

77

92

106

M.T. Bradley and Michel Pierre Janisse

Julian J. Szucko and Benjamin Kleinmuntz

Statistical versus Clinical Lie Detection

Lillian Lim Quon

The Polygraph, The Courts and Law Enforcement

### PUBLISHED QUARTERLY

©AMERICAN POLYGRAPH ASSOCIATION, 1981 0(2) P.O. Box 74, Linthicum Heights, Maryland 21090

Polygraph 1981, 10(2)

Measures

## INSPECTOR WILLIAM Y. DORAN ADDRESSES FEDERAL EXAMINERS\*

I want to welcome you to the FBI Academy on behalf of Judge Webster who had hoped to be with you today but his schedule would not permit. By way of background, I am not the polygraph examiner nor a scientist but as an administrator in the FBI laboratory I have had the opportunity to become involved with our polygraph program. I want to share with you today some of my observations and concerns in this regard.

Recently there has been a significant increase in the use of the polygraph not only as an investigative tool but also in business and government as a means of screening applicants or employees to improve personnel security.

At the same time, the courts have been looking more favorably on polygraph as admissible evidence. Although stipulation remains the most common circumstance for admissibility, both state and Federal courts have ruled admissibility over the object of the prosecution.

Interesting note: Recent 7th Circuit case <u>McMorris v</u>. <u>Israel</u> - Appellate court ruled that prosecution must stipulate unless justifiable and articulate reasons exist not to.

I know this is all encouraging to the members of the polygraph profession who are devoted to the purpose of gaining scientific and judicial acceptance. But as acceptance increases so also are the examiners' responsibilities and the number of pitfalls to be avoided.

Examiners by their very nature must be strong-willed, determined individuals. (I've even known one or two to be hard-headed.) If they are not, they seldom survive in this walk of life. Unlike other investigators who gather facts and evidence for a prosecutor, the polygraph examiner must conduct the examination, and then, based upon the complete process commit himself/herself to written conclusions regarding the veracity of the examinee--a conclusion or opinion which normally goes to the heart of the matter. When an investigator errs his/her reputation and prestige are not usually damaged. But if a polygraph examiner makes an error he/she Frequently, his/her personal reputation and that of the stands alone. profession are severely damaged. This is why it takes a strong person to make these determinations day after day and not collapse under the pres-This same strength and determination, so necessary, however may sure. work against an examiner - the hard head who for pride or some other reason is unwilling to occasionally accept inconclusive findings. It must be remembered that the inconclusive range serves a purpose - it is the safety No examiners zone and should be protected to avoid unnecessary errors. should render a judgment if he/she is not completely comfortable with his/ His/her charts must clearly support his/her conclusions her findings. based upon the probability of accuracy established by research.

<sup>\*</sup>William Y. Doran, Inspector, Deputy Assistant Director, Laboratory Division, Federal Bureau of Investigation, addressed the opening session of the annual seminar of the Federal Interagency Polygraph Committee held May 18-22, at the FBI Academy in Quantico, Virginia. The seminar was attended by 85 federal examiners, representing 14 different agencies.

## INSPECTOR WILLIAM Y. DORAN ADDRESSES FEDERAL EXAMINERS

Another area which must be avoided for the good of the profession is the spectre of constant opposing testimony of polygraph examiners in court. All examiner should be seekers of truth and not be swayed to the point of becoming "advocate experts." Unfortunately, we in government have little control over examiners in the provate sector. This is a problem that should be addressed by the professional group as a whole.

One potential pitfall of current concern to the FBI in the criminal arena is the use of polygraph in post-trial and pre-sentence situations. On occasion, we have honored the requests of judges who have asked us to conduct polygraph examinations of defendants following their conviction. The purpose was to provide investigative information to the courts which was previously unavailable. These incidents have involved confirmation of testimony, verification of mitigating circumstances, and verification of new testimony, events, or circumstances not previously known.

Use of polygraph by the Bureau in these situations in the past has presented no problems. Questions have arisen in the Bureau, however, regarding this practice. The concern is over the propriety of the practice since this utilization could be construed as challenging the validity of the jury's conclusions. The net result as it relates to the polygraph profession is that it could have a chilling effect on the criminal justice system. We in the Bureau view polygraph as an excellent investigative tool but don't desire to see it viewed as seeking to usurp the rightful responsibility and authority of the judicial system.

The matter of post conviction use of polygraph is being reviewed by the Bureau at this time for the purpose of establishing a definitive policy. Your agencies may wish to give this matter some thought also.

Again, we in the FBI are very glad to have you here with us and wish you a very pleasant, productive and informative stay. If there is anything we can do to improve your visit, please let us know. Thank you for your kind attention and I look forward to seeing you again in future years.

\* \* \* \* \* \*

Search for the truth is the noblest occupation of man; its publication is a duty. -Mad. de Stael.

\* \* \* \* \* \*

## THE COMPARISON OF TWO STIMULATION TESTS AND THEIR EFFECT ON THE POLYGRAPH TECHNIQUE

By

#### Steven L. Kirby\*

#### Abstract

This study was designed to compare two types of stimulation tests in the polygraph technique, identified as the standard card test (SCT) procedure (in which the subject is unaware of the fact that the examiner knows the identity of the chosen card prior to the administration of the test) and the known card test (KCT) procedure (in which the subject reveals the chosen card to the examiner prior to the administration of the test.)

The study assessed the responsiveness of each subject to their chosen card under both conditions. The study also evaluated whether or not subjects attempted to distort their responses to the chosen card under one test more than the other. Finally, the study concerned itself with whether or not examiner accuracy, in evaluating truth, deception or indefinite decisions on the issue questions, was affected by the stimulation test.

The results indicated no significant difference in the responsiveness of subjects to their chosen card between the two techniques. The study indicated that deceptive subjects attempted to distort the known card test to a more significant degree than deceptive subjects who were administered a standard card test. There was no significant difference between the two test conditions for truthful subjects distorting their responses to their chosen card. Finally, while overall examiner accuracy decreased in evaluating issue tests given after both stimulation tests, there was no significant difference between the two techniques in either the accuracy of the decisions or the rate of inconclusive decisions.

#### DESCRIPTION OF TESTS

Inasmuch as all polygraph records used in this study were conducted under the Reid Control Question Technique, some discussion is required to explain the technique in brief.

After accumulating sufficient factual data to conduct the examination, a series of relevant test questions (three or four) are constructed to which the subject will only answer yes or no. After a pre-examination interview with the subject, a series of tests are conducted.

<sup>\*</sup>Steven L. Kirby, an APA Member, is the President of Edward R. Kirby & Associates, Inc., 6525 West North Avenue, Oak Park, Illinois 60302. Reprints of this article are available from Mr. Kirby. Polygraph 1981, 10(2)

The tests consist of the aforementioned relevant questions, four irrelevant questions, and two control questions. An irrelevant question is a non-stressful question to which the subject's answer of "yes" will be a known truth. A typical irrelevant question would be "Do they call you John?" A control question is one which is unrelated to the matter under investigation, but is of a similar, though less serious nature, and one to which the subject will in all probability lie or at least his answer will give him some concern with respect to either its truthfulness or its accuracy(Horvath & Reid, 1971). A typical control question for a theft investigation would be "Did you ever steal anything in your life?" or "Besides what you told me about, did you ever steal anything in your life?" A control question is answered "no" by the subject. In the test, the irrelevant questions are asked in the first, second, fourth and seventh position; the controls are in the sixth and final question position, while the relevant questions are asked in the third, fifth, eighth and ninth position(Inbau & Reid, 1964).

In the Reid Control Question Technique, the previously described test is conducted, followed by a standard card test (described in detail later in this paper), followed by a third test identical to the first test in the series. Additional tests may be run after the third test and usually are conducted. However, the tests conducted after the third test had no bearing on this study.

After administering the full series of tests, the examiner then evaluates the test records. At the risk of over simplification, if a subject's physiological reactions are greater to the relevant questions than to the control questions, he is reported deceptive, and if his physiological reactions are greater to the control questions than to the relevant questions, he is reported truthful(Wicklander & Hunter, 1975).

In the administration of the standard card test, the examiner presents to the subject seven numbered cards face down in a pre-arranged order and instructs the subject to select one of the cards, study it, remember the number, and, not revealing which card was selected to return it to the deck. After making sure that the subject has indeed looked at the card in question (some subjects have been known not to look at the card in order to avoid any reaction), the examiner shuffles the cards, telling the subject that on the next test he will ask him whether or not he chose card number 3 or the other various cards in the deck. The subject is instructed to answer "no" to all of the cards, even the one he took, no matter how many times the subject is asked about that particular card number. The examiner then explains to the subject that by following this procedure he will be responding truthfully to all of the cards except the one he actually chose. The examiner continues by telling the subject that if the instrument is properly adjusted, it will reflect which card number the subject is not telling the truth to. At this time the examiner knows which card the subject selected due to the pre-arranging of the cards. The subject is not aware of this fact.

The examiner then proceeds with the test asking the basic question "Did you pick card number ?" The subject is asked the various cards in the deck with about a 15 second interval between questions. At the third question, the examiner asks the chosen card. This is followed by asking a different number, and then the chosen card question is repeated. If there is evidence of purposeful distortion, the chosen card is asked a third time. The chosen card is usually asked in the third, fifth, (and if distortion to the chosen card is shown), seventh position in the question sequence.

After the card test, the examiner turns off the instrument and informs the subject which card the examiner believes was chosen. When the subject confirms that this was in fact his selected card, the examiner then will generally proceed to verbally attempt to stimulate the subject by stressing that the instrument has determined which card the subject lied to, stressing the instruments infallibility. In the Reid Control Question Technique, the examiner then leaves the room for 3 to 5 minutes, allowing the subject to think over the specific issue test questions.

Upon re-entering the room, the examiner asks the subject if there are any questions to which the subject did not answer truthfully. If the subject makes no admissions, the examiner then proceeds to administer the remaining battery of tests. If the subject admits to a particular question (usually a control question), the examiner incorporates the admission or clarifying statement by rephrasing the question. Additional details of the standard card test stimulation can be found in Reid and Inbau(1977).

The administration of the Known Card Test differs from the Standard Card Test in the following manner: After the subject selects the card from the deck, he is instructed by the examiner to verbally inform the examiner which card he has chosen. The examiner then looks at the card to ensure that the subject has stated the correct number. At this point the card becomes "known." The examiner then instructs the subject to answer "no" to all of the cards, including the one that was selected, thereby answering truthfully to all of the cards except for the one he chose.

It is also explained to the subject that there should be a difference on the test between the known card and the other cards. The administration of the test is identical to the standard card test. Upon completion of the test, the examiner states to the subject that the test clearly indicates that there is a difference between reactions to the known card and those not chosen. If there is a discernable difference in one or more of the tracings, the examiner should show this to the subject, pointing out the differences on the actual chart tracings.

If there is evidence that the subject purposefully distorted the known card test, it is advisable to state only that this test served its purpose, rather than confirming in the subject's mind that his purposeful distortions caused the difference in reaction and therefore he "beat the test."

The examiner proceeds to verbaly stimulate the subject as previously described, leaving the laboratory for 3 to 5 minutes, and upon returning changing or re-phrasing any question to which the subject has made admissions.

The remainder of the examination should be conducted as previously described.

#### PURPOSE

It is the purpose of this paper to compare and discuss the two techniques and to determine if, in fact, the known card test has the stimulation value of the standard card test.

While it has been shown that examiner accuracy increases with the stimulation effect of a standard card test(Senese, 1976), there has been criticism that this test is a form of trickery and deceit(Star, 1977). In addition, if the examiner is not careful in the administration of the standard card test, there may be a discrepancy in the card the examiner feels the subject has selected and the card the subject actually has chosen. If in fact the examiner makes an error in determining which card the subject has chosen, he has probably destimulated the subject in that, he has proven to the subject that the polygraph test has not worked.

The known card test eliminates the possibility of either examiner or subject error in the selection of the card because both the examiner and subject know the card before the test is administered. In addition, since both parties to the examination known the card the subject chose, there can be no claims of trickery or deceit.

The areas to be compared are: (1) The degree of actual physiological response to the chosen card, (2) The percent cases wherein the subject attempted to distort the test and (3) The examiner accuracy in determining with and without the benefit of either the standard card test or known card test.

The three previously mentioned areas of comparison were computed for all subjects.

#### METHOD OF EXPERIMENT

The data used in this study consisted of forty sets of verified test results. Of these forty sets of records, all have been verified as to the subject's truthfulness or deception by either fully corroborated confessions or in one case, by proof that the offense under investigation never took place. All test records used were specific one issue tests.

Twenty of the subject's records in this study were verified deceptive, ten having been given a standard card test and ten a known card test. The other twenty subjects were verified truthful, ten given a standard card test and ten given a known card test.

All tests were conducted with the Reid Control Question Technique, wherein the card test was the second in a series of charts. Copies of the first chart were given to 10 examiners with experience, ranging from one to eight years, with a mean average experience of four and one-half years. each examiner made an independent analysis of the first control question chart and was asked to make an overall evaluation of the subject's truthfulness or deception. After a period of one month, the examiners were given the third charts and the examiners were asked to make similar diagnosis.

The same ten examiners, along with two additional examiners (both with approximately nine years of experience), were also shown the forty Polygraph 1981, 10(2)

# Steven L. Kirby

card charts (not knowing which were the standard or known card tests). The twelve examiners were asked to give a subjective opinion as to whether they observed a significant to moderate response, an erratic to minimal response, or an intentional distortion of the identity of the chosen card.

### RESULTS

Examiners judged 56% of the standard card test responses to the chosen card to be significant to moderate, 40% to be either erratic or minimal, and 4% to be purposely distorted. In comparison, the examiners judged 47% of the responses to the known card test significant to moderate, 41% to be either erratic or minimal and 12% of the known card test responses to be distorted.

## TABLE 1

Examiners evaluation of subjects response to Chosen Card -Known Card Test versus Standard Card Test

	Significant/Moderate	Erratic/Minimal	Distorted
KCT	47%	41%	12%
SCT	56%	40%	4%

In the analysis of verified truthful subjects given a standard card test, 56% of the responses to the chosen card were judged to be either significant or moderate and 44% to be erratic or minimal. None of those verified truthful subjects given a standard card test were suspected of intentionally distorting reactions to the chosen card.

Of those verified truthful subjects who were given the known card test, 55% of the responses were judged to have been either significant or moderate. Forty-three percent of the responses for this group were judged to be erratic or to have no emotional response. Two percent of the subjects in this group were thought to have distorted the reaction to the chosen card.

### TABLE 1A

Examiners evaluation of subject response to Chosen Card verified truthful subjects Known Card Test versus Standard Card Test

	Significant/Moderate	Erratic/Minimal	Distorted
KCT	55%	43%	2%
SCT	56%	44%	0%

Of subjects verified deceptive who were given a standard card test, 56% of their responses were rated as either moderate or significant, 36% of the responses were rated either erratic or were judged to be minimal reaction, 8% of the subjects were judged to have distorted the response to the chosen card.

In comparison, those verified deceptive subjects given a known card test were judged as follows: 39% significant or moderate reaction, 39% POlygraph 1981, 10(2) 67 erratic or minimal response, and 22% intentional distortion.

TABLE	1B
-------	----

Examiners evaluation of subjects responses to Chosen Card verified deceptive subjects Known Card Test versus Standard Card Test

	Significant/Moderate	Erratic/Minimal	Distorted
KCT	39%	39%	22%
SCT	56%	36%	8%

To test for a significant difference between the level of responsiveness to the chosen card, a statistical T-Test was performed. There was no significant difference at the 99% confidence level between the level of responsiveness of subjects given a standard card test and those given a known card test.

