a result of the polygraph examination, this information was reported in the summary without reference to the context of the confession. If the suspect subsequently pleaded guilty, this was reported in the summary, together with a comment as to whether or not he had pleaded guilty to a lesser charge in a plea-bargaining situation. The summary and available appendices on each case were submitted to each member of a panel of five legal experts.

The panel consisted of two prosecuting attorneys, one judge, and two defense attorneys. The composition of the panel was designed to ensure that the composite decisions would not be biased toward the defense or the prosecution. The panelists were instructed to disregard legal technicalities and rules of evidence and were asked to render an opinion, if possible, as to whether or not the suspect had committed the act of which he was accused. They reviewed each case file independently of each other and made one of five decisions: definitely innocent, probably innocent, undecided, probably guilty, or definitely guilty. If a panelist recognized any case and had personal knowledge of it, he disqualified himself. An alternate panelist (a defense attorney) reviewed those cases from which a regular panelist had disqualified himself.

When the case files were returned, the experimenter coverted the decisions to numbers on a 5-point scale ranging from +2 (definitely innocent) to -2 (definitely guilty). The five scores on each case were then summed to yield a total panel score, against which the results of the polygraph examination were compared. This procedure differed from the Bersh (1969) study which compared the polygraph against unanimous, dichotomous decisions by the panel.

Judicial outcome criterion

The outcome of each case within the judicial system was assigned to one of five categories. Independent guilty included those cases where the judicial outcome indicated guilt and the trier of fact was unaware of the outcome of the polygraph examination. Those included verdicts of guilty and pleas of guilty to either the original charge or to a reduced charge which included the specific act covered in the polygraph examination. Dependent guilty included those cases in which a suspect was convicted in a trial in which the results of the polygraph examination were entered as evidence. Inconclusive included cases in which the suspect was not tried because of insufficient evidence to prosecute, in which the suspect pleaded guilty to a reduced charge which had not been covered in the polygraph examination, or in which the case was still pending at the time the data were analyzed. Dependent innocent consisted of those cases where the prosecution dropped the charges following a finding by the polygraph examiner of no deception by the suspect. Independent innocent consisted of cases where the suspect was acquitted by a judge or jury following a trial in which no mention was made of the polygraph.

Behavior cues

The verbal and nonverbal patterns of behavior were carefully observed by the examiner during the pretest interview to determine the extent to which pretest behavior can predict the outcome of the polygraph examination. In addition to spontaneous cues arising from the suspect and his interaction with the examiner, the examiner asked specific questions known as the Reid questions, which have been previously found to be helpful in differentiating between truthful and deceptive persons (Horvath, 1973). Only the questions

appropriate to each situation were asked during the pretest interview.

The examiner made two separate predictions of the outcome of the test. The first prediction was made immediately after the advisement of rights at the outset of the pretest interview (i.e., after the examiner had observed the suspect for about ten minutes). The second prediction was made at the conclusion of the pretest interview after the suspect had explained what he knew about the incident under investigation and after the questions to be asked on the polygraph examination had been formulated and reviewed. At this point the examiner had observed the suspect for approximately 60-90 minutes. Both decisions were recorded on an 11-point scale ranging from +5 to -5, thereby showing not only the direction of the decision but also the examiner's confidence in those decisions.

It is possible that the close observation of the suspect's behavior and the deliberate predictions formulated by the examiner might change the examiner's interaction with the suspect to the extent that the suspect might feel that the examiner is biased. That might alter the suspect's physiological responsivity during the testing portion of the examination, resulting in an increased error rate. To check this hypothesis, the use of the Reid interview questions and the written predictions by the examiner were discontinued midway during the study in order to permit a comparison between the suspects for whom predictions were made and those for whom no predictions were made.

Personality variables

During the pretest interview several scales from the MMPI were verbally administered to each suspect. They included the Lie scale (L), K-scale,

Psychopathic Deviancy (Pd) scale, Hypochondriasis (Hs) scale, and Depression (D) scale. The scales could not be administered to seven of the suspects because of intellectual or language difficulties. The L, Hs, and D scales were discontinued after having been administered to 52 suspects in order to permit the use of the Hypomania (Ma) scale, the results of which will be reported elsewhere.

The L-scale was administered to determine if subjects with high scores would attempt deception on the polygraph more frequently than low scorers. The K-scale was administered partially to supplement the L-scale, but primarily because it is used to correct the Pd and Hs scales. The Pd scale was administered to examine the possibility that attempted deception by psychopaths is more difficult to detect with the polygraph than is deception by nonpsychopaths. The Hs scale was administered to check the hypothesis that false positive polygraph errors may occur in hypochondriacs more frequently than in nonhypochondriacs. The rationale for using the D-scale was that depression may be correlated with low autonomic responsivity, thereby increasing the number of inconclusive examinations.

Miscellaneous variables

A variety of populational variables potentially capable of affecting the autonomic responsivity of criminal suspects in a detection of deception situation were observed. The variables included the type of crime involved, sex, age, number of previous arrests, number of previous polygraph examinations, educational level, and the depth of religious beliefs. All data used for these analyses, other than crime type and sex, came from self reports by the suspects.

Results[1]

Of the 102 polygraph examinations on different criminal suspects, the examiner concluded the 14 (13.7%) were truthful when they denied committing the act of which they were accused, and 71 (69.6%) were deceptive. The remaining 17 examinations (16.7%) were inconclusive. Excluding the inconclusives, 16.5% of the examinations were truthful and 83.5% were deceptive.

Of the 92 independent cases, the examiner concluded that 13 (14.1%) were truthful, 66 (71.7%) were deceptive, and 13 (14.1%) were inconclusive. The scores of the deceptive subjects ranged from +3 to -36, with a mean of -13.6; the scores of the truthful subjects ranged from 0 to +22, with a mean of +8.5; and those of the inconclusive cases ranged from +1 to -22, with a mean of -6.0. The polygraph scores excluded a nonstandard polygraph channel used on some tests (the cardio activity monitor), and excluded any nonstandard charts such as the Yes Test, Yes-No Test, and SKY questions.

Reliability of chart interpretations

The relationship between autonomic responsivity recorded on the polygraph charts and examiner decisions is very complex, even for examiners who attempt to make rigorous decisions based upon careful, numerical evaluations of the polygraph charts. The polygraph charts in this study were given on-the-spot field evaluations which formed the basis for most examiner decisions. The first 77 sets of polygraph charts obtained in

---

[1]The significance of all results of statistical analyses was evaluated using a .05 rejection region, 2-tailed.

this study were later rescored by the polygraph examiner after about six months. The examiner could recall the decision for each case, but could recall neither the scores of the individual responses on each question pair on each chart, nor the total polygraph score.

The score-rescore correlations for each physiological measure and for the total polygraph scores are shown in Table 1. All of the correlations were reliable.