	Qura 1	COL		
KC	CT%	SC		
Significant/	Erratic/	Significant/	Erratic/	
Moderate	Minimal	Moderate	Minimal	
66.7	43.4	80.0	20.0	
63.2	36.8	65.0	35.0	
57.9	42.1	70.0	30.0	
41.2	58.8	52.6	47.4	
42.1	57.9	36.8	63.2	
72.2	27.8	42.1	57.9	
42.1	57.9	60.0	40.0	
64.7	35.3	52.6	47.4	
56.2	44.8	52.6	47.4	
31.6	68.4	44.4	65,6	
52.9	47.1	63.2	36.8	
57.1	42.9	77.8	32.2	
	KC Significant/ <u>Moderate</u> 66.7 63.2 57.9 41.2 42.1 72.2 42.1 64.7 56.2 31.6 52.9 57.1	KCT%   Significant/ Erratic/   Moderate Minimal   66.7 43.4   63.2 36.8   57.9 42.1   41.2 58.8   42.1 57.9   72.2 27.8   42.1 57.9   64.7 35.3   56.2 44.8   31.6 68.4   52.9 47.1   57.1 42.9	KCT%SCSignificant/Erratic/Significant/ModerateMinimalModerate66.743.480.063.236.865.057.942.170.041.258.852.642.157.936.872.227.842.142.157.960.064.735.352.656.244.852.631.668.444.452.947.163.257.142.977.8	KCT%SCT%Significant/Erratic/Significant/Erratic/ModerateMinimalModerateMinimal66.743.480.020.063.236.865.035.057.942.170.030.041.258.852.647.442.157.936.863.272.227.842.157.942.157.960.040.064.735.352.647.456.244.852.647.431.668.444.465.652.947.163.236.857.142.977.832.2

TABLE 2

Test of Significant difference relative to level of the subject's responsiveness to the Chosen Card - Known Card Test versus Standard Card Test

t = -0.97df = 11

A T-Test was also tabulated to determine significant difference between the number of times a subject intentionally distorted his chosen card under the different test conditions. There was a significant difference at the 99% confidence level that showed subjects were more likely to distort their chosen card on the known card test.

## Steven L. Kirby

					Т	ABLE	3					
Test o	f si;	gnifi	lcan	t di	ffer	ence	of	subj	ects	judge	ed to	have
distor	ted	their	re re	spon	ses	to tl	ne C	hose	n Cai	rd - 1	Known	Card
		1	ſest	ver	sus	Stand	dard	Car	d Tea	Bt		
Examiner	1	2	3	4	5	6	7	8	9	10	11	12
%КСТ	10	5	5	15	5	10	5	15	20	5	15	30
%SCT	0	0	0	5	5	5	0	5	5	10	5	10
t = 3.96 df = 11									·			

To ascertain whether a significant number of verified truthful subjects given the KCT were distorting their chosen card (a tactic associated with deceptive subjects), additional T-Tests were conducted comparing verified truthful subjects with each other and verified deceptive subjects with each other. There was no significant difference at the 99% confidence level between the 2% of the truthful subjects given a known card test who were judged to have distorted their tests versus the 0% distortion by truthful subjects given the standard card test. There was a significant difference at the 99% confidence level in comparing deceptive subjects who distorted the test when given a known card test is significantly greater than those deceptive subjects who distorted their response on the standard card test.

# TABLE 3A Test of significant difference of subjects verified truthful to have distorted their responses to the Chosen Card - Known Card Test versus Standard Card Test

Examiner	1	2	3	4	5	6	7	8	9	10	11	12
%KCT	0	0	0	5	0	0	0	5	0	0	0	0
%SCT	0	0	0	0	0	0	0	0	0	0	0	0
t = 1.49 df = 11												

#### TABLE 3B

Test of significant difference of subjects verified deceptive judged to have distorted their responses to the Chosen Card - Known Card Test versus Standard Card Test

Examiner	1	2	3	4	5	6	7	8	9	10	11	12
%SCT	0	0	0	5	5	5	0	5	5	10	5	10
%KCT t = 3.37	5	0	5	15	5	10	5	15	20	5	15	30
 df = 11						60						

In analyzing examiners accuracy on the first and third tests, using the two stimulation procedures, the following results were obtained:

The overall examiner accuracy in determining truth or deception on the first charts of those subjects subsequently given a standard card test was 58.5% correct, 21.5% incorrect, 20.0% inconclusive. The results of examiner analysis of the first chart of those subjects subsequently given a known card test were 66% correct; 17% error and 17% inconclusive.

Examiner accuracy on the chart following the known card test was 59.5% correct; 25% incorrect and 15.5% inconclusive.

## TABLE 4

## •Comparison of Examiner accuracy on evaluation of subject's tests before and after a Known Card Test and a Standard Card Test

First Test (Before Stimulation Test)

	KCT	SCT
Correct	66%	58.5%
Incorrect	17%	21.5%
Inconclusive	17%	20.0%
Third Test (After Stimulation)		
Correct	59.9%	59%
Incorrect	25%	23%
Inconclusive	15.5%	18%

Table 4A reflects the breakdown of examiner accuracy for those subjects verified truthful using both stimulation tests. Table 4B is a similar breakdown for those subjects verified deceptive.

In order to compare whether or not there was a significant difference in examiner accuracy in evaluating tests administered after either the known card test or the standard card test, a T-Test was conducted to determine if one group of records was more inherently difficult to interpret than the other. Excluding indefinites, examiners had an overall accuracy of 79.5% in evaluating the first test of those subjects subsequently given a known card test as opposed to a 72.9% overall accuracy in evaluating those subjects subsequent given a standard card test. The T-Test indicated no significant difference at the 95% confidence level between the two groups of charts.

## TABLE 4A

Comparison of Examiner Accuracy on evaluation of subject's verified truthful before and after a Known Card Test and Standard Card Test

First Test (Before Stimulation Test)

\_

.

	KCT	SCT
Correct	52%	45%
Incorrect	26%	37%
Inconclusive	22%	18%
Third Test (After Stimula	tion Test)	
Correct	53%	59%
Incorrect	30%	18%
Inconclusive	17%	23%

### TABLE 4B

Comparison of Examiner Accuracy on evaluation of subjects verified deceptive before and after a Known Card Test and Standard Card Test

First Test (Before Stimulation Test)

	KCT	SCT
Correct	80%	72%
Incorrect	8%	6%
Inconclusive	12%	22%
Third Test (After Stimulation	Test)	
Correct	68%	59%
Incorrect	20%	28%
Inconclusive	12%	13%

,

	- Standard Card lest versus	KIIOWII CALU IESL
Examiner	KCT	SCT
1	72.2	77.8
2	86.7	76.9
3	77.8	73.3
4	70.0	68.4
5	72.2	72.2
6	82.4	64.7
7	81.7	62.8
8	100.0	88.9
9	80.0	66.7
10	66.7	71.4
Total Average	79.5	72.9
t = 2.394		
df = 9		

TABLE 5Test of significant difference of examiner accuracy in evaluatingthe first test - Standard Card Test versus Known Card Test

That having been established, the examiners overall accuracy in evaluating the third chart (the test given after either the standard card test or the known card test) was examined. Contrary to a previous study (Senese, 1976), there was a decrease of 6.8% examiner accuracy in evaluating charts given after the standard card test. There was a 10.9% decrease in accuracy in evaluating charts given after the known card test as opposed to examiner accuracy in evaluating the first chart. A T-Test indicated that there is no significant difference at the 99% level in the decreases between the two stimulation tests.

TABLE	6
-------	---

Test of significance for the difference in the percent of increase or decrease of examiner accuracy before and after the Stimulation Test. Known Card Test versus Standard Card Test.

Examiner	KCT	SCT
1	+ 5.6	- 12.8
2	- 9.8	+ 10.6
3	- 25.2	- 2.7
4	- 1.6	- 24.0
5	- 9.0	+ 1.1
6	- 11.0	+ 2.0
7	- 17.9	- 6.3
8	- 5.6	- 3.9
9	- 9.4	+ 19.0
10	0.0	- 13.2
Total Average	- 10.9	- 3,8
t = 1.52 .		

df = 9

In comparing examiner accuracy before and after the stimulation test for truthful subjects, there was no significant difference at the 99% confidence level between the known card test or standard card test subjects.

Likewise, there was no significant difference at the 99% level of confidence in examiner accuracy before and after either stimulation test for those subjects verified deceptive.

## TABLE 6A

Test of significance for the difference in the percent of increase or decrease of examiner accuracy before and after a Stimulation Test verified truthful subjects - Known Card Test versus Standard Card Test

Examiner	KCT	SCT
1	0	+ 14.5
2	+ 38.6	+ 26.4
3	- 7.0	- 20.8
4	+ 5.5	- 10.0
5	- 15.5	+ 15.6
6	- 4.1	+ 22.3
7	- 32.2	+ 16.7
8	- 12.5	+ 12.3
9	- 16.6	+ 25.0
10	+ 40.0	- 19.5
Total Average t = 1.23 df = 9	38	+ 12.4

### TABLE 6B

Test of significant difference for percent of increase or decrease in examiner accuracy before and after a Stimulation Test - verified deceptive subject's Known Card Test versus Standard Card Test

Examiner	KCT	SCT
1	+ 11.1	- 30.0
2	- 37.5	- 14.3
3	- 40.0	- 22.2
4	- 10.0	- 37.7
5	0	- 20.0
6	+ 20.8	- 20.8
7	- 11.1	- 27.5
8	0	- 20.0
9	0	- 25.0
10	- 37.5	- 14.3
t = 1.61		
df = 9		

Polygraph 1981, 10(2)

=

Finally, since indefinite decisions were not computed in overall accuracy statistics (only definitive decisions were judged correct or incorrect) a T-Test was tabulated to determine whether or not the decreases in indefinites after the stimulation test were significant. Examiners indefinite decisions were reduced 2% after the standard card test (20% to 18%) and 1.5% after the known card test (17% to 15.5%). The T-Test revealed no significant difference at the 99% level of confidence in the decrease of examiner indefinite decisions from the first to third test after either the known card test or standard card test.

				10114	'					
Test	of	significant	difference	for	percent	of	indef	inite d	lecisions	
before	and	l after a St	imulation Te	est ·	- Known	Card	l Test	versus	s Standard	l
			Car	rd Te	est					

**TABLE 7** 

Examiner	KCT	SCT
1 2 3 4 5 6 7 8	0 + 10 - 5 + 5 - 5 + 15 - 5 0	- 10 - 15 - 10 - 5 + 15 + 10 0 - 10
9 10 t = 0.93	- 10 - 30	- 10 + 5
f = Q		

## DISCUSSION

The results clearly indicate that the only significant difference between the two techniques is that the known card test induces a greater percentage of deceptive subjects to distort their responses to the chosen card, which in itself is highly indicative of guilt(Reid & Inbau, 1976). The distortion is usually readily identifiable. It should be noted that truthful subjects given a known card test are not induced to attempt countermeasures. It is possible that by the nature of the known card test instructions, wherein the examiner stresses that he will have a "picture of the subject's lie" the deceptive subject is tempted to exaggerate his lie response since he feels it will be used for comparison purposes with the issue test. This verbal conditioning of the subject may in fact be the key to inducing the guilty to distort their responses, however, additional study would be needed to address the hypothesis.

Concerning the stimulation effect on those subjects who did not distort their responses to the chosen card, while there is no statistically significant difference, the standard card test produced a greater percentage of significant to moderate responses, particularly with deceptive subjects. This may indicate a greater fear of detection with the standard card test, though no specific conclusion can be drawn from this study.

#### Steven L. Kirby

Relative to examiner accuracy in evaluating records, under both procedures, examiner accuracy decreased in that examiners were more accurate by 7.4% in evaluating charts administered before the stimulation test. While there was no significant difference in the decrease, it does open the question of the overall effectiveness of any type of stimulation test. This is in direct conflict with Senese's study(1976b). It has also been the writer's personal experience that both the known and standard card tests have been an invaluable aid in stimulating more definitive results. Perhaps the problem is that in far too many cases, examiners are attempting to stimulate subjects who are already properly conditioned.

It has been expected and shown that in blind studies, examiner accuracy is approximately 87%(Horvath & Reid, 1971b, Hunter, 1971, Slowik, 1975). These studies allowed examiners to evaluate several issue tests, along with the stimulation tests. Even though this study was not meant to be, nor should it be construed as, a reliability or validity study, overall examiner accuracy on evaluating the first chart only, was 79.5% for the known card test and 72.9% for the standard card test. The high degree of accuracy in this study with very limited data(<u>i.e.</u>, only one chart), may be due to the examiner's expertise, but based on the decrease in accuracy after the stimulation tests (both known and standard) is probably the result of the use of very clear cut test records. If the latter is the case, then discernable patterns on the first chart should be indicative that the subject is properly stimulated.

In this study, the examiner's accuracy for those subjects verified truthful actually increased an average of 4.3% while examiner accuracy for those subjects verified deceptive dropped 15.8%. This data falls in line with the theory that some deceptive subjects who feel that the polygraph has detected their lie, give up, become less emotional, and therefore less responsive on each subsequent test(Reid & Inbau, 1977c). It is therefore suggested that stimulation tests may be overused in the polygraph technique and may be destimulating subjects, particularly deceptive subjects.

The final area of comparison concerns the effect each procedure has on the examiner's rate of indefinite decisions. Following both the known and standard card test procedures, indefinite decisions decreased. Neither technique proved significantly different from the other in this respect. This indicates that both techniques are equally effective in reducing inconclusive records. However, considering the decrease in accuracy, the reduction of inconclusives may be at the cost of overall accuracy.

In conclusion, since there is no significant difference in the effectiveness of the two techniques except that the known card test induces more purposeful distortion to the chosen card for deceptive subjects, it can be stated that the known card test is as effective as the standard card test.

In light of the public's growing distrust and negative feeling towards polygraph examinations(Harris, 1979), it behooves all examiners to be as straightforward and professional as possible. This study shows that technique need not suffer by altering the standard card test which has been the subject of criticism. Since this can be done without effecting accuracy, the known card test should be employed in place of the standard card test; but should not be used in cases where definitive records are available on the initial chart.

#### Comparison of Two Stimulation Tests

#### Footnotes

1. Horvath, Frank S. and Reid, John E., "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception." <u>Journal of Criminal Law</u> and Police Science 62(1971): 276-376.

2. Inbau, Fred E. and Reid, John E., "The Lie Detector Technique: A Reliable and Valuable Investigative Aid." <u>American Bar Association Jour</u>nal 50(5)(May 1964).

3. Wicklander, Douglas E. and Hunter, Fred L. "The Influence of Auxiliary Sources of Information in Polygraph Diagnoses." <u>Journal of Po</u>lice Science and Administration 3(4)(1975): 405-409.

4. Reid, John E. and Inbau, Fred E. <u>Truth and Deception; The Poly-</u> graph (Lie Detector) <u>Technique</u>, 2d ed. Baltimore, Maryland: Williams & Wilkins Co., 1977, pp. 42-43.

5. Senese, Louis C. "Accuracy of the Polygraph Technique With and Without Card Test Stimulation." Journal of Police Science and Administration 4(1976): 274-276.

6. Star, Jack. "This Man Says You're a Liar." <u>Chicago Magazine</u> (September 1977): 170.

7. Senese, pp. 274-276.

8. Reid and Inbau, 1977b, pp. 102-109.

9. Senese, 1976b, pp. 274-276.

10. Horvath and Reid, 1971b, pp. 276-376.

11. Hunter, Fred L. and Ash, Philip. "The Accuracy and Consistency of Polygraph Examiners Diagnosis." Journal of Police Science and Administration 1(3)(1971): 370-375.

12. Slowik, Stanley M. and Buckley, Joseph P. "Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure and GSR Recordings." <u>Journal of Police Science and Administration</u> 3(3)(1975): 305-310.