Table 1

Score-Rescore Correlations of

Polygraph Measures

Measure	r
Respiration	.80
SRR	.92
Cardio	.91
Total Score	.92

Another way of looking at the score-rescore results is to compare dichotomous decisions based solely upon the numerical scores. If the original decisions of the polygraph examiner had been based exclusively upon the scores of the polygraph charts, the decisions would have been somewhat different. The examiner occasionally took other factors into account in addition to the polygraph score when making his decision. These additional factors included the degree of consistency of the responses across questions, charts, and physiological measures, together

with the behavior patterns of the suspect. Table 2 shows the frequencies of deceptive, inconclusive, and truthful decisions actually made in the first 77 cases compared to the decisions that would have been made if the decisions had been based exclusively upon the numerical scores.

Table 2

Frequencies of Actual Examiner Decisions

Versus Decisions Based Exclusively On

Numerical Scores and Rescores

	Decision		
	Truthful	Inconclusive	Deceptive
Actual Decision	10	12	55
Numerical Score	6	10	61
Numerical Rescore	3	21	53

It also shows the decisions that would have been made if the decision had been based exclusively upon the rescores of the polygraph charts. It is evident that the examiner differed from the actual numerical scores by arriving at an inconclusive determination in several cases where the charts clearly showed "deception," either because the subjects were classified by him as "guilt complex" responders or because the pattern of physiological responses was sufficiently inconsistent that the examiner felt that numerical score alone was misleading.

Inspection of Table 2 also discloses that when the polygraph charts were later rescored, the number of inconclusives increased at the expense

of both deceptive and truthful decisions. To test the significance of that shift, the rescores were subtracted from the original polygraph scores. The mean change score of -1.43 was significantly different from zero, $t(76) = 3.23$. Thus, the rescores were regressed toward zero.

Table 3 shows the relationship between decisions based exclusively on the original numerical scores of the polygraph charts and decisions based exclusively on the rescores of the same charts.

Table 3

Decision Relationships Based on the Original Numerical Score and the Rescore of Polygraph Charts

Rescores	Original Scores			
	Deceptive	Inconclusive	Truthful	Total
Deceptive	53	1	0	54
Inconclusive	6	10	4	20
Truthful	0	1	2	3
Total	59	12	6	77

It will be noted that in spite of the more conservative nature of the rescores, 65 (84.4%) of the 77 rescores resulted in the same decision. In no case was a decision reversed; the main changes were from a decision to inconclusive.

Table 4 shows the intercorrelations between the various physiological measures from the original numerical scoring of all polygraph charts obtained

from the first 77 subjects in this study. The only 2 correlations which were not significant were those obtained when respiration was compared to the cardiovascular and SRR scores.

Table 4

Intercorrelations of the Polygraph Measures

	SRR	Cardio	Total
Respiration	.185	.089	.552*
SRR		.395*	.863*
Cardio			.632*

*$p < .05$

In order to further check the reliability of the chart interpretation, all 102 sets of polygraph charts were independently scored by another examiner (David C. Raskin), who had not been informed of the case facts, the appearance or behavior of the subjects, nor any other fact pertaining to the examination. Table 5 shows the relationship between the original examiner's decisions and those decisions that would have been made on the basis of the score of the independent evaluation in which a score of +6 or higher was indicative of truthfulness, -6 or lower was considered deceptive, and between ±5 was inconclusive.

It can be seen that in 77 of the 102 cases (75.5%) both examiners agreed that the charts were truthful, deceptive, or inconclusive. Excluding the cases where either or both of the examiners felt that the charts were inconclusive, the two examiners agreed on 68 of the 69 decisions (98.6%).

Table 5

Comparison of the Original Examiner's Decision and the Blind Evaluation of the Charts By An Independent Examiner

Independent Evaluation	Original Examiner's Decision			
	Truthful	Deceptive	Inconclusive	Total
Truthful	12	1	2	15
Deceptive	0	56	6	62
Inconclusive	2	14	9	25
Total	14	71	17	102

As noted earlier, the original examiner's decisions sometimes took into account factors other than the numerical score of the polygraph charts. Table 6 shows the comparison between the original examiner's numerical scores and the blind evaluation of the charts by the independent examiner. For both sets of scores +6 or higher was judged truthful, -6 or lower was considered deceptive, and between ±5 was inconclusive.

It can be seen that in 86 of the 102 cases (84.3%) both examiners agreed that the charts were truthful, deceptive, or inconclusive. Excluding the cases where either or both of the examiners felt that the charts were inconclusive, the two examiners agreed on the interpretation of 71 of the 71 sets of charts (100%).

Since both examiners scored the charts numerically, a Pearson Product-moment correlation coefficient was calculated to determine the predictability

Table 6

Comparison of the Original Examiner's Chart Evaluation and the Blind Evaluation of the Charts By An Independent Examiner

Independent Evaluation	Original Examiner's Scores			
	Truthful	Deceptive	Inconclusive	Total
Truthful	10	0	5	15
Deceptive	0	61	1	62
Inconclusive	1	9	15	25
Total	11	70	21	102

between the two sets of scores. The correlation was .911 (df=100), which was significant. This means the 83.0% of the variability of the scores is predictable.

Table 7 presents a frequency distribution of the difference scores between the two examiners. The single most common score was a difference of 1 point. The median absolute difference between each pair of scores made by the two examiners was 3.3, and the standard deviation was 3.2. The mean score of the original examiner was -8.99, and that of the independent examiner was -7.02. The difference between the two means was significant when tested by the correlated t-test (t=4.18, df=101).

Accuracy of decisions

Panel criterion. Ten of the 102 original examinations were eliminated from this portion of the study because they were related to other cases

Table 7

Frequency Distribution of the Absolute Values of the Difference Scores Between the Evaluation of the Charts By the Original Examiner and a Blind Evaluation of the Charts By An Independent Examiner. (N=102)

Difference Scores (Absolute Value)	Frequency	Cumulative Frequency	Cumulative Percentage
0	8	8	7.8
1	21	29	28.4
2	13	42	41.2
3	11	53	52.0
4	10	63	61.8
5	8	71	69.6
6	11	82	80.4
7	7	89	87.2
8	1	90	88.2
9	3	93	91.2
10	3	96	94.1
11	4	100	98.0
12	0	100	98.0
13	1	101	99.0
14	1	102	100.0

within the study. Table 8 shows the comparison of the summed panel scores with the original examiner's decisions for the 92 independent cases.

Table 8

Comparison of the Summed Panel Scores With the Decisions of the Original Polygraph Examiner

Sum of Panel Scores		Polygraph Outcome			
		Truthful	Deceptive	Inconclusive	Total
Innocent	10	0	0	0	0
	9	0	1	0	1
	8	1	0	0	1
	7	0	1	0	1
	6	0	1	2	3
	5	1	2	0	3
	4	2	1	0	3
	3	0	1	2	3
	2	1	3	1	5
	1	1	4	0	5
Undecided	0	4	5	2	11
	-1	2	4	2	8
	-2	1	3	0	4
	-3	0	1	0	1
	-4	0	11	2	13
	-5	0	8	0	8
	-6	0	7	2	9
	-7	0	5	0	5
	-8	0	5	0	5
	-9	0	2	0	2
Guilty	-10	0	1	0	1
Total		13	66	13	92

Table 9 shows the number of agreements and disagreements between the polygraph and the panel at various numerical cut-offs required for a panel decision.

Table 9

Rate of Agreement Between the Polygraph Examiner's Decisions and the Panel at Various Cut-offs Required For a Panel Decision

	Panel score required for decision									
	±1	±2	±3	±4	±5	±6	±7	±8	±9	±10
Agreements with polygraph	53	48	44	43	30	21	14	9	3	1
Disagreements with polygraph	17	11	7	6	5	3	2	1	1	0
Total	70	59	51	49	35	24	16	10	4	1
% Agreement	76%	81%	86%	88%	86%	88%	88%	90%	75%	100%
χ^2	18.5	23.2	26.8	27.9	17.9	13.5	9.00	6.40	1.00	1.00
probability	.001	.001	.001	.001	.001	.001	.01	.02	n.s.	n.s.

It can be seen from Table 9 that when the polygraph was compared simply against the direction of the panel sum, excluding only the 22 cases where the panel total was zero, the polygraph decisions agreed with the panel 53 times out of 70 (76%). A chi square test showed that the result was significant (χ^2=18.5, df=1). Fourteen of the 17 disagreements were on suspects the panel considered innocent. As the total panel score

required for a decision was increased, the rate of agreement with the polygraph fluctated around 87%. Chi square tests showed the results were significant for all levels of the panel criteria through ±8.

Another way of looking at the data is to limit panel decisions to those cases where the panel was unanimous. Because there was only one case where all panelists said that a suspect was definitely guilty or innocent, the panel was considered unanimous when all panelists made a decision in the same direction without regard to the confidence of the decision. Of the 22 cases where there was a unanimous panel and the polygraph examiner made a decision, they agreed 19 times (86%). The three disagreements were subjects considered innocent by the panel and found deceptive by the examiner.

Because of the relatively small number of cases in which all five panelists made a decision in the same direction, the examiner's decisions were also compared against those panel decisions in which at least four panelists made a decision in the same direction, the fifth panelist either agreeing or being undecided. Table 10 shows the breakdown of such cases. Excluding the 6 cases in which the examiner made no decision, the examiner agreed with the panel in 36 of 41 cases (88%). All of the disagreements were instances where the examiner found a suspect deceptive and the panel found him to be innocent.

Finally, the examiner's decisions were compared against a criterion in which at least a majority of the panelists agreed on the guilt or innocence of each suspect, the remaining panelists having either no opinion or the opposite opinion. The results are shown in Table 11.

Table 10

Comparison of Examiner Decisions and Panel Decisions

Consisting of at Least 4 Panelists in Agreement

Examiner Decision	Panel Decision		
	Guilty	Innocent	Total
Deceptive	33	5	38
Truthful	0	3	3
Inconclusive	4	2	6
Total	37	10	47

Table 11

Examiner Decisions Compared to Panel Decisions Consisting

of at Least Majority Agreement

Examiner Decisions	Panel Majority Decision			
	Guilty	Innocent	Inconclusive	Total
Deceptive	43	9	14	66
Truthful	0	5	8	8
Inconclusive	4	3	6	13
Total	47	17	28	92

Excluding the 35 cases where either the examiner or the panel was undecided, the examiner agreed with the panel majority in 48 of 57 decisions (84%). Again, the disagreements consisted of subjects judged innocent by

the panel and found deceptive by the examiner. However, more than half the suspects found truthful by the examiner were considered inconclusive by the majority of the panel.

Table 12 shows the agreement between the examiner's numerical evaluation of the polygraph charts (as opposed to his actual decisions) and the panel majority.

Table 12

Comparison of Decisions Based on Examiner's Numerical

Score and a Panel Majority Decision

Examiner's Numerical Score	Panel Majority Decision			
	Guilty	Innocent	Inconclusive	Total
Deceptive	42	8	17	67
Truthful	0	3	6	9
Inconclusive	5	6	5	16
Total	47	17	28	92

Excluding the 39 cases where the examiner or the panel was undecided, the examiner's scores agreed with the panel majority in 45 of 53 cases (85%).

It is also of interest to compare the blind evaluation of the polygraph charts by an independent examiner against the criterion of the panel majority. Table 13 shows the results. In those cases where both the panel and the reviewing examiner made definite decisions, the independent examiner agreed with the panel majority in 44 of 51 cases (86%).

Table 13

Comparison of Decisions Based on Independent Examiner's Numerical Score and a Panel Majority Decisions

Independent Numerical Evaluation	Panel Majority Decision			
	Guilty	Innocent	Inconclusive	Total
Deceptive	39	6	13	58
Truthful	1	5	7	13
Inconclusive	7	6	8	21
Total	47	17	28	92

As was the case in the preceding comparisons, more than half of the suspects found truthful by the independent examiner were inconclusive with the panel, and 6 of the 7 disagreements consisted of deceptive outcomes on subjects considered innocent by the panel.

Judicial outcome criterion. The agreement between the polygraph examiner's decisions and the judicial outcome in those 41 cases in which the judiciary was uninformed that a polygraph examination had been administered is shown in Table 14. In the 37 cases in which the examiner made a definite decision, it agreed with the judicial outcome 31 times (84%). All of the 6 disagreements occurred when a person acquitted by the judicial system had been found deceptive by the examiner.

Table 15 shows the relationship between the examiner's numerical evaluation of the polygraph charts and the judicial outcome. The decisions based on numerical scoring agreed with the judicial outcome in 30 of the

Table 14

Comparison of Examiner Decisions and Independent Judicial Outcomes

Examiner's Decisions	Judicial Outcome		
	Guilty	Innocent	Total
Deceptive	30	6	36
Truthful	0	1	1
Inconclusive	3	1	4
Total	33	8	41

Table 15

Comparison of Examiner's Decisions Based on Numerical Scores and the Independent Judicial Outcomes

Examiner's Numerical Evaluation	Judicial Outcome		
	Guilty	Innocent	Total
Deceptive	29	5	34
Truthful	0	1	1
Inconclusive	4	2	6
Total	33	8	41

35 cases (86%) where the score was sufficient for a definite decision.

Table 16 shows the relationship between the blind evaluation of the polygraph charts by an independent examiner and the judicial outcome.