13. Reid and Inbau, 1977c, pp. 98-100.

14. Harris, Louis & Associates, Inc. "The Dimensions of Privacy." A <u>National Opinion Research Survey of Attitudes Towards Privacy</u>. Conducted for Sentry Insurance, 1979, p. 35.

[Note: The writer thanks Patricia Parker for her editing and Joseph Buckley for his assistance in all phases of this study. The writer is also indebted to Eugene Gerke for the statistical tabulations. Finally, without the help of Mr. John E. Reid and his Chicago staff, this paper could not have been written.]

## ACCURACY DEMONSTRATIONS, THREAT, AND THE DETECTION OF DECEPTION: CARDIOVASCULAR, ELECTRODERMAL AND PUPILLARY MEASURES

By

## M.T. Bradley and Michel Pierre Janisse

#### Abstract

Subjects, half of whom committed a mock crime, were examined with both Control Question and Guilty Knowledge tests in an attempt to detect guilt or innocence. Skin resistance, heart rate, and pupil size were the physiological measures employed. Prior to the polygraph test the effectiveness of the physiological detection apparatus was demonstrated to each subject. They were led to believe they had been detected in either none, one, two or three of three trials in this demonstration. Subsequent detectability of subjects varied as a result of the manipulation with the Control Question test such that detectability increased as the level of demonstrated effectiveness Results with the Guilty Knowledge test were less increased. Another manipulation, threat of punishment, did not clear. affect detection results but did alter heart rate change such that those threatened, whether guilty or innocent, received rank scores more in the guilty direction than those not threa-Skin resistance was the most efficacious measure in tened. both tests while pupil and heart measures varied in detection accuracy depending on the test employed.

Responses of physiological systems under the innervation of the autonomic nervous system have been measured and used in both laboratory and field investigations to detect attempts at deception by subjects or suspects being questioned (Orne, Thackray, & Paskewitz, 1972). The use of physiological measurement as a reflection of cognitive events, especially lying, has a long history and has gained wide acceptance in applied fields such as criminal investigation.

Field investigators (e.g., Barland & Raskin, 1975) and laboratory investigators (Lykken, 1974) have assumed that the effectiveness of lie

Address requests for reprints to either author: M.T. Bradley, Division of Social Science, University of New Brunswick, P. O. Box 5050, St. John, New Brunswick, E2L 4L5, Canada; or Michel Pierre Janisse, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

©1981, The Society for Psychophysiological Research. Reprinted with permission of the publisher from PSYCHOPHYSIOLOGY, 1981, 18, 307-315.

The authors wish to acknowledge the advice of Anis Mikhail, Roland Penner, Robert Stelmack and John Schallow on earlier versions of this manuscript.

detection techniques would be enhanced after demonstrations of accuracy. Such demonstrations typically consisted of card tests which, unknown to the subject, had been arranged to make the testing technique appear infallible (Reid & Inbau, 1977). Systematic studies of this manipulation's effectiveness have not been conducted in the field and laboratory studies have not only failed to support the above assumption but have yielded the opposite result. Ellson, Davis, Saltzman, and Burke (cited in Gustafson & Orne, 1965) found that subjects who were given feedback that their lying had been detected on the initial trial were actually more difficult to detect on the following trial than subjects who were told they had not been detected. Davis (1961) is accounting for this decrease in detectability suggested that guilty subjects, after demonstrations, could be classed in three categories. Those certain of detection would be resigned, pay little attention to the detection proceedings, and as a consequence would be autonomically unresponsive. Those who were uncertain would be more attentive and therefore differentially more responsive to critical items. Finally, those subjects who were convinced the device was ineffective would be relaxed and difficult to detect. Davis (1961) indicated that a single trial demonstration of ineffectiveness would create uncertainty and result in subjects being more detectable.

Bradley and Janisse (1979), agreeing that a single trial demonstration was probably inadequate for the subject to form a strong impression of the detector's effectiveness, gave a series of three card tests. Through deception, their subjects were given a demonstration that the lie detector measuring the pupillary response was either 0%, 33 1/3%, 66 2/3%, or 100% effective. On the subsequent fourth trial detection results over groups formed an inverted U shaped function with the 0% and 100% groups more difficult to detect than the uncertainty groups.

Although the above results explain prior laboratory findings, including those from a study by Barland and Raskin (1973), they are at odds with results predicted by field investigators. The present study was designed to identify possible reasons for these differences. A mock crime paradigm involving threat of punishment was used in an attempt to simulate actual field conditions (Podlesny & Raskin, 197). All subjects were interrogated by two techniques, a Control Question test (Backster, 1969) commonly used in field investigations and the Guilty Knowledge test (Lykken, 1974) commonly used in laboratory studies. Measures of skin resistance (SR) and heart rate were included in addition to the pupillary response because of their frequent use in the detection of deception paradigms.

Punishment or threat of punishment has received little attention in the detection of deception literature. For the sake of realism, it would seem worthwhile not only to include a punishment or threat variable, as have Barland and Raskin (1975) and Lykken (1959), but also to manipulate this variable to see if it does alter detection rates. Since in actual criminal interrogations, both guilty and innocent suspects may entertain the possibility of imprisonment, half of each group in the present experiment were threatened with shock as a consequence of a judgment of guilt.

Of the three measures used, SR has been consistenty effective in virtually every laboratory study of deception (Barland & Raskin, 1973). Heart rate measures have not been so effective but have yielded detection rates better than chance (Orne et al., 1972; Podlesny & Raskin, 1978; and Raskin & Hare, 1978) and may be sensitive (responsive) to different

factors in detection (Clark, Note 1). Janisse (1977) noted that there is a dearth of published studies using the pupillary response in the detection of deception (e.g., Berrien & Huntington, 1943; Heilveil, 1976), but recently, Bradley and Janisse (1979) and Janisse and Bradley (1980) utilized it effectively in a card test form of the Guilty Knowledge test. The present study, through the inclusion of all three measures, provided the opportunity to assess their relative effectiveness.

## METHOD

### Subjects

Subjects were 192 Caucasian male university students who took part in the experiment to fulfill an Introductory Psychology course requirement.

#### Apparatus

The pupillary response was measured by a Whittaker Space Sciences Eye View Monitor and Television Pupillometer designed to provide an accurate assessment of pupil diameter and to record the data in digital form on a Kennedy incremental tape recorder, Model 1600/360. This apparatus was set up in a white experimental room illuminated by five 100-watt bulbs placed directly above the subject and approximately  $3 \ 1/2$  meters from a visual target. The bulbs were connected to a 25-ampere constant voltage transformer to provide a steady non-fluctuating power source to prevent change in illumination due to power surges in the external electrical supply. Heart rate was assessed by a Whittaker Pulse Watch which was designed to output data on the same Kennedy recorder as the pupillometer. Skin resistance (SR) was recorded on a two-pen chart drive with one pen serving as an event marker (Lafayette Datagraph Psychogalvanometer, Model # 77010). The medial phalanges on the first and third fingers served as recording sites and were fitted with Lafayette (cat. # 76602) electrodes after the fingers had been cleaned with a cotton swab dipped in alcohol. The electrodes. attached without electrolyte, were curved to conform to finger shape and were held in place by velcro wraps. The SR recordings were made in the AC mode to eliminate problems associated with slow drift encountered with DC recording. This was judged an adequate method of recording since scores were based on the relative magnitude of responses within each subject.

Verbal stimuli were presented via a two-channel Sony tape recorder and the onset of each stimulus was marked on both the digital tape and paper chart by pressing connected hand buttons. A battery powered shock device using an automobile coil was equipped with electrodes embedded in a cloth wrap so that these electrodes could be mounted snugly on a subject's leg. The first room the subject entered contained a Minute Minder bell timer.

#### Procedure

Subjects who had left their names and telephone numbers some weeks earlier during a session of filling out personality inventories were contacted by telephone and told to report to a specific room which was located near the detection laboratory. All subjects upon entering the room found a note taped to the table directing them to close the door and set a bell timer for five minutes to allow them the privacy and time to read and

carry out the instructions which had been placed on the table before their arrival. Innocent subjects read that they were to sit in the room for the time interval and then open the door to wait for the arrival of the interrogator. Guilty subjects read that the room they were in was part of a store and that hidden at the back of a shelf under the table was one dollar in a medium sized white envelope. Each subject was to steal the dollar and hide it in his right front pants pocket. After committing the crime they concealed the instructions, waited for the time interval to end, opened the door and sat waiting for the interrogator.

All instructions informed subjects that they would be accused of committing a theft and would be subjected to a polygraph examiner using measures of physiological activity to determine their guilt or innocence. Subjects weere informed that during the course of the procedure from the moment of being accused to the end of the experiment they were to deny any involvement in our knowledge of the crime. They could of course give their name and answer freely questions about matters not directly relevant to the crime.

Half of the guilty and half of the innocent subjects were informed that they would receive a painful but not permanently damaging electric shock if judged guilty. In reality no electric shock was ever given.

Several precautions were taken to keep the interogator blind to the actual guilt or innocence of a given subject. The instructions had been folded, stapled and randomly inserted into 192 portfolios by an assistant who retained a master sheet until after the completion of the experiment. These portfolios contained coded information on the shock and certainty conditions but not on the guilt or innocence of the subjects. The money was kept in the first room at all times by checking the envelope and restocking it as necessary. The presence or absence of money could inform the examiner whether the previous subject had been guilty but this was after the polygraph test and could not lead to any predictions about the next subject. Finally, the interval spent in the room was timed equally for all subjects so that the examiner could not guess who carried out the more time-consuming guilt instructions.

The examiner, cued by the bell of the timer, arrived at the open door, asked for the subject's name, accused him of the crime and led him to the test room. Once there the subject was shown the equipment and was given a little information about the lie detection procedure. Subjects in the threat of shock condition were shown the shock apparatus and electrodes and were reminded that they could receive shock whether guilty or innocent if they were judged guilty. It was also explained that those judged guilty if actually guilty would forfeit the stolen money. 0f course, with a judgment of innocent a subject would be free to leave. At this point subjects were reminded that they could withdraw from the experiment and still retain their experimental credits. In either instance, continuing or not, subjects promised not to divulge details of the experiment. Those who continued then filled out a state anxiety scale (Spielberger, Gorsuch, & Lushene, 1970).

In preparation for the polygraph test the questions on the Control Question test were reviewed with the subject. The crucial part of the test contained crime-relevant questions paired with control questions. Prior to the review, subjects were reminded that they must, on the one

hand, answer "No" to all crime-relevant questions because the innocent were truly innocent while the guilty were attempting to conceal their guilt. On the other hand all subjects were to feel they could discuss questions about their past life so that ambiguities could be clarified, and make it possible to answer these questions with a "No." This was done to make the verbal responding comparable between each of the crime-relvant and control question pairs and was achieved by making a standard minor modification, "Except for what you have told me ...," in the control question. The questions and sequence were as follows:

1. Is your name ?

2. Regarding the money, do you intend to answer all quustions about it truthfully?

3. Are you convinced that I will only ask questions on this test that you have already okayed?

- 4. Before the age of 18 did you ever steal any money?
- 5. Did you take the money?
- 6. While you were in school did you ever take anything of value?
- 7. Did you steal the money from the room?

8. Before the age of 18 did you ever cheat anyone who trusted you?

9. Do you know what happened to the money?

Questions on the Guilty Knowledge test were not reviewed and subjects were only informed that they would be asked a series of questions about items containing certain information only the guilty subject knew. This information had been presented in their instructions and an example of a question set is presented below.

Regarding the money taken. 1. Was it ten dollars? 2. Was it five dollars? 3. Was it one dollar? 4. Was it twenty dollars? 5. Was it fifteen dollars?

After being prepared for the polygraph test, which included strapping on the shock electrodes in the threat of shock condition, subjects took a series of card tests ostensibly designed to assess their detectability. They were given three trials and after each were informed the number the equipment indicated they had attempted to conceal. Depending on the group, subjects were detected on none, one, two or three trials to create the impression that the detection measures were ineffective, sometimes effective, or perfectly effective. In order to carry out this demonstration, the experimenter rigged the card tests so that unbeknownst to the subjects they chose a predetermined number. In this manner the experimenter would feed back the correct number for a hit or some other number for a miss regardless of the actual performance of the equipment. Subjects kept a concealed score of the results of this test by writing down the number they chose and the number indicated. They did this because the examiner had informed them that he was very interested in studying this information in the future. In actuality the record served as a manipulation check and prevented the subject from being inattentive or forgetful of the results.

Following the above manipulation subjects filled out the state anxiety scale for the second time and were readied for the two test Polygraph 1981, 10(2) techniques, the order of which was counterbalanced across all subjects. The Control Question test was presented three times in a row which is typically the minimum number of times for an investigation. The Guilty Knowledge test was presented once. After the polygraph test the subject filled out both the State and Trait scales of the Spielberger anxiety inventory. Under the assurance that the information would not be examined prior to the judgment of guilt or innocence the subjects rank ordered the questions in each question technique assigning a one to the question they subjectively felt resulted in the largest autonomic response, a two to the second most evocative question, and so on for both question techniques. They then answered a post-experimental questionnaire, were given the judgment of guilt or innocence, were briefed, were reminded of their promise not to reveal critical information, and were released.

The questions in both techniques had been prerecorded for presentation so that the examiner's vocal inflections would not vary across subjects. The recorded questions were spaced at 20-sec intervals and it took 8 min to deliver the 24 questions on the Guilty Knowledge test. The 9 questions on the Control Question tet were recorded at 20-sec intervals and took 3 min to present. There was a pause while the tape was rewound before the second and third presentations. The Control Question test was individualized by dubbing in the subject's name and the phrase "Except for what you have told me" as necessary for the control questions. The six questions for the series of three card tests used for the effectiveness manipulation had been pre-recorded at 10-sec intervals and it took 1 min to deliver each test.

## Data Analysis

The data were analyzed on different levels. Detection scores most relevant to field work were derived by a modification of procedures described by Barland and Raskin (1975) for the Control Question test and by Lykken (1959) for the Guilty Knowledge test. With the Control Question test each control and relevant question pair was assigned a score of 1, 0, or -1, depending on whether the response to the control question was larger, the same, or smaller. The test had 3 such question pairs and was repeated 3 times for a scoring range between +9 and -9. Subjects with scores of +2 or greater were classed as innocent while those with -2 or less were classed as guilty. Subjects with scores between these values were not judged in reference to detection but categorized as inconclusives. With the Guilty Knowledge test each of the 4 question sequences received a score of 2, 1, or 0 depending on the relative magnitude of the response to the critical item. If the response to the critical item was largest it received a score of 2, if second largest, 1, and finally 0 for any other response magnitude. In this test with 4 5-item sequences, plus a buffer item per sequence, a score of 4 or more was classed as deceptive and less than 4 as innocent. The frequency of detection was reported from these scores.

The numerical data derived by the procedures described above were subjected to analyses of variance to test the major hypotheses. The basic design was a 4 x 2 x 2, four levels of manipulated effectiveness, two levels of punishment, and two levels of guilt.

All of the above analyses used the data from three dependent measures: pupil size change, heart rate change, and skin resistance. Pupil Polygraph 1981, 10(2) size and heart rate values were calculated for each second of recording in mm and bpm, respectively.

Pupil size change was derived by subtracting the average of the 3 sec of baseline immediately prior to the appropriate question from the highest value obtained in the 10 sec following the question. In a similar manner heart rate change was obtained by subtraction of the baseline. The response value, however, was derived from the lowest rate in the 15 sec following the question. SR was continuously output on a chart and the responses of interest were measured in mm of pen deflection in a 10-sec period immediately following the question.