Table 16

Comparison of Decisions Based on Independent Examiner's Numerical Scores and Independent Judicial Outcomes

Independent Numerical Evaluation	Judicial Outcome		
	Guilty	Innocent	Total
Deceptive	27	4	31
Truthful	0	3	3
Inconclusive	6	1	7
Total	33	8	41

The independent examiner agreed with the judicial outcome in 30 of the 34 cases (88%) where he made a definite decision based on the numerical evaluation. As in the above analyses, all of the disagreements occurred on suspects acquitted by the judicial process and found deceptive by the numerical evaluation of the charts.

It is of interest to compare the judicial outcome and the panel majority criteria. Table 17 shows the extent of agreement between the two criteria against which the accuracy of the polygraph was assessed. In the 35 cases where a definite decision was made by both criteria, there was agreement 31 times (89%). Three of the four disagreements were when persons acquitted by the judicial outcome were considered guilty by the majority of the panel. Only 3 out of the 8 acquitted by the judicial process were considered innocent by the panel.

Table 17

Comparison of Panel Majority Decisions and Judicial Outcomes

Judicial Outcome	Panel Majority Decision			
	Guilty	Innocent	Inconclusive	Total
Guilty	28	1	4	33
Innocent	3	3	2	8
Total	31	4	6	41

Boundaries for the inconclusive region

A post facto manipulation of the boundaries of the inconclusive region was made for the numerical evaluations by the original examiner and for those by the independent examiner. Fig. 1 and Fig. 2 present the percent accuracy of truthful and deceptive decisions and the percent inconclusives for boundaries ranging from zero to ±12 for the two sets of scores using the panel majority decisions for determining accuracy. It can be seen that as the width of the inconclusive region was increased, there was a slight increase in accuracy of deceptive decisions, a moderate increase in accuracy of truthful decisions, and a corresponding rise in the number of inconclusives for both sets of scores. In general the accuracy of truthful and deceptive decisions remained fairly close to 85%. However, the relatively small number of truthful decisions and the less than perfect accuracy of the panel criterion argue against placing a great deal of confidence in the observed changes in percent accuracy of the truthful decisions. There seems to be a lack of any

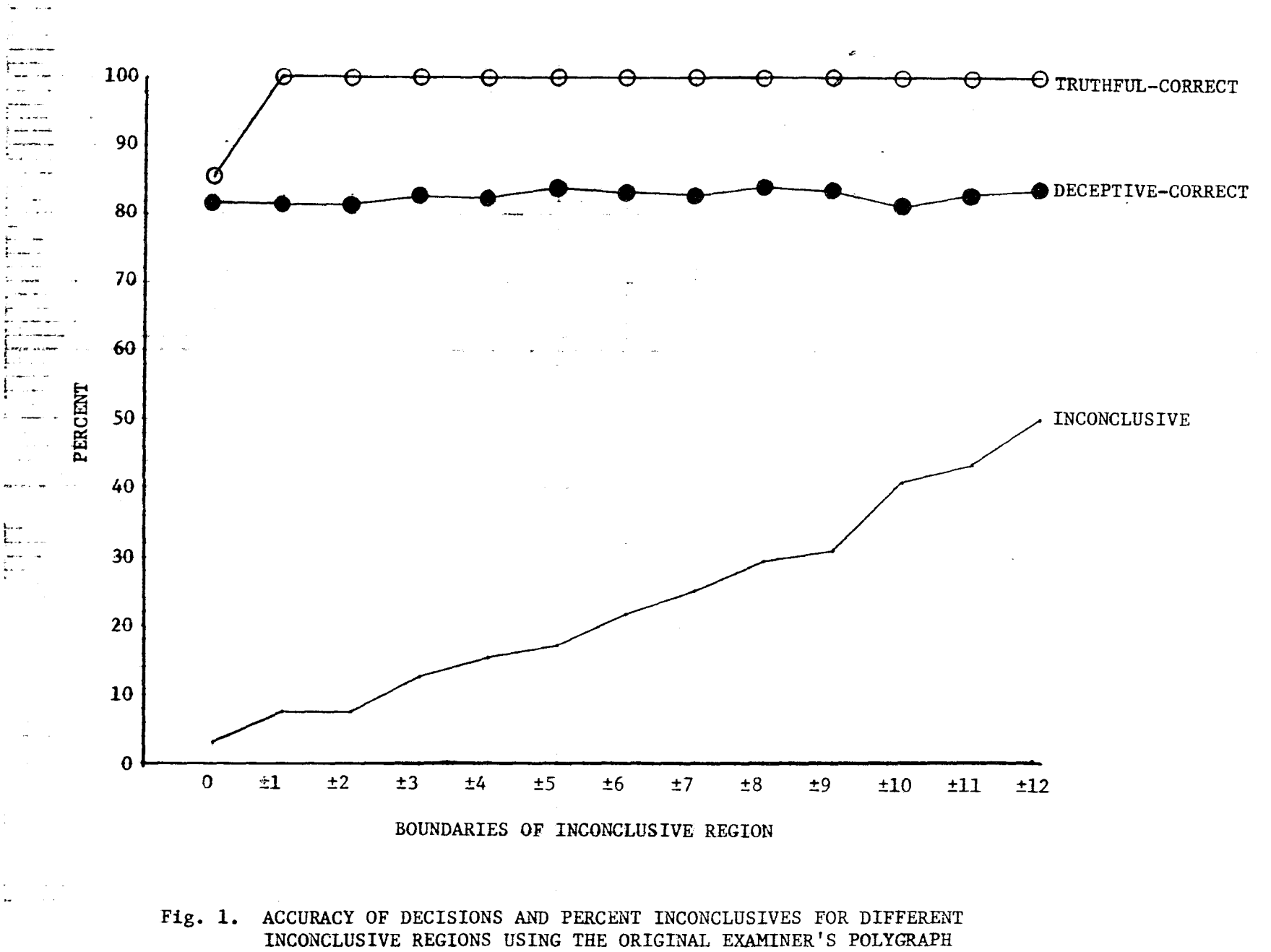

Fig. 1. ACCURACY OF DECISIONS AND PERCENT INCONCLUSIVES FOR DIFFERENT INCONCLUSIVE REGIONS USING THE ORIGINAL EXAMINER'S POLYGRAPH SCORES AND MAJORITY PANEL DECISIONS FOR GUILT AND INNOCENCE.

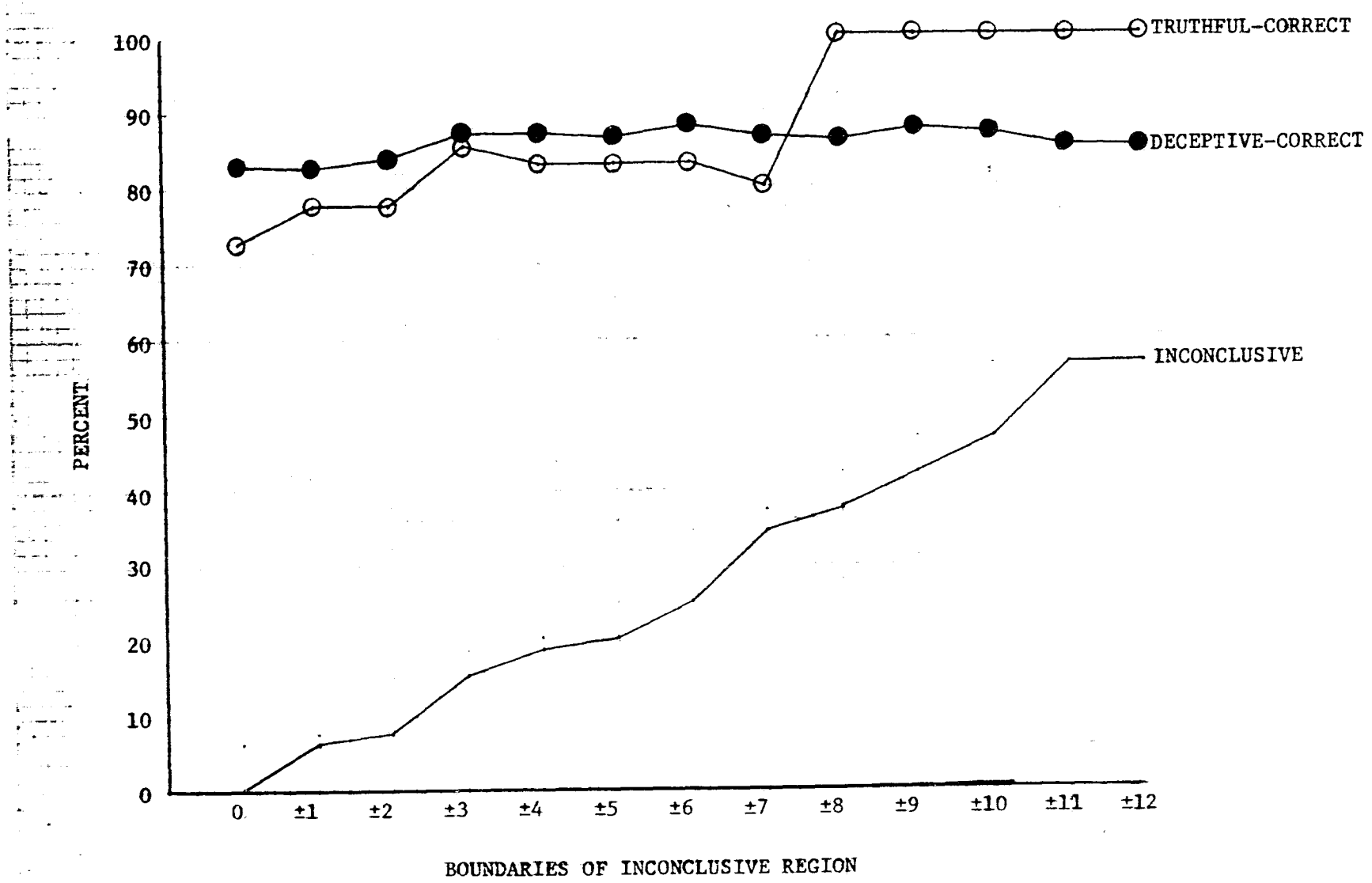

Fig. 2. ACCURACY OF DECISIONS AND PERCENT INCONCLUSIVES FOR DIFFERENT INCONCLUSIVE REGIONS USING THE INDEPENDENT EXAMINER'S POLYGRAPH SCORES AND MAJORITY PANEL DECISIONS FOR GUILT AND INNOCENT.

strong indications that the traditional ±5 boundaries for the inconclusive region should be changed. Perhaps data from studies currently underway in our laboratory will shed additional light on the question.

In order to assess the power of the technique to identify guilty and innocent subjects, the extremeness of the scores was calculated for both types of subjects as defined by the panel majority criterion. The mean of the total scores for the independent examiner for the first three charts was -9.0 for guilty subjects and -2.2 for innocent subjects. The mean scores were significantly more extreme for the guilty subjects, $t(62)=4.08$.

Behavior symptoms

The relationship between the predictions based upon the examiner's observation of the suspect's behavior and the actual outcome of the polygraph examination for each of the two predictions was assessed. The initial observation of the suspect's behavior correctly predicted 17 of the 25 decisions (68%) made on the basis of the polygraph charts, which was not significant ($\chi^2=3.24$). Seven of the 8 incorrect predictions consisted of predicting that a person later found deceptive would be truthful ($p<.05$). Thus, if there were any examiner influence on the suspect's behavior during the initial portion of the interview, there was no evidence that it increased the likelihood of having the suspect respond as if he were nondeceptive. When the initial behavioral predictions were compared to the panel majority in those cases where the panel made a definite decision, they agreed only 10 times out of 18 (56%), which was not significant ($\chi^2=.22$).

The second prediction of the outcome of the examination, made immediately prior to the actual test, was a more accurate and significant predictor of test outcome, yielding 20 correct predictions out of 22 (91%), $\chi^2=14.7$. However, only 11 of 16 predictions (69%) agreed with the decision made by the panel majority which was not significant ($\chi^2=2.25$).

Effectiveness of physiological measures

In order to determine the relative effectiveness of the respiration, SRR, and cardio measures in detecting deception, the records of 32 suspects who had either made full confessions following the polygraph examination (n=21) or who had later pleaded guilty to the original charge without any plea bargaining (n=11) were selected as representing confirmed deception to the relevant questions.

The polygraph examiner had found 29 of the 32 suspects deceptive, with the remaining three examinations inconclusive. When the charts were blindly evaluated by the independent examiner, 27 charts were found deceptive, 1 was judged truthful, and 4 were inconclusive. To ensure that the scoring was as objective as possible, all analyses in this section are based upon the scores obtained by the independent reviewer. In order to standardize the scores for these analyses, the scores for only the first three charts were used.

The 3-chart independent evaluations showed mean scores of -1.6 for respiration, -7.5 for the SRR, and -1.8 for the cardio. The mean total score was -10.9 at the end of the third chart. An analysis of variance of the scores of the three components yielded a significant difference among the components, $F(2/62)=17.84$. A Newman-Keuls test showed that

the SRR was significantly better than both respiration and the cardio, and there was no difference between respiration and cardio. Analyses of variance found that all three components had scores significantly different from zero at the end of the third chart. Thus, each of the measures was found to be effective in determining deception.