#### RESULTS

Accuracy of Detection

The classification of subjects into categories of deceptive and nondeceptive in relation to whether they were actually guilty or innocent is displayed in Table 1 for each of the tests (Control Question and Guilty Knowledge) and the three physiological measures. The accepted level of significance was p < .05.

In considering the Control Question test it should be noted that an additional category of inconclusive was used. Examination of the x-squared analyses in the table revealed that scores were non-randomly distributed across the six categories for SR and heart rate change measures. When inconclusives were excluded from the analyses it was found that correct decisions were made more frequently than incorrect decisions with SR,  $x_1^2 = 59.78$ , and heart rate change,  $x_1^2 = 5.96$ .

With the Guilty Knowledge test SR was most effective in correctly classifying subjects. Success in this test was due mainly to the correct classification of innocent subjects rather than to the identification of the guilty. A lowering of the cutoff criterion for a guilty judgment would result in more correct guilty classifications but at the expense of classifying more innocent subjects as guilty.

Hypotheses Tests on Detection Scores

The detection scores for each of the three dependent measures in each of the two tests were individually examined in a three factor analysis of variance. The factors were levels of manipulated effectiveness, punishment, and guilt.

<u>Control Question</u>. No significant main effects were found for the levels of manipulated effectiveness for any of the three dependent measures. The punishment factor was significant (F(1/176) = 5.08) for change in heart rate such that subjects not threatened with shock tended to receive scores more in the innocent direction(X = .67) than those threatened with shock (X = -0.41). Both SR  $(F(1/176) = 83.95; X_1 = 1.92, X_G = -1.98)$  and heart rate change  $(F(1/176) = 8.72; X_1 = 83, X_G = -.57)$  discriminated between guilty and innocent subjects.

Manipulated effectiveness was found to interact with guilt for SR scores (F(3/176) = 4.04) such that the more effective the apparatus appeared to be, more innocent subjects scored as innocent and more guilty

83

subjects scored as guilty. This interaction is displayed in Figure 1. No other interactions with SR or other dependent measures were found to be significant.

TABLE 1 Frequency of Detection

## Skin Resistance

Subjects	Contro	ol Que	estion	Guilty K	nowledge Test
	D[a]	ND	I	D	ND
Innocent	58	13	25	57	39
Guilty	9	56	31	11	85

 $X_{2}^{2} = 64.62*$   $X_{4}^{2} = 46.11*$ 

-	٠	-	~	٠		~ 1		
Pun	1	1	S	1	70	Ch	on.	αe
I U P	· .	τ.	D	-	20	0.1	an	<u> </u>

	$X_{2}^{2} = 7$	.06*		X <b>2</b> = 15.15*	
Guilty	19	32	45	17	79
Innocent	34	20	42	43	53
		Heart	Rate	Change	
	$X_{2}^{2} = 0$	.50		$X_{\prime}^{2} = 3.84*$	
Innocent Guilty	33 30	21 25	42 41	32 19	64 77

D[a] = deceptive, ND = non-deceptive, I = inconclusive. \* p < .05.

<u>Guilty Knowledge Test.</u> Manipulated effectiveness was significant for SR (F(3/176) = 4.175). A trend analysis on the SR data revealed significant quadratic (F(1/176) = 5.68) and cubic trends (F(1/176) = 4.14), p < .01, but no linear component.

None of the dependent measures significantly discriminated between the punishment conditions. All of the dependent measures discriminated significantly between guilt and innocence: SR, F(1/176) = 72.87; X<sub>1</sub> = 1.84, X<sub>G</sub> = 3.90; pupil change, F(1/176) = 5.08; X<sub>1</sub> = 2.34, X<sub>G</sub> = 2.88; heart rate change, F(1/176) = 13.98; X<sub>1</sub> = 2.29, X<sub>G</sub> = 3.23. None of the interactions, either two-way or three-way, for any of the three dependent measures were significant.

Subjective Ranks

The mean subjective estimates of physiologically responsive (1 = most physiologically responsive to 9 = least physiologically responsive) to the 6 crucial questions, 3 control and 3 crime-relevant, in the Control Question test, are displayed in Table 2. The differences between the guilty

and innocent sugjects were significant on the multivariate package of 6 crucial questions and on all of the individual univariate tests for each question. The F(1/176) values ranged from a low of 68.52 to a high of 144.77. Interestingly, the accuracy of judging guilty or innocence from the subjective rank orders (164 correct, 1 inconclusive, 27 incorrect) treated in the same manner as physiological detection scores actually exceeded (x-squared = 8.6) the accuracy obtained with the SR (114 correct, 60 inconclusive, 22 incorrect).





#### DISCUSSION

## Manipulated Effectiveness

Clear support for the hypothesis proposed by field investigators  $(\underline{e.g.}, \text{Reid \& Inbau}, 1977)$  and by Lykken (1974) was found for the Control Question teste using the skin resistance measure. The detection scores diverged in the appropriate directions (positive for innocent, negative for guilty) as demonstrated effectiveness improved. From a technical standpoint users of this test would be justified in fostering the belief that the detection equipment is highly effective.

#### TABLE 2

Mean subjective ratings of crucial questions on the Control Question Test

	Mean Ratings								
	Contr	ol Que	stions	Crime Relev Questions		vant			
Subjects	4	6	8	5	7	9			
Innocent Guilty	2.53 5.01	3.29 5.37	2.87 5.41	5.26 2.37	5.21 2.77	6.00 3.56			

This experiment did little to clarify the relationship between the effectiveness and the Guilty Knowledge test. The significant quadratic trend with the skin resistance measure did show that detection as predicted by Davis (1961) and demonstrated with the pupil by Bradley and Janisse (1979) was maximized in uncertainty conditions. However, this

appeared as a main effect rather than in a guilt by effectiveness interaction which suggested that the manipulation was affecting both guilty and innocent subjects. Theoretically, the scores of innocent subjects should have resulted in a horizontal function because such subjects were unaware of the critical items and thus could not be differentially influenced on such items by the effectiveness demonstrations. The failure to fully replicate the Bradley and Janisse (1979) study may be related to differences in experimental design. In contrast to their experiment, the present research employed the mock crime paradigm and a threat of punishment manipulation while the former study used only a simple card test.

## Detection

Because all subjects were examined with both types of test techniques it may seem approrpriate to compare the 84% correct (60% including inconclusive) level achieved on the Control Question tets with the 74% level achieved with the Guilty Knowledge test. For many reasons, such a comparison should be avoided. The present study used a relatively weak version of the Guilty Knowledge test. Davidson (1968) and Podlesny and Raskin (1978), for example, achieved accuracies of 98% and 90% respectively in studies using more items of information. Conditions were also less than optimal for the Control Question test. Podlesny and Raskin (1978) with Control Question tests using a "semi-objective" scoring technique, a seven point judgment scale for rating each question pair, repetition of trials for subjects judged inconclusive, and composite scores derived from several physiological measures, achieved 89% levels of accuracy. Experiments which optimize conditions as far as possible for each type of test (Podlesny & Raskin, 1978) or better still, which manipulate these and other variables, would provide the most appropriate information for judging the relative accuracy of the two techniques.

### Threat of Shock

In an attempt to provide motivation to avoid detection, while also manipulating that motivational variable, half of the innocent and guilty subjects were threatened with shock if judged guilty. This form of motivation did not prove effective on the detection scores except with heart rate change on the Control Question test. The finding of a single measure affected by threat of shock should not be considered as anomalous since such a result fits well with recent studies in the literature using similar measures. Stanners, Coulter, Sweet, and Murphy (1980) found heart rate sensitive to shock threat while pupil size and GSR did not evidence Dumoff (1978), using an ego threat rather than a shock such effects. threat to increase emotional arousal, found that heart rate was affected while pupil size was not. In both of the above studies pupillary responses were sensitive to varied levels of mental effort. A further analysis of continuous physiological data available in the present study revealed that shock threat was related to differences in heart rate change for both tests while it was not related to pupil differences. Such results suggest that investigations of threat of punishment in relation to cardiovascular measures may be a profitable avenue of research in psychophysiological detection of deception.

One caveat for the interpretation of shock results is that while state anxiety, assessed periodically throughout the examination, was higher for guilty subjects than innocent subjects, there was no extra POlvgraph 1981, 10(2) 86 increment in anxiety for those under threat of shock. Possibly the shock threat manipulation was not effective. Interestingly, other factors could have been at work since a post-experimental questionnaire revealed that fewer subjects in the shock group expected to be detected. A speculation is that subjects may have ameliorated shock threat effects by not admitting or denying the possibility of detection.

### Subjective Ratings

A controversy over the utility of Control Question tests in the detection of deception has now raged for some years across several journals (e.g., Dawson, 1980; Lykken, 1974, 1978, 1979; Podlesny & Raskin, 1977, 1978; Raskin & Podlesny, 1979). At issue is whether the Control Question test can, as they were designed to do, evoke the attention and thus the autonomic response of innocent subjects to the control questions while focusing the guilty suspect's attention on the "crime relevant" questions. The rationale is that control questions in the test are of an emotional nature because they ask suspects about possible misdemeanors or crimes in which they could have been involved in the past. The innocent person, it is hypothesized, will be responsive because the implications of such questions reflect ill on his character. In effect, not only is he accused of the particular crime but now he is being regarded as a highly suspicious character. An aggravating factor is that the questions are of such a general nature that it is easy to feel uncomfortable with one's answer. Tn comparison, the relevant questions are easier to answer because the innocent suspect knows he is truthful. The guilty person ignores the control questions because he realizes he is being tested for the specific crime and questions about his past life are irrelevant. Lykken (1978) has decided it is naive to think control questions would elicit the attention of, or cause much concern for, innocent suspects whereas Raskin and Podlesny (1979) have decided that control questions should cause great concern for innocent subjects. The controversy has not been easy to resolve because of the somewhat confusing story on detection accuracy (Bersh, 1969; Horvath, 1977; and Lykken, 1979). In the present experiment data were collected which bear directly on how the suspects perceived the control and relevant quustions. After the polygraph test, but before they received the judgement of guilty or innocence, subjects rank ordered the series of questions on the Control Question test assigning a one to the question they believed evoked the largest autonomic response and a nine to the least autonomically evocative question. This rating was made with no knowledge of the test results or actual autonomic responsiveness and, very importantly, under the assurance that the ratings would not be used in the In fact, the suspect did not physically yield the judgment process. rating sheet until the judgment was delivered and in the case of shock threat the shock electrodes were removed.

The mean subjective ratings for the 6 crucial questions, 3 control and 3 relevant, shown in Table 2, suggest that innocent subjects perceive themselves as responding more to control questions while guilty subjects perceive themselves as responding more to relevant questions. Taken at face value, these results are consistent with predictions made by proponents of the Control Question technique (Podlesny & Raskin, 1977) and are contrary to the criticism that it is unreasonable to assume that innocent subjects will be more concerned about control questions than relevant questions (Lykken, 1979). However, an alternative explanation is that the subjective ratings merely reflect the subjects' expectations about how

they should respond and do not indicate true concerns during the test. While the latter interpretation is consistent with the low and non-significant correlations between the subjective ratings and the physiological responsivity during the test, the present data cannot conclusively support either of the two alternative theoretical explanations.

The physiological data are of more direct relevance to the issue of whether the control questions served their intended fucntion. The SR data shown in Table 1 indicates that the majority of the innocent subjects (56/96) responded predominately more to the control questions, a sizable minority (31/96) responded similarly to control and relevant questions, and only a small minority (9/96) responded predominately more to the relevant questions. Of the innocent subjects who responded differentially to control and relevant questions, 86% responded more to the control questions. Practically the opposite pattern of results was obtained with the guilty subjects. Thus, the control questions served their intended function for most, but certainly not all, of the subjects in the present study. Of course, whether these results would generalize to an applied setting where the emotional concerns may differ qualitatively from those experienced in the laboratory remains a question for future study.

## Summary and Recommendations

Manipulation of demonstrated effectiveness altered detection levels in both the Control Qeustion test and the Guilty Knowledge test. Threat of shock had an effect on continuous heart rate change but not on the other measures. Skin resistance was the most effective measure for both tests.

The results of this study not only support an assumption commonly made by field workers employing Control Question tests but may even provide some resolution to a problem engendered by this assumption. The assumption that card test demonstrations portraying detection devices as highly effective enhance accuracy has led to the ethically questionable practice (Lykken, Note 2) of informing suspects no matter what the outcome of the card test that they have been detected. According to our findings such a procedure may be unnecessary. High levels of control test accuracy are found after card test demonstrations indicating 66% to 100% effectiveness (see Figure 2). This range embraces the median accuracy levels of 73% (Orne et al., 1972) found in card test studies as well as the levels of detection found in mock crime studies. A straight-forward procedure of giving a suspect several card tests with feedback of the actual results should provide, in a probabilistic sense, not only an accuracy enhancing effectiveness demonstration but an ethical one as well. The cost of such a procedure is that by chance, some suspects would receive feedback indicating detection effectiveness outside of the optimal range. Whether the examiner would then continue with a polygraph test or decide that the suspect should not be examined is another question, one that is beyond the scope of this paper.

In contrast to the results obtained with the Control Question test, the present study did not isolate a specific range of demonstrated effectiveness that promotes optimal detection with the Guilty Knowledge test. However, it was clear that demonstrations of relatively low effectiveness (33%) did not detract from its subsequent effectiveness. One possibility is that other than demonstration of complete ineffectiveness, the Guilty

Knowledge test is impervious to such manipulations. Another possiblity is that there is an interaction between effectiveness and salience of items. One prediction that might be tested is that low salience items would be affected very much by effectiveness manipulations while high salience items would not.

Another recommendation of this study is that effort should be aimed at identifying the optimal conditions for various physiological indices. Although the heart rate and pupil measures did not perform as well as SR, there is promise that these measures could perform well in specific situations such as an emotional interrogation (heart rate) or an interrogation concentrating on information (pupil).

Threat of shock had little effect on detection accuracy. Such a result could encourage the laboratory study of detection since threat was designed to be a step on a continuum shared with legal penalties at the most noxious extreme. If basic detection results are not altered over levels of this continuum, laboratory studies would, for reasons of control, remain the most attractive vehicles for the study of the physiological detection of deception. On the other hand, the possibility exists that threat of shock in the laboratory is either different in kind or such a small step on the continuum compared to the real life situation that a gap exists. It would seem worthwhile to encourage the continued development and testing of more powerful threat manipulations to more thoroughly understand the difference between field and laboratory situations. Imaginative experimenters have already provided realistic laboratory situations appropriate for these questions. Balloun and Holmes (1979) led subjects to believe that they would be expelled from the university if detected after they had cheated on an exam. Kugelmass and Lieblich (1966) told police cadets that they would not be able to continue their education if they could not avoid detection. Intuitively, such threats seem powerful and if included in studies with appropriate ocntrols (non-threatened individuals), may indicate whether the threat variable is important in the process of detection.

#### References

Backster, C. <u>Technique fundamentals of the tri-zone polygraph test</u>. New York: Backster Research Foundation, 1969.

Balloun, K.D., & Holmes, D.S. Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. Journal of Applied Psychology, 1979, <u>64</u>, 316-322.

Barland, G.H., & Raskin, D.C. Detection of deception. In W.F. Prokasy & D.C. Raskin (Eds.) <u>Electrodermal activity in psychological re-</u> search. New York: Academic Press, 1973, pp. 418-477.

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

Berrien, F.J., & Huntington, G.H. An exploratory study of pupillary responses during deception. Journal of Experimental Psychology, 1943, 32, 443-449.

Bersh, P.J. A validation of polygraph examiner judgments. <u>Journal</u> of Applied Psychology, 1969, 53, 399-403.

Bradley, M.T., & Janisse, M.P. Pupil size and lie detection: The effect of certainty on detection. Psychology, 1979, 4, 33-39.

Davidson, P.O. Validity of the guilty knowledge technique: The effects of motivation. Journal of Applied Psychology, 1968, 52, 62-65.

Davis, R.C. Physiological responses as a means of evaluating information. In A.D. Biderman & H. Simmer (Eds.) <u>The manipulation of human</u> behavior. New York: Wiley, 1961, pp. 142-168.

Dawson, M.E. Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. <u>Psy</u>-chophysiology, 1980, 17, 8-17.

Dumoff, M.G. Discriminating emotion and mental effort with autonomic measures: Pupil size and heart rate as differential measures of cognition and anxiety. Unpublished doctoral dissertation, University of Manitoba, 1978.

Gustafson, L.A., & Orne, M.T. The effects of verbal responses on the laboratory detection of decepiton. Psychophysiology, 1965, 2, 10-13.

Heilveil, I. Deception and pupil size. Journal of Clinical Psychology, 1976, 2, 10-13.

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

Janisse, M.P. Pupillometry: <u>The psychology of the pupillary res</u>ponse. Washington, D.C.: Hemisphere, 1977.

Janisse, M.P., & Bradley, M.T. Deception, information and the pupillary response. Perceptual & Motor Skills, 1980, 50, 748-750.

Kugelmass, S., & Lieblich, S. Effects of realistic stress and procedural interference in experimental lie detection. Journal of Applied Psychology, 1966, 50, 211-216.

Lykken, D.T. The GSR in the detection of guilt. <u>Journal of Applied</u> Psychology, 1959, 43, 385-388.

Lykken, D.T. Psychology and the lie detector industry. <u>American</u> Psychologist, 1974, 29, 724-739.

Lykken, D.T. The psychopath and the lie detector. <u>Psychophysiology</u> 1978, 15, 137-142.

Lykken, D.T. The detection of deception. <u>Psychological Bulletin</u>, 1979, <u>86</u>, 47-53.

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.S. Greenfield & R.A. Sternbach (Eds.) <u>Handbook of</u> 90 Polygraph 1981, 10(2) psychophysiology. New York: Holt, Rinehart & Winston, 1972, pp. 743-785.

Podlesny, J.A., & Raskin, D.C. Physiological measures and the detection of deception. Psychological Bulletin, 1977, 84, 782-799.

Podlesny, J.A., & Raskin, D.C. The effectiveness of techniques and physiological measures in the detection of deception. <u>Psychophysiology</u>, 1978, 15, 344-359.

Raskin, D.C., & Hare, R.D. Psychopathy and detection of deception in a prison population. Psychophysiology, 1978, 15, 126-136.

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

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

Spielberger, C.D., Gorsuch, A.L., & Lushene, R.E. <u>Manual for the</u> <u>State-Trait Anxiety Inventory</u>. Palo Alto: Consulting Psychologists Press, 1970.

Stanners, R.F., Coulter, M., Sweet, A.W., & Murphy, P. The pupillary response as an indicator of arousal. <u>Motivation and Emotion</u>, 1980, <u>3</u>, 319-340.

#### Reference Notes

1. Clark W.R. <u>A comparison of pupillary response</u>, <u>heart rate and GSR</u> <u>during deception</u>. Unpublished manuscript. Department of Psychology, Ball State University, Muncie, Indiana, 1975.

2. Lykken, D.T. Personal communication, September 24, 1980.

\* \* \* \* \* \*

Polygraph 1981, 10(2)

91

## STATISTICAL VERSUS CLINICAL LIE DETECTION

### Bу

Julian J. Szucko and Benjamin Kleinmuntz University of Illinois at Chicago Circle

## Abstract

The question of the polygraph's validity for detecting lying has broad policymaking implications and is clearly within the realm of psychological inquiry. Studies conducted to date have tended to ignore the single most important variable of polygraph validity: the human judge. This study focuses on clinicians' interpretations of polygraph protocols and shows that clinicians perform less accurately than statistical ana-Statistics outperformed human judges because they used lvses. information optimally and applied decision rules consistently. while clinicians tended to add error variance to their protocol interpretations. Unfortunately, current empirical evidence sugests that the prospects for improving clinicians' consistencies are not very promising: the authors therefore recommend the possibility of applying statistical methods to interpreting polygraph data. They conclude by suggesting that psychologists become more active researchers in this area of investigation, a research domain that is perperly within their scientific purview.

There is currently a lively debate in the lie detection literature between those who advocate the use of the polygraph examination as a valid technique to discriminate between truth and deception and the skeptics who do not. The first are well represented in a recent survey by Podlesny and Raskin (1977), who report that the false positive rate (<u>i.e.</u>, proportion of person incorrectly identified as lying) is within the range of 2% to 8% and who conclude that a number of physiological measures are accurate indices for discriminating between truth and deception. This position is seriously challenged in a reply to these authors by Lykken (1979), who asserts that false positive polygraph expectancies are within the unacceptable range of 36% to 39% and who argues that the theory of the lie test[1] is so naive and implausible that one should not accept the claims of very high accuracy that are sometimes made by the technique's advocates (see also the rejoinder by Raskin & Podlesny, 1979).

The authors express very special thanks to L. Rowell Huesmann and Herbert Stenson for their valuable comments on the statistical analyses in this article and additional thanks to Kenneth R. Hammond and David T. Lykken for their invaluable substantive and editorial contributions.

Requests for reprints should be sent to Benjamin Kleinmuntz, Department of Psychology, University of Illinois at Chicago Circle, Chicago, Illinois 60680.

<sup>©1981</sup> by the American Psychological Association. Reprinted by permission of the publisher and authors.

Empirical evidence has been brought to bear on both sides of the issue, sometimes support the technique's discriminability (Horvath & Reid, 1971; Hunter & Ash, 1973; Kubis, 1950; Slowik & Buckley, 1975; Wicklander & Hunter, 1975; see also Kubis, Note 1, on computer feasibility studies) and sometimes not supporting it (Bersh, 1969; Horvath, 1977; see also Lykken, 1981; Orne, Thackray & Paskewitz, 1972).

The question of whether polygraph interrogation is a valid tool for detecting deception is of interest to the broad scientific community and has policymaking consequences potentially more important than many efforts currently occupying psychologists. Its potential importance is due to its wide use in personnel screening and so-called periodic honesty checks in industry and its occasional acceptance as evidence in the courts. The technique is of interest to scientific psychology because it is clearly psychological; as Lykken (1974) once observed in an article in this Journal, "If persons who administer and evaluate Stanford-Binets or Rorschachs or MMPIs are psychometrists, then the polygrapher is a psychometrist also" (p. 725). In addition, Lykken (1974) has noted that the "polygraph test involves the study of autonomic ... responses to psychological stimuli" (p. 725) and therefore should concern those psychologists intested in psychophysiological responsivity.

But how shall we proceed in establishing the validity of this psychologically important technique? Basically two research strategies have been followed to date. One strategy has been to make the validity of the technique's physiological measures the central focus of study, and the other has been to attempt to evaluate, with varying degrees of statistical sophistication, the predictive powers of the examiners using the technique. Unfortunately, neither approach has addressed the single most important psychological variable of the lie detection equation - the inferential process of the clinical interpreter. After all, it is the clinician as information processor (or his or her actuarial surrogate) who must combine, analyze, and synthesize polygraph protocols prior to arriving at "deceptive" versus "nondeceptive" decisions. It is of central importance, therefore, that special attention be given to the decision-making activities of the judges who interpret polygraph protocols. This special attention has already been applied successfully to study human judgment in general (Einhorn & Hogarth, 1981; Einhorn, Kleinmuntz & Kleinmuntz, 1979) and clinical judgement (Kleinmuntz, 1975) in particular. Furthermore, in the tradition of contemporary human-information-processing research (see Newell & Simon, 1972), such attention has contributed important insights into the cognitions underlying problem solving.

In what follows, we attend to the inferences and clinical information processes of judges whose task it is to interpret the physiological polygraph data of theft and no-theft subjects. We do so within the framework of the statistical versus clinical controversy, a controversy that has continued to interest clinical psychologists and others since the appearance of Meehl's (1954) influential monograph on the topic almost 30 years ago. The fundamental question that Meehl posed regarding the relative accuracy of statistical and clinical prediction was essentially "Which is better?" This question was rephrased somewhat in the title of a later article by Meehl (1957) as follows: "When shall we use our heads instead of the formula?" The answer to both of these questions can be established only by empirical investigations, which are best conducted by means of a basic design proposed by Wiggins (1973): "The same input data are given

to a clerical worker (or a computer) and a clinician, and each is asked to make specific predictions concerning a socially relevant criterion" (p. 182). Our study uses such a design. Its outcome, as well as the outcomes of future similar studies, may well establish that for polygraph interpretation it is best to use the formula (or the head). The outcome of this and similar studies may also become part of Meehl's (1965) famous "box score" tabulation of studies comparing statistical and clinical data-combination methods (see also Goldberg, 1968; Kleinmuntz, 1967; Lindzey, 1965; Sawyer, 1966).

To establish the relative accuracy of clinical and statistical prediction methods, we used as inputs the physiological polygraph data of theft and no-theft experimental subjects. We then compared human protocol interpreters with statistics. We employed the Brunswikian lens model equation (Brunswik, 1952, 1956), which has been used extensively to investigate human judgment by Hammond (1955, 1966) and his associates (Hammond, Hursch, & Todd, 1964; Hammond, Stewart, Brehmer, & Steinmann, 1975) and which Petrinovich (1979) has described as an alternative psychological research paradigm, and a theory of signal detectability analysis (see Swets, Tanner, & Birdsall, 1961), which permitted us to examine still more closely and from another vantage point the clinical information processing of the protocol interpreters. Signal detection theory has been used pre-viously in similar judgment studies (see Lusted, 1968, pp. 98-140; Stenson, Kleinmuntz, & Scott, 1975) and was used in this study to gain insight into human judges' inferential styles. In the remainder of the article, we present a brief consideration of the polygraph examination, a description of the experiment, the results of the data analysis, and a discussion of the implications of our work for lie detection and human information processing as well as for scientific psychology.

### Polygraph Examination

A brief description of the polygraph procedure will be helpful here, since much of what follows asssumes such knowledge. The first phase of the examination is the pretest interview, which is conducted without the aid of the polygraph. This is a structured interview designed to obtain biographical data and to evaluate the subject's attitudes toward dishonesty, as well as to assess his or her attitudes toward the test itself. While the primary purpose here is to develop a data base from which to formulate questions for the polygraph phase, the responses and behaviors are also treated as interpretable data (Horvath, 1977).

The second phase is generally considered the polygraph test proper. It is only during this phase that the subject's physiological reactions are monitored. The usual field examination requires the continuous recording of three to four channels of physiological data. The variables measured typically include galvanic skin response (GSR), blood pressure, abdominal respiration, and thoracic respiration. The latter two measure the amount of external stomach and chest movement by means of a system of attached tubes and bellows. Blood pressure is continuously monitored through a similar system that uses a sphygmomanometer cuff, usually attached to the bicep (for a further description of this instrumentation, see Reid & Inbau, 1977).

In the standard polygraph phase, the examiner asks about 12 questions, each requiring a simple yes-no answer. These questions are formulated during the pretest interview and are reviewed with the examinee just prior to the polygraph monitoring, a review that is intended to clarify unnecessary ambiguities. As the questions are presented both verbal and physiological responses are recorded on the polygraph chart.

The above question sequence is usually repeated three or four times, with a stimulation test inserted between the first and second presentation. The stimulation or "card test" procedure is used to convince the subject of the infallibility of the polygraph. During this test, the subject is presented with several numbered cards and is instructed to select one. The examiner then tries to identify the chosen card, informing the subject that he or she will do so on the basis of the polugraph tracings. However, the examiner actually uses any number of deceptions (<u>e.g.</u>, memorizes the position of each card) to ensure that he or she correctly identifies the target card. At this point the subject may also be cautioned that he or she "can't beat the machine."

The specific questions the subject is required to answer vary depending on the reasons for the examination. Yet although the content may vary, most current examinations include three general types of questions. These have been designated as case-irrelevant, case-relevant, and control questions. Case-irrelevant questions deal with established biographical data (e.g., name, age, current address) and are designed to obtain a normal or baseline response level; case-relevant questions deal with the specific issues under investigation; and control questions attempt to force the subject to lie about some normatively shared transgression (e.g., "Did you ever steal anything in your life?"). The control questions permit the examiner to observe the subject's physiological correlates of mild emotional arousal.

In evaluating the polygraph charts, most interpreters use their intuition to look for signs of differential autonomic disturbance. If the disturbance associated with the relevant items seems to be greater or more persistent than that associated with the control questions, then the subject is judged to be deceptive. On the other hand, if the disturbance associated with the control questions appear to be greater, then it is assumed that the subject is being truthful. Most interpreters use a global evaluation method without specific measurement or scoring, although a method which requires that the autonomic differences between relevant and control items be assigned numerical values or scores has been introduced (Backster, 1963). These scores can then be used in evaluating the subject's responses. In the experiment to be reported below, we gave the judges only the physiological data as evidence, not the entire examination protocol, and we compared their intuitive, global, or clinical processing of these data with that achieved by statistical formulas and methods using the same data.

#### The Experiment

The present research required three distinct types of participants: subjects, examiners, and interpreters. The first were the individuals from whom the polygraph charts were obtained; the second were the polygraph examiners who administered the tests; and the third were the judges who evaluated the polygraph data and made decisions regarding the truth or falsity of the subjects' responses. The terms interpreter and judge are thus used interchangeably.

## Statistical versus Clinical Lie Detection

The polygraph charts were obtained from 30 psychology undergraduate subjects who volunteered for an advertised polygraph experiment. These undergraduates were randomly assigned to theft and no-theft conditions, so that 15 subjects served in each condition. Those in the former condition were directed to an office on campus, where they were to search through a desk containing a number of undesirable items plus a five-dollar bill, and were instructed to steal anything they desired. After their departure from the office, the experimenter searched the desk to determine what was stolen. All 15 theft condition subjects elected to steal the five-dollar bill.

The no-theft subjects were instructed to take a brief stroll around campus before returning to the experimenter's office. All subjects had previously been told that they would subsequently be administered a polygraph test. All 30 were taken, individually, to a downtown lie detector firm, where they were tested within an hour of fulfilling their theft or no-theft conditions. Subjects in both conditions, furthermore, were told that "intelligent and well-adjusted persons can 'pass' the test without being found guilty." The intent of this instruction, according to Gustafson and Orne (1963), is to involve the subjects in the experiment so that their motivational absorption in the outcome of their performance in the experiment resembles as closely as possible field conditions. The consequence of this instruction in our experiment was that one half of the subjects (i.e., the theft condition, or untruthful, subjects) would attempt to deceive the polygraph exzaminer and the other half (i.e., the no-theft condition, or truthful, subjects) would not. Upon completion of the experiment, all subjects were thoroughly debriefed by the experimenter.

The polygraph test was administered by four examiners-trainees at the polygraph firm. The examiners, who were not told that these were experimental subjects, asked the standard pretest, control, crime-relevant, and crime-irrelevant questions. The raw physiological data were then submitted to six experienced polygraph interpreters or judges, who were instructed to arrive at independent judgments of "truthful" or "untruthful." These judgments were to be made on an 8-point rating scale that was defined by "definitely truthful" at the low end of the continuum and "definitely untruthful" at the high end. The judges knew that half of the subjects had stolen something, but did not know what was stolen by whom or that the examinees were experimental subjects. The reason we informed the judges about the proportion of guilty subjects was to provide information that would allow them to adjust their intuitive selection ratios accordingly, although there is some evidence to suggest that such information does not influence decision making (Stenson et al., 1975).