Table 18 shows the effectiveness of each physiological measure in detecting deception by showing the number of correct, incorrect, and inconclusive results, as determined by the sign of the score for each component summed over three charts. For the individual components a

Table 18

Percent Accuracy of Respiration, SRR, and Cardio Measures in Identifying Deception (N=32)

	Respiration	SRR	Cardio	Total (±5)	Total (0)
Correct	69	91	59	78	84
Incorrect	25	9	12	9	16
Inconclusive	7	0	28	12	0

sum of zero was considered inconclusive. The accuracy of the 3-chart total polygraph score was first determined using the ±5 limits of the inconclusive region. The results were then recalculated when the limits of the inconclusive region were narrowed to zero. A chi square test on the number of correct identifications by each physiological measure also indicated that each measure was capable of independently detecting deception.

Biographic variables and autonomic responsivity

Type of crime. The first 77 cases were categorized according to the type of crime. Because of the small number of truthful subjects, the analyses were limited to those subjects called deceptive by the examiner. The scores reported are based upon the original examiner's rescores of the first three charts. The four crime categories were: (1) sex crimes (n=8, mean score = -11.9); (2) drug crimes, including drunk driving (n=7, mean score = -11.3); (3) crimes of violence in which the victim was physically injured or killed (n=15, mean score = -12.2); and (4) crimes of financial gain, including armed robbery in which the victim was not physically hurt (n=23, mean score = -9.2). A one-way analysis of variance was conducted to test for the differences between the means of the groups. There was no significant difference in polygraph scores as a function of the type of crime, $F(3/49) = 0.57$.

Another method of categorizing crimes is to identify the four categories of sex, drug, personal confrontation between the criminal and his victim (including armed robbery), and instances in which the criminal avoided confrontation with the victim (theft, forgery, etc.). Using this classification, the number of subjects and the mean scores for the sex and drug crimes remained the same as before; but the mean score for the 21 subjects in the confrontation group was -11.0, and the 17 subjects in the nonconfrontation group had a mean of -9.2. A one-way analysis of variance on the means also failed to find any significant difference, $F(3/49) = 0.28$. It is apparent that the type of crime that had been committed had no discernible effect upon the responsivity of the suspects classified as deceptive.

Sex. The mean score for the 64 males was -9.5 and the mean for the 13 females was -7.5, which was not significant, $t(75)=0.92$.

Education. Of the first 77 cases the 20 deceptive subjects who had not graduated from high school had a mean polygraph score of -13.1, the 19 deceptive subjects who were high school graduates had a mean polygraph score of -8.6, and the 11 deceptive subjects who had some college education had a mean polygraph score of -10.7. A one-way analysis of variance showed that there was no significant difference in the polygraph scores among the three groups, $F(2/47)=1.73$.

Previous arrests. The first 50 deceptive suspects for whom data were available were classified as having no previous arrests (n=12) or having at least one previous arrest (n=38). The two means of the polygraph scores were -9.5 and -11.3, respectively, which were not significantly different, $t(48)=0.72$.

Religiousness. Ten of the deceptive suspects reported during the pretest interview that they had strong religious convictions. An additional 21 deceptive suspects reported having weak or no religious convictions. The means of the 3-chart rescores by the original examiner were -9.4 and -10.6, respectively, which were not significantly different, $t(24)=0.43$.

Previous polygraph examinations. Fifteen of the first 50 deceptive suspects for whom data were available reported that they had previously been examined on the polygraph by another examiner, concerning either the same incident as the present examination or unrelated crimes. They had a mean score of -11.0 as compared to a mean of -10.7 for the 35 deceptive suspects who denied having ever been previously examined on the polygraph. The difference was not significant, $t(48)=0.13$.

Age. The first 50 deceptive suspects were assigned to one of three categories on the basis of age: the youngest quartile (mean age = 19.0), the middle two quartiles (M=26.8), and the oldest quartile (M=42.6 years). The mean polygraph scores were -8.8, -12.3, and -8.9, respectively, which were not significantly different when tested on a one-way analysis of variance, $F(2/47)=1.38$.

Personality variables

L-scale. The reason for administering the L-scale was to investigate the relationship between the "lie scale" and whether a suspect attempted deception on the polygraph. Because of the difference in the absolute values of the polygraph scores of truthful and deceptive subjects, the analyses were done separately for the truthful and deceptive subjects. The 34 deceptive subjects from whom data were obtained were divided into high and low L-scale groups on the basis of the median L-scale T-score. A t-test on the means of the polygraph scores (-10.2 and -8.6, respectively) found that the difference was not significant, $t(32)=0.60$. A similar analysis of the polygraph scores of the 8 truthful subjects also found that the means of 3.2 for the high-L group and 4.8 for the low-L group were not significantly different, $t(6)=0.62$.

K-scale. Since the K-scale is a more sophisticated fake-good detector than the L-scale, a similar analysis was made using the median K-scale score to separate truthful and deceptive subjects into high-K and low-K groups. Of the total of 48 deceptive subjects, the mean polygraph scores for the high-K and low-K groups were -12.0 and -9.0, respectively. The difference was not significant, $t(46)=1.40$. A

similar analysis for the total of 9 truthful subjects, with means of 3.2 and 4.0 for high-K and low-K, respectively, also showed that the difference was nonsignificant, $t(7)=0.34$.

Pd-scale. Because of the greater theoretical and practical importance of the effect of psychopathy upon detectibility of deception, the analyses conducted with these data are reported in more detail than for the previous variables. There were a total of 36 suspects considered guilty by the panel majority, and to whom the psychopathic deviate scale was administered. In order to sharpen the contrast between the high and low Pd-scorers, the 14 highest and 12 lowest scores were assigned to the psychopath and nonpsychopath groups, respectively. The Pd-scale T-scores of the psychopathic group ranged from 76 to 100, with a mean of 85.4. The percentile of the mean T-score was 99.98. For the nonpsychopath group, the Pd-scale T-scores ranged from 41 to 55, with a mean of 50.3. The percentile of the mean was 51.2.

Table 19 shows the mean scores for the psychopaths and the nonpsychopaths for each physiological measure at the end of the third chart, as scored by the independent examiner. No significant differences were found between the two groups for respiration, SRR, or all measures combined. However, the cardio scores for psychopaths were significantly smaller than those of the nonpsychopaths, $F(1/24)=5.08$. All means in Table 19 were significantly different from zero except for the mean cardio score for psychopaths and the respiratory scores for each group and for both groups combined.

Table 19

Mean Scores of Psychopaths and Nonpsychopaths For Each Physiological Measure at the End of the Third Chart

	Respiration	SRR	Cardio	Total
Psychopaths	-0.9	-6.6	-0.4	-7.8
Nonpsychopaths	+0.5	-5.1	-2.5	-7.1
Total	-0.2	-5.8	-1.4	-7.4

Hs -scale. The hypothesis that hypochondriasis may increase the possibility of false positive errors could not be tested because of an insufficient number of confirmed false positive errors in this study. It had also been hypothesized that deceptive hypochondriacs might have larger autonomic responses than deceptive nonhypochondriacs. To test this hypothesis, all 34 deceptive subjects to whom the scale had been administered were assigned to a high Hs or a low Hs group on the basis of the median K-corrected T-score. A t-test showed that the means of -7.8 and -10.9, respectively, were not significantly different, $t(32)=1.202$. Similarly, there was not a significant difference for the mean polygraph scores of the 8 truthful subjects, with means of 5.0 and 3.0 for the high Hs and low Hs subjects, respectively, $t(6)=0.85$. The correlation between the Hs-scale T-scores and the polygraph scores of the 34 deceptive subjects was not significant, $r(32)=-.17$.