### Data Analysis

The physiological polygraph data were converted from analog to digital form.[2] Then, using the digitized scores, we performed a Hotelling T-squared test to determine whether the physiological polygraph data contained the information that is necessary to differentiate between truthful and untruthful responses. The results of the T-squared analysis indicated that the combined physiological measures significantly (p < .01) discriminated between the truthful and untruthful groups, T-squared(10, 88) = 32.18. We also performed a discriminant function analysis to discover the optimal weights for each of the physiological indices or cues in the polygraph protocol. The outcome of this analysis disclosed that 80% of the

Polygraph 1981, 10(2)

96

protocols could be classified correctly into their truthful and untruthful categories. Moreover, the point-biserial correlations between the criterion of truthfulness-untruthfulness on the one hand and GSR (.23), thoracic respiration (.25), and abdominal respiration (.22) on the other were significant (p < .05). Thus, we established that there was sufficient information in the physiological polygraph data to differentiate between truthful and untruthful subjects. The question then becomes whether human polygraph interpreters can use the information in these data to arrive at equally accurate differentiating judgments, and if not, why not? To answer this question, we analyzed the polygraph data according to Tucker's (1964) modification of the lens model equation. The diagrammatic representation of the lens model question (see Hoffman, Slovic, & Rorer, 1968) is shown in Figure 1.

The Tucker modification of the lens equation is

$$\boldsymbol{r}_{\boldsymbol{a}} = \boldsymbol{G}\boldsymbol{R}_{\boldsymbol{e}}\boldsymbol{R}_{\boldsymbol{s}} + \boldsymbol{C}\sqrt{1-\boldsymbol{R}_{\boldsymbol{e}}^2}\sqrt{1-\boldsymbol{R}_{\boldsymbol{s}}^2}, \qquad (1)$$

where  $r_{e}$  is the validity of the judge's decisions, as measured by the correlates between judgments,  $Y_{s}$ , and the criterion,  $Y_{e}$ ; G is the correlation between judgments,  $Y_{s}$ . and the criterion  $Y_{e}$ , corrected for attenuation when  $R_{e}$  or  $R_{s}$  is less than 1.00 (G is sometimes also defined as the subject's knowledge);  $R_{e}$  is the criterion's predictability from the cues, as measured by the multiple correlation between the cues,  $X_{i}$ , and the criterion,  $Y_{e}$  (the term  $R_{e}$  is sometimes called task uncertainty);  $R_{s}$  is the judge's ability to control the application of his or her knowledge, as measured by the multiple correlation between the cues,  $X_{i}$ , and the judgments,  $Y_{s}$  (this term has also been called consistency or cognitive control; and C is the correlation between the variances in the environment and in the subject's judgment that are unaccounted for by the linear component, G.

Essentially, then, this analysis, which we performed by dichotomizing the 8-point rating scale so that 1-4 defined the "truthful" and 5-8 defined the "untruthful" categories, separates empirical validity, or  $r_{\bullet}$ , into cue validity, or  $R_{\bullet}$ ; knowledge, or G; and application of knowledge, or  $R_{\bullet}$ . By evaluating cue validity, it is possible to determine whether the physiological polygraph data are a valid deception index independent of the interpreter's ability to use this information.

Table I summarizes these results and shows that although the overall validity coefficients  $(r_{\mathbf{Q}})$  of Judges 2, 3, 4, and 5 differed significantly from .00, the best interpreter's validity, that of Judge 3, was .43. This table also shows that Judges 1 and 6 did not differentiate significantly between the protocols of theft and no-theft subjects and that, for this group of judges, greater amounts of experience, on the average, did not improve their validities. This result is reflected in the equal average values of  $r_{\mathbf{Q}}(.23)$  achieved by the high- and low-experience judges. Interestingly, the highest overall validity coefficient, or accuracy, was achieved by Judge 3, who had only three months of experience in interpreting polygraph data; the next to lowest validity was achieved by Judge 6, who had eight years of experience. One can also see in Table I that a simple linear combination of the cues  $(R_{\mathbf{Q}})$  outperformed all the judges without exception, again demonstrating that the essential information

Statistical versus Clinical Lie Detection



Figure 1. The lens model.

regarding discriminability between truthful and untruthful groups is contained in the physiological data.

One finding that reflects more favorably on the clinicians' interpretative skills than the preceding was that Judges 6, 3, and 5, in decreasing order of accuracy, were able to identify as important such specific polygraph cues as thoracic respiration, blood pressure, and abdominal respiration. These skills are reflected in the coefficients for G, presented in Table I, of .76, .58, and .41, respectively. These coefficients are indices of the judges' ability to identify valid relationships between some physiological measures and the criterion, but do not support some polygraphers' claims that deception can be discriminated from other emotion-laden behaviors on the basis of specific physiological cues. It is not possible in the present analysis, moreover, to know precisely which cues are the most differentiating.[3]

Interpreter/experience	۲.	R,	G	R,	с	Ad	β'
1/3 months	.02	.52**	17	.47•	.09	.55	02
2/3 months	.23°	.52**	.17	.44	.25*	.64°	10
3/3 months	.43**	.52**	.58**	.49	.37**	.75*	15
4/1 year	.27**	.52**	.18	.55**	.31**	.64*	.12
5/2 years	.33**	.52**	.41**	.51**	.30**	.68"	.00
6/8 years	.08	.52**	.76**	.37	08	.68*	11
Low experience M	.23	.52	.311	.46	.24		
High experience M	.23	.52	.45	.47	.23 <sup>1</sup>		

TABLE I Lens Model Equation and Signal Detection Theory Analyses of Polygraph Interpreter Judgments

Note. See text for definitions of r., R., and so on.

<sup>1</sup> Mean was calculated using absolute values, since sign only specifies direction of relationship. \*p < .05. \*p < .01.

The interpreters' relatively unimpressive performance vis-a-vis the statistical analysis conducted so far is partly explained by  $R_{e}$ , which sets the maximum limit for achievement because it represents the best possible linear combination between cues and criteria; that is,  $R_{e}$  sets the limit if one assumes that nonlinear combinations of cues will not add to predictive accuracy. Additionally, as the lens model equation indicates, when  $R_{s}$  (cognitive control) is less than its maximum value of 1.00, the

accuracy or validity  $(R_A)$  of the clinicians' judgments is lowered accordingly. As can be seen in Table I, the values of  $R_s$  range from .37 to .55. Therefore, these low values of cognitive control exert a considerable adverse influence on  $r_A$ , which is each judge's overall validity coefficient. On the other hand, if each judge were an "ideal" decision maker, in the sense of applying optimal weights consistently,  $R_s$  and G would be at their maximum values of 1.00. In point of fact, this "ideal decision rule" is described by our discriminant function analysis and by R, both of which involve correlations between the physiological measures and the truthful-untruthful criterion.[4]

Since the use of dichotomized ratings in the lens model equation resulted in a loss of 8-point rating scale data, a theory of signal detectability (Swets et al., 1961), or TSD, analysis was performed. This has the advantage of yielding indices similar to those of the lens model while preserving the total rating scale continuum. Particularly, TSD computes the area under the receiver operating characteristics (ROC) curve, or Ad, which is an index of a judge's discriminative ability between signal and noise judgments. Thus, Ad is like the index of knowledge, or G, of the lens model. In the TSD analysis, truthful responses were defined as voise trials and untruthful responses as signal-and-noise trials. Another TSD measure used was  $\boldsymbol{\beta}'$ , which reflects each judge's bias in calling protocols truthful or untruthful along the 8-point continuum of possibilities. That is, negative values of  $\beta'$  reflect a bias in favor of calling responses untruthful, whereas positive values reflect a bias in favor of calling responses truthful.[5]

The term  $\beta$  also indicates whether accuracy can be improved by shifting the decision criterion toward a untruthful or truthful end of the 8-point rating scale. It is therefore a useful statistic for future coaching of judges to improve their performance. In other words, by virtue of the fact that a TSD analysis allows one to learn each judge's decision strategy bias, one can teach the judge to correct for his or her decision bias by giving feedback about the direction of the bias. Such feedback would consist of a statement such as, "Don't classify a response as untruthful without more evidence than you have used in the past."

In Table I, values of .50 for Adrepresent no discriminability, and values of 1.00 reflect perfect discriminability. Our obtained values for Adrange from .55 to .75, which means that all judges except one could significantly ( < .05) discriminate between truthful and untruthful protocols. But the values for  $\mathcal{P}'$  reported in Table I indicated that this discriminability is achieved at a considerable personal cost to the hapless individual who may in fact be telling the truth but is misclassified as untruthful (i.e., the false positive or false alarm rate is high). The negative values of  $\mathcal{B}'$  in table I reflect this bias by showing, for example, that four of the six judges called many truthful protocols untruthful.

These interrelationships are presented somewhat more clearly in Table 2, which shows how different decision criterion placements affect decision accuracy. Table 2 compares these placements by presenting the theoretical cumulative hit rates and false alarm rates (HRs and FARs) of the interpreters for placements at each of the points on the rating scale. For example, in the row labeled 8, it is assumed that only those subjects who received a rating of 8 would be considered untruthful, while subjects who received a rating of 7 or below would be reported truthful. Likewise, in

the row labeled 2, it is assumed that subjects who received a rating of 2 or above would be labeled untruthful, while subjects who received a lower rating, in this case 1, would be reported truthful. Accordingly, each row in the table reports the HRs and FARs under the assumption that the examiner's cutoff point separating the untruthful and truthful categories would be located at that rating. Table 2 thus reveals a somewhat clearer picture than Table I of the overall accuracy of each interpreter's decisions.

TABLE 2

	Cu	mulati	ve Hit	Rates	(HRs)	) and 1	False A	larm	Rates	(FARs)		
	In	t. 1	Int	. 2	In	t. 3	Int	. 4	Int	t. 5	In	t. 6
Rating	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR	FAR	HR
8	.00	.00	.00	.02	.00	.00	.00	.02	.00	.07	.00	.02
7	.02	.11	.02	.20	.00	.07	.00	.13	.00	.20	.00	.27
6	.36	.44	.24	.40	.09	.24	.09	.38	.07	.40	.24	.60
5	.71	.73	.49	.71	.36	.78	.27	.53	.33	.67	.76	.82
4	.93	.98	.80	.87	.78	1.00	.78	.80	.80	.80	.89	.87
3	1.00	.98	.93	.96	.87	1.00	.89	.91	1.00	1.00	1.00	.98
2	1.00	1.00	1.00	.98	1.00	1.00	1.00	.98	1.00	1.00	1.00	1.00

Note. Int. = interpreter; HR = proportion of untruthful responses classified correctly; FAR = proportion of truthful responses misclassified as untruthful. Correct rejections, or truthful responses classified correctly, and misses, or untruthful responses classified as truthful, are not presented, since these can easily be calculated from the reported values (e.g., correct rejections = 1 - FAR; misses = 1 - HR).

Our results show that at the upper extremes of the scale, at points 7 and 9 (ratings that also reflect the interpreters' confidence about their deception judgments), the judges' classifications of deception are almost always correct (i.e., the FARs are negligible). However, at a rating of 5, which most likely corresponds to the judges' actual or empirical decision placements, the proportion of misclasified responses, or FARs, increases substantially. Interpreter 6, who correctly identified the greatest proportion of untruthful protocols, misclassified an almost equal proportion of truthful ones. Here the effects of response bias, or  $\boldsymbol{\beta}'$ , on the decisions of Judges 5 and 6 are apparent (see Table I): Their A values were identical, and yet Judge 6 made more errors than Judge 5. The latter judge's errors were specifically due to the fact that he was biased toward classifying truthful responses as untruthful.

#### Implications and Conclusions

What can be conclude about the accuracy of polygraph protocol interpretation and about the relative merits of statistical versus clinical prediction for this task? First, our lens model equation and TSD analyses indicate that polygraph interpreters can only poorly distinguish between truthful and untruthful responses and therefore at a cost to innocent persons who are misclassified. Moreover, our discriminant function and Hotelling T-squared analyses show that the information necessary for making discriminations between truthful and deceptive subjects is contained in the polygraph protocols but that judges do not use this information optimally. Furthermore, increased experience does not correspond to greater interpretive accuracy. This lack of accuracy of human information processors seems to be due to their inability to apply decision rules consistently rather than to a lack of knowledge about the correct cues.

Our findings are even more discouraging about the future of human information processing of polygraph data when we consider the reports of some researchers (Hammond, Summers, & Deane, 1973; Goldberg & Rorer, Note 2) who have shown that it is exceedingly difficult to improve human consistency in applying decision rules. And even more discouraging than this is the fact that if knowledge, or clinical wisdom, were to improve with experience and hence approximate perfect knowledge, this eventuality would still not overcome the adverse effects of inconsistency due to the fact that the upper limits on validity are imposed by the criterion's predictability from the cues, as measured by the task uncertainty, or R $\boldsymbol{e}$ . However, our TSD analysis suggests that some improvement can be obtained by decreasing the judges' biases. Unfortunately, the large overlap of the signal and noise distributions indicates that though judges can discriminate accurately to a limited degree, they are not very good at making such discriminations. This suggests that only minimal improvement is possible from an adjustment of the response bias.

Regarding the question of whether it is better to use the head rather than the formula, our results strongly suggest that human judges are illequipped to interpret polygraph protocols. Perhaps this is because they have limited memory stores, or perhaps it is because they are inconsistent rule appliers. In either case, it is apparent that human interpreters may best be replaced by statistical or other mechanical methods of data combining. What we are suggesting based on our findings, therefore, is that the formula is better than the head and that lie detector tests should be interpreted actuarially rather than intuitively.[6]

However, it is necessary to issue a caveat to those who wish to apply the formula in interpreting polygraph data. The present study was conducted under experimental rather than field conditions. Consequently, there is no way of knowing whether the same protocol interpretations that we observed would have been obtained under field conditions. Moreover, the judges of this study were given only physiological polygraph data to interpret; ordinarily they would integrate such data with clinical and Therefore, what needs to be empirically established other information. next is whether similar results would be obtained in field studies in which judges were given the opportunity to interpret complete protocols. Under these conditions subjects would be real participants in the examination rather than psychology undergraduate volunteers, and judges would be given access to more than just physiological data. The results of such field studies would be meaningful additions ot Meehl's "box score" tabulation comparing the relative accuracy of intuitive versus statistical processing of input data and would permit a more definitive answer than does the present research.

Finally, we hope that our foray into this domain will generate a flurry of studies using similar approaches. But if it does not, we trust that, at the very least, we have demonstrated that it is important for psychologists to become involved in the area of polygraph interrogation, which is now largely left - by default - to practitioners of the art and to researchers with a vested interest in the thriving polygraph industry. As we indicated at the outset, polygraph interrogation involves psychological testing, uses psychophysiological technology, and draws on knowledge that psychologists have helped to shape. As such, it is (or should be) within the purview of psychology and the scientific community at large and must be subjected to the rigorous criteria of scientific falsifiability.

### Footnotes

[1] Lykken (1974, p. 726) makes an important distinction between lie detection, which is the most commonly used form of polygraph interrogation, and the guilty knowledge test. The first of these is the method referred to throughout this article and is the technique that asks subjects one or more relevant questions ("Did you steal the money?") and then attempts to ascertain whether their answers to these questions are or are not deceptive; the second method is fully described by Lykken (1981) and varies from lie detection in a number of important ways.

[2] Abdominal and thoracic respiration were represented by maximum and minimum amplitudes as well as duration of cycle. Blood pressure was represented by the same amplitudes plus pulse rate. GSR, for which there is no regularly recurring cycle, was represented by a simple amplitude measure and was sampled at the rate of one measurement per second. All measurements were taken manually in millimeters for a 15-second interval that began at the start of each question, and all measurement data were obtained from the polygraph charts.

[3] A separate hierarchical analysis would need to be performed to know which physiological cues are most salient.

[4] The optimal discriminant weights for the ideal decision rule are -.25, -.18, -.06, -.14, -.22, .07, -.07, .19, .11, and -.04, respectively, for three thoracic respiration measures (maximum and minimum amplitude and duration), three abdominal respiration measures (maximum and minimum amplitude and duration), three blood pressure indices (maximum and minimum amplitude and duration), and the GSR amplitude.

[5] The terms  $A_d$  and  $\beta'$  are used as the nonparametric equivalents of the d' and  $\beta$  measures of the theory of signal detectability because the assumptions of normality of the ROC curve and its variance cannot be made.

[6] David Lykken (Note 3) has pointed out to us that "replacing a polygraph interpreter by a computer will not be much help to the poor innocents whose hearts pound harder when they are asked 'Did you kill Fred?' than when they are asked 'Did you ever do anything naughty when you were younger?'" This conviction of Lykken's is related to his distinction already cited in Footnote 1 - between lie detection and the guilty knowledge test. He believes that the "basic theory of the lie test is naive and often wrong (i.e., that truthful subjects often are more aroused by the relevant than by the control questions, just like the deceptive subjects) and therefore that lie tests should better not be interpreted at all."

#### References

Backster, C. "The Backster chart and reliability rating method. Law and Order, 1963, 11, 63-64.

Bersh, P.J. "A validation study of polygraph examiner judgements." Journal of Applied Psychology, 1969, <u>53</u>, 399-403. Brunswik, E. <u>Conceptual</u> framework of psychology. Chicago: University of Chicago Press, 1952.

Brunswik, E. <u>Perception and the representative design of experiments</u> (2nd ed.). Berkeley: University of California Press, 1956.

Einhorn, H.J. & Hogarth, R.M. "Behavioral decision theory: Processes of judgment and choice." <u>Annual Review of Psychology</u>, 1981, <u>32</u>, 53-88.

Einhorn, H.J., Kleinmuntz, D.N., & Kleinmuntz, B. "Linear regression and process-tracing models of judgment." <u>Psychological Review</u>, 1979, 86, 465-485.

Goldberg, L.R. "Seer over sign: The first 'good' example?" Journal of Experimental Research in Personality, 1968, 3, 168-171.

Gustafson, L.A., & Orne, M.T. "Effects of heightened motivation on the detection of deception." Journal of Applied Psychology, 1963, <u>47</u>, 408-411.

Hammond, K.R. "Probabilistic functioning and the clinical method." Psychological Review, 1955, 62, 255-262.

Hammond, K.R. "Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology." In K.R. Hammond (ed.), <u>The psychology of Egon Brunswik</u>. New York: Holt, Rinehart & Winston, 1966.

Hammond, K.R., Hursch, C.J., & Todd, F.J. "Analyzing the components of clinical inference." Psychological Review, 1964, 71, 438-456.

Hammond, K.R., Stewart, T., Brehmer, B., & Steinmann, D. "Social judgment theory." In M. Kaplan & S. Schwartz (eds.), <u>Human judgement and</u> decision processes. New York: Academic Press, 1975.

Hammond, K.R., Summers, D.A., & Deane, D.H. "Negative effects of outcome-feedback on multiple-cue probability learning." <u>Organizational</u> Behavior and Human Performance, 1973, 9, 30-34.

Hoffman, P.J., Slovic, P., & Rorer, L.G. "An analysis of variance model for the assessment of configural cue utilization in clinical judgment." Psychological Bulletin, 1968, 69, 338-349.

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

Horvath, F.S., & Reid, J.E. "The reliability of polgyraph examiner diagnosis of truth and deception." Journal of Criminal Law and Criminology, 1971, 62, 278-281.

Hunter, F.L., & Ash, P. "The accuracy and consistency of polygraph examiners' diagnoses." Journal of Police Science & Administration, 1973, 1, 370-375.

Kleinmuntz, B. "Sign and seer: Another example." Journal of Abnor-<u>mal Psychology</u>, 1967, <u>72</u>, 163-165. Polygraph 1981, 10(2) Kleinmuntz, B. "The computer as clinician." <u>American Psychologist</u>, 1975, 30, 379-387.

Kubis, J.F. "Experimental and statistical factors in the diagnosis of consciously suppressed affective experience." <u>Journal of Clinical</u> Psychology, 1950, 6, 12-16.

Lindzey, G. "Seer versus sign." <u>Journal of Experimental Research in</u> Personality, 1965, 1, 17-26.

Lusted, L.B. <u>Introduction to medical decision making</u>. Springfield, Illinois: Charles C. Thomas, 1968.

Lykken, D.T. "Psychology and the lie detector industry." <u>American</u> Psychologist, 1974, 29, 725-739.

Lykken, D.T. "The detection of deception." <u>Psychological</u> <u>Bulletin</u>, 1979, 86, 47-53.

Lykken, D.T. <u>A tremor in the blood</u>: <u>Uses and abuses of the lie de-</u> tector. New York: <u>McGraw-Hill</u>, 1981.

Meehl, P.E. <u>Clinical vs. statistical prediction: A theoretical</u> <u>analysis and a review of the evidence</u>. Minneapolis: University of Minnesota Press, 1954.

Meehl, P.E. "When shall we use our heads instead of the formula?" Journal of Counseling Psychology, 1957, 4, 268-273.

Meehl, P.E. "Seer over sign: The first sound example." Journal of Experimental Research in Personality, 1965, 1, 27-32.

Newell, A., & Simon, H.A. <u>Human problem solving</u>. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Orne, M.T., Thackray, R.T., & Paskewitz, D.A. "On the detection of deception." In N.S. Greenfield & R.A. Sternbach (eds.). <u>Handbook of</u> Psychophysiology. New York: Holt, Rinehart & Winston, 1972.

Petrinovich, L. "Probabilistic functionalism: A conception of research method." American Psychologist, 1979, 34, 373-390.

Podlesny, J.A., & Raskin, D.C. "Physiological measures and the detection of deception." Psychological Bulletin, 1977, 84, 782-799.

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

Reid, J.E., & Inbau, F.E. <u>Truth and deception</u>: <u>The polygraph (lie-detector) technique</u> (2nd ed.). <u>Baltimore</u>, <u>Maryland</u>: <u>Williams & Wilkins</u>, 1977.

Sawyer, J. "Measurement and prediction, clinical and statistical." Psychological Bulletin, 1966, 66, 178-200. Slowik, S.M., & Buckley, J.P. "Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings." Journal of Police Science & Administration, 1975, 3, 305-309.

Stenson, H., Kleinmuntz, B., & Scott, B. "Personality assessment as a signal detection task." Journal of Consulting & Clinical Psychology, 1975, 43, 794-799.

Swets, J.A., Tanner, W.P., Jr., & Birdsall, T.G. "Decision process in perspective." <u>Psychological Review</u>, 1961, <u>68</u>, 301-340.

Tucker, L.R. "A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch and by Hammond, Hursch, and Todd." Psychological Review, 1964, 71, 528-530.

Wicklander, D.E., & Hunter, F.L. "The influence of auxiliary sources of information in polygraph diagnoses." <u>Journal of Police Science & Ad-</u> ministration, 1975, 3, 405-409.

Wiggins, J.S. Personality and prediction: Principles of assessmentpersonality. Reading, Massachusetts: Addison-Wesley, 1973.

\* \* \* \* \* \*

He who conceals a useful truth is equally guilty with the propagator of an injurious falsehood. Augustine.

## By

### Lillian Lim Quon

#### Introduction

According to the <u>Bible</u>, deception was first practiced by Cain, who having killed his brother, Abel, told God his brother's whereabouts were unknown to him. God, of course, could detect the lie and Cain was cursed and a mark placed upon him.

Lacking omniscience, but still confronted by lying as a common place event, society has since made continuing efforts in the art and science of lie detection. Some of those early efforts may have had a scientific basis. For example, early methods to detect deception included having an accused lick a red hot iron or blade. If the suspect's tongue remained uninjured he had passed the test. A similar detection device involved having the suspect chew on dry rice and then spit the rice out. If the rice remained dry, the suspect was judged guilty. If the accused in either case believed in the test and in his own guilty, his fear of discovery could inhibit his salivation, making him more susceptible to being burned in the first instance and incapable of wetting the rice grains placed in his mouth in the second instance.[1]

Those methods, along with the rack and thumbscrew, are no longer popular and were never really reliable. In contrast, the polygraph technique is founded on scientific principles which are not crossing the "twilight zone" between the "experimental and demonstrable stages."[2] This article will discuss the admissibility of polygraph results in court, the use of the polygraph in criminal investigations and highlight some of the new research efforts underway on the polygraph technique.

#### The Courts

In 1923, the court in the germinal case of <u>Frye v. United States</u> (D.C. Cir. 1923) 293 F. 1013, ruled the results of a systolic blood pressure deception test were inadmissible. The results had been offered by the defendant in a murder case and indicated his truthfulness in denying knowledge of the charged crime. In rejecting the evidence, the court stated:

"We think the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made."

Lillian Lim Quon is Deputy Attorney General, State of California. The views expressed herein do not necessarily reflect the policy or position of the Office of the Attorney General of California. Reprinted by permission of the author and publisher from <u>Peace Officer Law Report</u> (California Department of Justice): April 1981.

Ironically, Mr. Frye was later exonerated after the arrest and confession of the actual murderer some three years later.[3]

Although 48 years have passed since the decision in <u>Frye</u>, the <u>Frye</u> rule requiring a scientific principle or discovery, for purposes of admissibility, to be sufficiently established to have gained general acceptance in the particular field in which it belongs is still applied to the field of modern polygraphy. The general acceptance rule has, in the California courts, precluded the admission of polygraph results into evidence absent a stipulation of all parties.

In 1957, the California Supreme Court in the case of <u>People v. Carter</u> (1957) 48 Cal.2d 737, stated: "Lie detector tests do not as yet have enough reliability to justify the admission of expert testimony based on their results." (Id. at p. 752, emphasis added.) Again in 1958, the California Supreme Court stated in the case of <u>People v. Jones</u> (1959) 52 Cal.2d 636, "The courts have consistently held that whether the test is a polygraph test or a sodium amytal or sodium pentothal test, the results are not such as to be admissible for or against the defendant because of a <u>lack of scientific certainty</u> about the results. (Id., at p. 653, emphasis added.) The holdings in <u>Carter</u> and <u>Jones</u> were then cited with favor by the California Supreme Court in 1974 in the case of <u>People v. Thornton</u> (1974) 11 Cal.3d. 738, 763-764.

In holding polygraph test results and the willingness or refusal to take a polygraph test are inadmissible, the California Supreme Court did not evaluate what progress, if any, has been made in the field of polygraphy. However in People v. Adams (1975) 53 Cal.App.3d 109, the Court of Appeals did acknowledge that unquestionably substantial progress had been made in improving the equipment and operator techniques used in administering the polygraph test. However, based solely on the facts of that case and the state of the law in California, the Court of Appeal concluded the results of the polygraph test administered to the defendant should not (Id., at p. 115.) have been admitted into evidence. In the court's opinion an inadequate foundation had been laid for the admission of polygraph results because the defendant took a unilateral test from a polygrapher of his own choosing. The court quoted from an article titled "The Friendly Polygrapher," authored by Martin T. Orne, stating:

"'Whereas the usual polygraph examination is carried out in a situation where the polygrapher is at arm's length -- in the employ of a law enforcement agency, a potential (or actual) employer or in some similar relationship, where his decision will inevitably have a direct effect on a suspect's future -- the context in which the friendly polygrapher carries out his test is inevitably different. In the latter case the suspect realizes that his attorney has employed the polygraph examiner to help in the preparation of his defense. For the innocent person this may matter very little; however, for the guilty individual it alters the situation considerably. The guilty individual when tested by a friendly polygrapher knows that the results of the test if he is found deceptive will not be used against him. The only kind of findings which his attorney will utilize are ones where his innocence is being corroborated by the polygraph. As a consequence, the client's fears about being detected are greatly reduced. As we have been able to show in the laboratory, and as is acknowledged

by all polygraph experts, a suspect's fear of detection is the major factor in assuring his augmented physiological response while lying. It is precisely this aspect of the situation which is most dramatically altered when the polygraph is employed by the defendant's attorney. The respect and perhaps even deference accorded to the client by the polygraph examiner will tend to convince the client that the polygraph is really attempting to help his cause and thereby make him less afraid and less detectible, even if he is guilty ... '" (Id., at pp. 116-117.)

The Court of Appeals discounted the defendant's offer to take a second test from an independent expert appointed by the court because foundational testimony in the Adams case indicated the more times a person is tested the less reliable are the test results. (Id., at p. 518.) The court also found the defendant's ingestion of liquor and valuum prior to the occurrence of the charged crime raised serious doubts as to whether the defendant could have sufficient mental recall to make meaningful responses to a polygraph test concerning his conduct when affected by alcohol and drugs. The court further found defendant's ingestion of a tranquilizer prior to his polygraph test diminished the reliability of the results and found defendant's status as a medical doctor would have increased his ability to "beat the test". (Id., at pp. 118-119.) For all those reasons, the Court of Appeals concluded:

The stare decisis doctrine requires the trial court and us to follow the holdings of the Supreme Court, absent such factual changes as to negate their applicability." (Id., at p. 119.)

In the case of In re Joaquin S. (1979) 88 Cal.App.3d 82, the Court of Appeals extended the prohibition on polygraph evidence in a trial to proceedings collateral to the guilt or innocence of the accused. In Joaquin S., a juvenile committed to the California Youth Authority attempted to introduce into evidence the results of two polygraph tests at a proceeding to vacate or modify his commitment under Welfare and Institutions Code section 778. The offer of proof was as follows: "Joaquin was questioned by a polygraph operator concerning his participation in the two robberies, the basis for his commitment to the CYA; the polygraph operator would testify Joaquin had answered truthfully in stating he had not committed the robberies." (Id., at p. 82.) The offer of proof additionally included qualifying the polygraph operator and supplying evidence as to the reliability of polygraph results. The Court of Appeal found the proffered evidence inadmissible, stating: "The polygraph is not a truth machine. Ιt does not supplant the finder of fact or breathe life into a defense already rejected by the trial court." (Id., at p. 85.) The Court of Appeal held that even under the relaxed standards of admissibility applicable to a postcommitment proceeding, proffered evidence must still be relevant and reliable, concluding lie detector evidence had yet to pass the test of relevant and reliability in the California courts. (Id., at p. 83.)

Because polygraph results are inadmissible, the California Supreme Court has held a defendant is not entitled to pretrial discovery of the results of a polygraph examination taken by the victim of the charged crimes. The Supreme Court found the material was not discoverable because the results would neither lead to any additional evidence nor aid the defendant in the preparation of trial. (<u>Ballard v. Superior Court</u> (1966) 64 Cal.2d 159, 170-179.)

The use of polygraph results in pretrial proceedings evaluating <u>probable cause</u> for a defendant's arrest remains an open question. In <u>People</u> v. Lara (1974) 12 Cal.3d 903, an investigating officer testified the material witness on probable cause had submitted to a lie detector test and having done so substantiated the officer's belief in the witness' credibility. The California Supreme Court in Lara observed that no authority had been cited for holding such collateral use of the polygraph test for investigative purposes to be improper. However the Supreme Court avoided reaching the issue because it found sufficient evidence independent of the polygraph test provided probably cause to arrest. (Id., at p. 909.)

Although the California courts have consistently refused to recognize the validity of polygraph results, those test results have been deemed properly admitted into evidence pursuant to the stipulation of the prosecutor and defense counsel. <u>People v. Houser</u> (1948) 84 Cal.App.2d 686, was the first California case which held the submission to a polygarph test by a qualified examiner and admission into evidence of the results were proper subjects for stipulation. The Court of Appeal in <u>Houser</u> suggested the evidence was admissible pursuant to stipulation simply because the defendant had thereby waived any claim of error on appeal. (<u>Id</u>., at p. 685; see People v. Reeder (1976) 65 Cal.App.3d 235, 239.)