D-scale. The hypothesis that depressed persons might show smaller autonomic responses in a detection of deception situation was tested by separating the 52 subjects for whom D-scale scores were available

into high D and low D groups on the basis of the median D-scale T-score. The mean absolute values of the polygraph scores of the two groups were 7.9 and 8.9, respectively, and were not significantly different, $t(50) = -0.53$.

Discussion

The significant agreement between the polygraph examiner's decisions and the decisions of the panel when only scores of zero were considered inconclusive is remarkable given the paucity of information available to the panel in many instances. As the score required for a panel decision was increased, the agreement between the polygraph and the panel was generally 85% or higher. The same level of agreement was also found using the criterion of the panel majority and also with the judicial outcome criterion.

Generally speaking, with the bulk of the criteria there was no false negative polygraph outcome. The only disagreements were with deceptive polygraph results when the panel judged a suspect to be innocent or when the judiciary acquitted the defendant. Of the disagreements, one has been confirmed as a false positive polygraph error. In the remaining disagreements, there is not sufficient evidence to determine whether the polygraph technique or the criteria were in error. The fact that there was only 89% agreement between the panel decisions and the judicial outcomes points out that the criteria were less than ideal and raises the question of whether the polygraph decision or the criterion for guilt or innocence was wrong when they disagreed.

The fact that all of the disagreements occurred on persons judged to be innocent was not surprising when the judicial criterion was

employed, since a suspect must be acquitted unless the evidence against him shows guilt beyond a reasonable doubt. There are several possibilities as to why the panel disagreed with the polygraph outcome. First, the fact that all of the panelists were legally-trained professionals may have prompted them to give the benefit of the doubt to some of the suspects. The available evidence in many of the cases submitted to the panel fell far short of that required for conviction in courts. The defense attorneys commented that although they attempted to arrive at a decision without reference to legal implications and restrictions, they felt that in a number of cases they were biased towards innocence because they were well aware of the legal penalties called for in each case by a decision of guilty.

A second hypothesis is that the dossiers presented to the panel contained biased information, since the suspect's account was always available; whereas the information concerning the police investigation of the suspect's version was often inadequate. The project investigators responsible for assembling the case facts reported that although the attorneys were interested and cooperative, they were so busy that interviews with them were much shorter than desirable. The defense attorneys' files generally lacked documentary investigative material. The caseloads of the prosecuting attorneys were so heavy that the prosecutors often could not remember critical details of the investigations or the names of the investigators who would be able to furnish detailed information. The police reports were often limited to the initial on-the-scene investigation, and reports of subsequent investigations by detectives or the county attorney's office usually could not be located. The

difficulties involved in trying to trace down the appropriate documentation from various sources were compounded by the inexperience of the project's investigators, who were university students. The case information presented to the panel was often limited to unsubstantiated testimony or hearsay and often consisted of the initial police report describing the complaint, the initial link between the incident and the suspect (usually an accusation by the complainant or photographic identification by the victim), and a paraphrase of the suspect's version. Unless the suspect confessed or later pleaded guilty, the panel could rarely be provided with additional information.

A third hypothesis is that it may have been more difficult for the polygraph examiner to detect that a person was truthful than to detect his deception. There are several independent lines of evidence bearing on that hypothesis. Using the panel majority criterion for guilt and innocence and the polygraph scores of the independent examiner, the mean score for the guilty subjects was more extreme than the mean score for the innocent subjects. However, a large proportion of the subjects found truthful by the polygraph were judged inconclusive by the panel.

Those data are consistent with an earlier study (Barland & Raskin, 1975) conducted with college students in a laboratory situation in which the polygraph scores for innocent subjects were less extreme than those for guilty subjects. A later experiment (Raskin, 1975) used a control-question technique with convicted criminals and psychopaths in a mock-crime situation. Although 95.5% accuracy of decisions was obtained, the two errors were false positives. However, there was no difference in the extremeness of scores for guilty and innocent subjects. A recent

study in our laboratory using a sample of subjects from the general population in a mock-crime situation found a higher rate of false negative as compared to false positive results and other indications of better results in identifying truthful subjects.

In the present study when decisions were made, they tended to be similar in accuracy for deceptive and truthful polygraph decisions. That is to say that if one assesses the confidence which can be placed in a decision, then relatively high confidence can be placed on both truthful and deceptive polygraph outcomes. That result is consistent with the findings by Bersh (1969) which indicated that the agreement between a panel majority and polygraph decisions was 88.1% for subjects found deceptive and 86.9% for those found truthful by the polygraph examiner.

Although the 9 errors by the original examiner were false positives when using the panel majority criterion, they comprised only 9.8% of the total number of cases. When the charts were independently evaluated, 6 of the 7 errors were false positives, and they comprised only 6.5% of the cases. Since 73.4% of the panel majority decisions indicated that the subject was guilty, it is surprising that the large majority of errors were false positives. There were considerably fewer opportunities for the polygraph to yield false positive errors, given the small proportion innocent determinations by the panel. However, the rates of polygraph errors were approximately equal for deceptive and truthful polygraph decisions.

Although there seems to be some suggestion that control-question techniques may produce a relatively greater proportion of false positives than false negatives, those findings should be viewed within the framework

of their practical application. When the results are considered in terms of data which form part of the basis for the determination of guilt or innocence, then a high degree of confidence may be placed on truthful as well as deceptive polygraph outcomes. However, such data should not be used as the sole basis for arriving at a conclusion of guilty. In accord with our usual judicial procedures, determinations of guilty should be arrived at using considerable evidence in addition to a deceptive outcome on a polygraph test. Furthermore, it might be argued that a truthful polygraph result is adequate to cast a reasonable doubt and form the basis for dismissal of charges or acquittal. However, such decisions are best left to the courts.

Reliability of chart interpretation

When the original examiner rescored his charts after an interval of about six months, the score-rescore correlations ranged from .80 for respiration to .92 for total polygraph scores. These correlations are generally consistent with previous research in which numerically-scored polygraph charts were rescored by five examiners other than the original examiner (Barland & Raskin, 1975).