Although parties may stipulate to the admission of polygraph evidence, the Court of Appeal in <u>People v. Cooper</u> (1979) 95 Cal.App. 3d 844, has held there is no abuse of prosecutorial discretion in the District Attorney refusing to so stipulate as a matter of policy. In <u>Cooper</u> the Court of Appeal made reference to the "serious problems of reliability inherent in the process of polygraphy", finding it could not criticize a refusal by the District Attorney to stipulate, as a matter of policy, to the admission of polygraph test results. (Id., at p. 850.)

Despite what the courts have characterized as reliability problems with the polygraph, as yet no California case has found defense counsel incompetent for stipulating to the admission of such evidence. In People v. Reeder, supra, 65 Cal.App.3d 255, the Court of Appeal found defense counsel was not incompetent for stipulating with the District Attorney, prior to trial, that the defendant and the victim each be administered a polygraph test by a licensed examiner, and that the details and results of the tests together with the opinions of the examiners be received in evi-The trial was a credibility contest between the victim and the dedence. fendant. The victim claimed the defendant had raped her and forced her to commit oral copulation upon him. The defendant admitted meeting the victim, but denied sexually attacking her. Other than the polygraph evidence, there was no corroboration of the victim's testimony. As it turned out, the defendant's polygraph examiner found the defendant to be untruthful and the victim's examiner found the victim to be truthful. Therefore despite the significance of the polygraph evidence in the trial and the resulting conviction based on the victim's testimony and the polygraph results, the Court of Appeal concluded the competence of counsel was not truly at issue because the defendant had authorized the stipulation and had suggested to his counsel that he submit to such a test. (Id., at pp. 238 - 240.)

A review of the California cases which have evaluated the admissibility of polygraph results makes it evident polygraph evidence will remain

inadmissible in a criminal case absent stipulation of the parties until an adequate and convincing foundation is laid on behalf of the polygraph test's validity and reliability. The general acceptance test set out in <u>Frye</u> and adopted by the California courts in conservative in nature and deliberately intended to interpose a substantial obstacle to the admission of evidence based upon new scientific principles. (<u>People v. Kelly</u> (1976) 17 Cal.3d 24, 31.)

The reluctance of the courts to accept polygraph evidence is understandable because polygraph results at least appear to impinge directly on the fact finding process of guilty or innocence. The polygraph is potentially too persuasive a means of advocacy and as a result may be given undue weight by the jury in a criminal case.[4] Additionally, the results of a polygraph examination are treated as suspect because of the opportunity for subjective evaluation by the polygraph operator in reaching his conclusions and because of speculation pathological liars and others who are able to distort their physiological responses can in some measure beat the test.[5] These doubts concerning the polygraph militate against the admission of polygraph evidence in a court of law. It appears those doubts will remain until disspelled by a substantial accumulation of scientific evidence demonstrating the reliability and validity of a polygraph examination.

## The Polygraph as an Investigative Tool

Although polygraph results, as previously discussed, may not be admitted into evidence absent a stipulation of the parties, the polygraph has been used with great frequency in the course of investigating cases for purposes of prosecution. Today, the polygraph is a much more sophisticated instrument than the systolic blood pressure deception test discussed in <u>Frye v</u>. <u>United States</u>, <u>supra</u>, 293 F. 1013. That test employed only one physiological measurement in assessing the truthfulness of the subject. The modern polygraph records, at a minimum, changes in breathing, heart beat, blood pressure and galvanic skin reflex. Some researchers indicate that up to 26 physiological parameters may be used in detecting deception.[6]

In theory, the act of lying causes a reaction in the nervous system which produces higher rates of pulse, blood pressure and breathing and changes the skin's electrodermal response. In a study sponsored by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, United States Department of Justice, the researchers concluded polygraph examinations utilizing the control question or guilty knowledge technique were approximately 90% accurate when properly conducted and evaluated.[7] Other studies also indicate a high degree of accuracy and consistency among "qualified" examiners.[8]

It follows as a matter of common sense, the additional information supplied by a polygraph examination taken by a fit subject and conducted by a qualified operator assists an investigator in evaluating the credibility of victims, suspects, informants and witnesses. Understanding the limitations of the test and recognizing the test as not infallible will aid the investigator in using the polygraph effectively as an investigatory tool.

All commentators reviewed by this author agreed the most important Polygraph 1981, 10(2)

factor involved in the use of polygraph is the ability, experience, training and integrity of the examiner. The studies indicating the accuracy of polygraph results appear to base their conclusions on the examinations administered by a limited test group of operators. However, in 1979 there were from 4,000 to 7,000 polygraphers practicing in the United States.[9] Those thousands of polygraphers operate within the benefit of a national standard for the education and licensing of polygraph examiners. Therefore prudence dictates an investigator should only rely on the accuracy of the polygraph where he can safely posit confidence in the qualifications and experience of the person who conducted the test.

A good examiner will recognize but not give undue weight to subjective factors, including behavioral cues, observe external stimuli and be on the alert for any number of factors which may impair the effectiveness of the test. Purposeful evasion of detection is one such factor. Subjects may attempt to evade detection by ingesting drugs, varying their breathing pattern, engaging in minute muscular movements, concentrating on topics other than the polygrapher's questions and even covertly doing damage to the polygraph instrument.[10] One commentator has suggested he would beat the lie detector as follows:

"After the first control question, I might suspend breathing for a few seconds, then inhale deeply and sigh. While the second control is being asked, I might bite my tongue, hard, breathing rapidly through my nose. During the third control question, I might press my right forearm against the arm of the chair or tighten the gluteus muscles on which I sit. A thumbtack in one's sock can be used covertly to produce a good reaction on the polygraph. So too can a thumbnail when dug under the nail of the forefinger."[11]

A good examiner, with the assistance of the investigating officers, will secure all available information concerning the charged crime and the accused. He can then conduct a more effective pretest interview and formulate adequate relevant, irrelevant and control questions to be used during the actual test. It is important that test questions be unambiguous and unequivocal to the subject. A polygrapher given sufficient background information can avoid ambiguity in his questions. Further, he can better formulate questions which would pose the most threat to the guilty subject and thus elicit the greatest physiological response. Therefore cooperation and the free flow of information between the examiner and the police are important factors in obtaining an effective polygraph examination.[12]

It has been suggested, as previously noted, that the results of a polygraph given by a "friendly polygraph" should be treated as unreliable. A friendly polygrapher is one who examines a suspect on a confidential basis for the defense attorney. The theory postulates that there is no physiological response to lying per se. Rather, the physiological changes recorded by the polygraph are dependent on the test subject's desire to avoid detection and his anxiety concerning the consequences of discovery. The theory goes on to suggest the test subject who has no fear of any consequences is an unfit test subject. The polygraph test given to a subject on a confidential basis provides no adverse consequences because the results of such a test would not be disclosed to law enforcement unless those results were favorable to the accused. It would therefore

follow the results of a polygraph test administered by a polygrapher on a confidential basis, irrespective of the polygrapher's qualifications, integrity and the adequacy of the background information available to him would be immediately suspect.[13] It would also seem to follow any subsequent test's reliability would be diminished even though administered by an independent examiner because a subject who has successfully avoided detection on one polygraph exam would be confident he could beat the polygraph and therefore less anxious.

A study which sought to test the theory concerning the "friendly polygrapher" has indicated there is no difference in the frequency of truthful outcomes of defense and law enforcement examinations conducted by the same examiner.[14] However, since the study apparently did not take into account the accuracy of the test reesults, and because the study was based on a limited test sample, it is still necessary that further research take place before the dangers of relying on a "friendly" polygraph examination be discounted.

The fitness and preparation of the subject for polygraph testing are almost as important as the qualifications of the polygraph operator. The polygraph technique appears to be more effective if the police do not tell the test subject the details of the charged crime.[15] Under those circumstances, the polygraph examiner can employ a peak of tension test, formulating questions which would engender a measurable response only from a subject who knows the details of the charged crime.[16]

Additionally, it is generally recommended the police only briefly interrogate or not interrogate the subject prior to the test. It is possible a truthful subject may become so distraught at interrogation that he may either give inconclusive reactions to the polygraph examination `or give false positive reactions on the test.[17] It has been reported that when polygraph examinations are given in order to detect the perpetrator of a crime, as opposed to the use of the polygraph for business and employment purpose, there is a somewhat higher risk of false positives. Apparently the risk of false positives increases where the test subject is well educated, without a prior criminal history, and is particularly concerned about his reputation in the community.[18] A truthful person's anger or resentment may also produce responses on the polygraph test which can be misinterpreted as false positives. It is therefore important this risk of error be diminished through minimizing any interrogation or other accusatory processes prior to the polygraph examination.[19] It has been recommended that little or no interrogation of the subject occur within four hours prior to the examination.

It has been suggested a subject will produce inconclusive test results if he is a person so unintelligent or uneducated as to believe, as a general principle, that the detection of deception through the polygraph is not possible.[20] Much like a self-fulfilling prophecy, a person's belief the polygraph does not work eliminates any fear on his part that his deceptions will be detected by the examination. Absent any fear or anxiety, the simple act of lying would not elicit the physiological response indicating deception.

The fitness of persons suffering from certain mental disorders, for purposes of an effective polygraph examination, is also still in controversy.[21] It would appear persons whose mental disorder is of the type which prevents them from appreciating the difference between a truth and a lie would not be fit subjects for testing. Similarly, persons whose sense of social responsibility and ability to relate to society are markedly diminished may be inadequate test subjects.

A polygraph diagnosis may also be impossible if the test subject has a physical condition or illness which distorts his physiological responses. Persons who are in custody prior to the polygraph examination should be monitored so as to avoid their ingestion of drugs.[22] It would also be helpful to know the medical history of the test subject and whatever medication he may have recently taken. Given that information, the polygraphers can then make an informed decision concerning the advisability of postponing the test until such time as the test subject is in good health and free of medication.

In addition to a qualified polygrapher and a fit subject, an appropriate physical setting for the administration of the polygraph examination promotes a more effective test. Therefore if the test is being administered at the police station or District Attorney's office or some other site not necessarily designed for the conduct of such tests, a suitable environment for the examination should be provided. The key to doing so is in eliminating factors which might distract the subject during the test. Thus, a private and quiet room should be available for the examination. The subject should not be able to see or hear activities which may be going on outside the room and the subject should not be in any physical discomfort.[23]

The polygraph is no substitute for careful and thorough police work. [24] That appears to be quite true. First, without adequate background information concerning the accused and the charged crime, a polygrapher cannot formulate satisfactory test questions. Second, a polygraph examination may produce inconclusive results where the test subject is purposely evading detection. Inconclusive results and errors of diagnosis may also result from any number of other factors, some of which have been discussed here, which affect the fitness of the subject for testing. A1though research studies indicate the polygraph is highly accurate, the test is not infallible since it is dependent on both the expertise of the operator and the fitness of the subject. The results of the polygraph test can be used to corroborate other information available to the police and assist in the effective allocation of law enforcement resources by suggesting fruitful avenues for investigation. As such, the polygraph is an invaluable aid to criminal investigations and the preparation of cases for trial.

## Some New Developments

The field of polygraphy is a fertile area for research validating the polygraph examination and research improving instrumentation and methodology. The use of the mass spectrograph, an ultra sound gas analyzer and a microwave monitor in measuring respiratory responses is being studied. New instruments are being developed which measure capillary and pupillary responses to emotional activity and monitor cardiological activity.[25] It has also been suggested other physiological parameters, including instantaneous pulse rate, brainwaves and electrical changes in the muscles be evaluated for purposes of detecting deception.[26] Research testing the validity of using behavior cues with the polygraph charts in reaching a diagnosis continues. Some studies suggest examiners should restrict their basis for decisions to the polygraph charts and recommend numerical scoring of polygraph charts.[27]

The polygraph silent answer test is being utilized by some polygraphers. In that test, the subject makes no verbal response to the test questions. The physiological response of the subject is recorded without distortion from the test subject who may purposely or inadvertently sigh, take in large breaths or employ other respiratory maneuvers under cover of his oral answer.[28]

Research into observable patterns of purposeful distortions continues.[29]

The previously discussed utility of the polygraph in law enforcement investigations merits continued research and improvements in instrumentation and methodology. As long as the "lie" is with us, a method to detect that lie should not be neglected.

## Footnotes

1. Lee, The Instrumental Detection of Deception (1953), pp. 3-26.

2. In Frye v. United States (D.C.Cir. 1923) 293 F. 1013, the Court in rejecting the results of a systolic blood pressure deception test stated: "Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." (Id., at p. 1014.)

3. Ansley, Legal Admissibility of the Polygraph (1975), p. 15.

4. Id., at p. 11.

5. United States v. Wilson (D.Md. 1973) 361 F.Supp. 510, 514; Abrams, "The Validity of the Polygraph With Schizophrenics," in 3 Polygraph, pp. 328-337; Reid and Inbau, <u>Truth</u> and <u>Deception</u>: <u>The Polygraph</u> ("Lie-Detector") Technique (1979), p. 215.

6. Slowik and Buckley, "Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure and GSR Recordings," 3 J. Police Sci. & Adm., (1979), p. 305; California District Attorneys Association, Arson Investigation, (1980) p. XX-1.

7. Raskin, Barland and Podlesny, <u>Validity and Reliability of Detec</u>tion of Deception, (1978), p. 23.

8. Slowik and Buckley, <u>op</u>. <u>cit</u>., <u>supra</u>, p. 309; Hunter and Ash, "The Accuracy and Consistency of Polygraph Examiners' Diagnosis, 1 J. <u>Police</u> <u>Sci. & Admin</u>. (1979), p. 370; Horvath and Reid, "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception," 62 <u>J. Crim. L.C. & P.S.</u>, 62 (1971), p. 281.

9. Lykken, A Tremor in the Blood (1981), p. 1.

10. Reid and Inbau, op. cit., supra, pp. 194-215.

11. Lykken, op. cit., supra, p. 239.

12. Reid and Inbau, <u>op</u>. <u>cit</u>., <u>supra</u>, pp. 417-419; California District Attorneys Association, Arson Investigation, <u>op</u>. <u>cit</u>., <u>supra</u>, p. XX 3.

13. Ansley, op. cit., supra, pp. 114-116.

14. Raskin, Barland and Podlesny, op. cit., supra, pp. 6, 24.

15. See fn. 12, supra.

16. Reid and Inbau, op. cit., supra, p. 59.

17. See fn. 12, supra.

18. Raskin, Barland and Podlesny, op. cit., supra, at pp. 22, 24.

19. California District Attorneys Association, Arson Investigation, <u>op. cit.</u>, <u>supra</u>, p. XX 3.

20. Reid and Inbau, op. cit., supra, p. 215.

21. Ibid.

22. See fn. 19.

23. Reid and Inbau, op. cit., supra, p. 6.

24. Reid and Inbau, op. cit., supra, p. 418.

25. Ansley, <u>op</u>. <u>cit</u>., <u>supra</u>, pp. 257-293; Reid and Inbau, <u>op</u>. <u>cit</u>., <u>supra</u>, p. 377.

26. Reid and Inbau, op. cit., supra, p. 385.

27. Raskin, Barland and Podlesny, op. cit., supra, pp. 23, 25.

28. Horvath and Reid, "The Polygraph Silent Answer Test" (1971) J. Crim. L.C. & P.S., pp. 285-293.

29. Ansley, op. cit., supra, pp. 199-209.

\* \* \* \* \* \*

Polygraph 1981, 10(2)

115