The fact that the original polygraph scores were more extreme than the re-evaluations six months later could have been caused by the influence of extraneous information such as behavior symptoms on the original evaluation of the charts, or it could have been caused by the pressure to make a decision at the time of the original examination. However, the facts that there were no decision reversals, merely an increase in the number of inconclusives and that 84.4% of the trichotomous outcomes remained the same indicate that the re-evaluations were

highly reliable in terms of actual outcome.

The independent evaluations of polygraph charts by another examiner also yielded a high degree of reliability. The correlation of .91 between the two examiners is a demonstration of great consistency in the assignment of scores and indicates that the outcomes may be determined accurately by means of the polygraph charts alone without any contact with the subject or access to case information. Furthermore, the decisions made by the independent examiner agreed with those of the original examiner in 98.4% of the instances and agreed 100% with decisions based on original examiners' scores. Obviously, application of the numerical scoring technique by examiners experienced with that technique yields highly consistent results.

Most leading polygraph examiners have usually stated that while the SRR is an accurate discriminator between truth and deception in the laboratory or with a peak of tension test, it is the least effective physiological measure with criminal suspects when a control question test is utilized (e.g., Arther, 1971; Lee, 1953, p. 21; Marston, 1938, p. 59). Reid and Inbau suggested that the differential effectiveness of the SRR is produced by different sources of cognitive awareness in laboratory settings as opposed to emotional responses in real-life situations (Reid & Inbau, 1966, p. 237). Some examiners have claimed that the SRR's potential effectiveness in field situations is negated by the extreme lability of the SRR measure, which makes it difficult to score, and its susceptibility to extraneous cognitive activity which makes it ambiguous to interpret. In the present study, the opposite was found to be true, i.e., the SRR was by far the easiest to score, and with the

32 confirmed deceptive subjects it was by far superior to the two other measures.

Various hypotheses have been discussed for the lack of favor of the SRR with field examiners using control-question tests (Barland & Raskin, 1973). Recent reports by other examiners based upon careful evaluation of all physiological measures are beginning to show that the SRR is not the least effective (Slowik & Buckley, 1975).

These findings suggest that the traditional distrust of the SRR may have been based in part upon examiner bias or other factors such as poor examination techniques, inadequate instrumentation, and lack of proper training in chart interpretation.

The intercorrelations of the polygraph measures were relatively low. Respiration responses were not correlated with SRR and cardio responses; however, all three physiological measures were able to identify deception in the 32 confirmed deceptive subjects. These facts suggest that the examiner obtains useful, independent information from each physiological measure and gains in ease of decision-making by recording multiple physiological indices.

Behavior symptoms

Behavior symptoms turned out to be surprisingly accurate in predicting the polygraph outcome, though not guilt or innocence as determined by the panel. The initial prediction was not significantly accurate when all predictions were considered. However, 91% of the later predictions that a subject would be deceptive were confirmed by the exmaination. Only 50% of the predictions of nondeception were confirmed by the examination. Not unexpectedly, the second prediction of polygraph outcome, based upon observation of the suspect for about 60 to 90 minutes, was more accurate.

It is also apparent that the systematic observation of behavior did not increase the number of errors or inconclusives, since there was no difference in the scores of those whose behavior was systematically observed as compared to those subjects in whom the behavioral observations were unsystematic. Of course, it is possible that even where there was no systematic observation of behavior, the examiner's clinical impressions obtained during the pretest interview may have subtlely influenced the subjects' autonomic responsivity.

Personality and biographical variables

With one exception, none of the various biographical and personality variables examined in this study were related to the autonomic responsivity of subjects judged deceptive. Thus, it did not matter whether the subject scored high or low on most of the MMPI scales which were administered. It also made no difference whether the person was a first-time offender or a long-time recidivist, undergoing his first polygraph examination or not, a grade-school dropout or a college graduate, male or female, young or old, deeply religious or not. It made no difference what type of crime he had committed.

The only statistically significant finding concerning personality variables was that the cardiovascular scores of psychopaths were smaller than those of the nonpsychopaths. This finding is contrary to previous research with psychopaths, both in standard laboratory situations (Hare, 1970) and in detection of deception situations (Raskin, 1975). That research has found either no difference in cardiovascular responsivity

between psychopaths and nonpsychopaths or that psychopaths tend to be stronger cardiovascular responders than nonpsychopaths. The most parsimonious explanation of the cardiovascular analysis of this experiment is that the difference was probably a random difference due to sampling error. Certainly, when looking at all physiological measures combined, the deception of the psychopaths was as readily detected as that of the nonpsychopaths. That is consistent with the only extensive laboratory investigation which found psychopaths at least as readily detectable as nonpsychopaths (Raskin, 1975).

This is not to say that personality and such variables have no effect. Possibly the MMPI diagnostic procedure was too crude to permit optimum categorization of the subjects. Possibly the polygraphs were too insensitive or the numerical scoring system too crude to pick up subtle differences in patterns of physiological responsivity. However, this study does clearly show that when they are assessed against external criteria of guilt or innocence, the personality and background variables examined here did not have any discernible relationship with the outcome of the examination.

Three factors which limit the generalizability of the results of this study include the use of only one polygraph examiner, the inadequacy of the criteria against which the examiner's decisions were compared, and the small number of suspects diagnosed as truthful by the polygraph and innocent by the external criteria.

Any criterion against which the results of polygraph examinations can be compared has some problem. The most obvious problem is that ground truth can never be completely determined outside of a laboratory

situation, especially on persons cleared of suspicion by the polygraph examination. Other problems include the possible lack of independence between the polygraph examination and preceding and subsequent events. Another problem involves the variability of the subject-examiner interaction in terms of the content of the pretest interview, the type of examination conducted, the wording of the test questions, and the number of charts obtained. However, given those limitations, the overall results are consistent with the proposition that carefully administered control question polygraph examinations are highly accurate in assessing the credibility of criminal suspects.

References

Abrams, S. Polygraph validity and reliability: A review. Journal of Forensic Sciences, 1973, 18, 313-326.

Arther, R. O. The GSR unit. Journal of Polygraph Studies, 1971, 5(6), 1-4.

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

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

Bersh, P. J. A validation study of polygraph examiner judgments. Journal of Applied Psychology, 1969, 53(5), 399-403.

Hare, R. D. Psychopathy: Theory and Research. New York: Wiley & Sons, 1970.

Lee, C. D. The instrumental detection of deception. Springfield, Ill.: Charles C. Thomas, 1953.

Marston, W. M. The lie detector test. New York: Richard K. Smith, 1938.

Orne, M. T., Thackray, R. I., & Paskewitz, D. A. On the detection of deception: A model for the study of the physiological effects of psychological stimuli. In N. Greenfield & R. Sternbach (Eds.), Handbook of psychophysiology. New York: Holt, Rinehart & Winston, 1972. Pp. 743-785.

Raskin, D. C. Psychopathy and detection of deception in a prison population. Report No. 75-1, Contract 75 NI 99-0001, Department of Psychology, University of Utah, June 1975.

END