The Effects of Incentives on the Detection of Deception

Charles R. Honts, Ph.D. and Barbara Carlton, Ph.D.

August 1990

Department of Defense Polygraph Institute
Fort McClellan, Alabama 36205-5114
Telephone: 205-848-3803
FAX: 205-848-5332

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
May 7 1990

3. REPORT TYPE AND DATES COVERED
Final Report (From May 90 to Aug 90)

4. TITLE AND SUBTITLE
The Effects of Incentives on the Detection of Deception

5. FUNDING NUMBERS
DoDPI90-P-0001

6. AUTHOR(S)
Charles R. Honts &
Barbara L. Carlton

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Defense
Polygraph Institute Bldg 3195
Fort McClellan, AL 35205

8. PERFORMING ORGANIZATION REPORT NUMBER
DoDPI90-R-0003

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Defense
Polygraph Institute
Building 3195
Fort McClellan, AL 36205-5114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DoDPI90-P-0001
DoDPI90-R-0003

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Public release, distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A mock crime experiment was conducted to explore the effects of manipulating motivation to deceive on the physiological detection of deception using the control question test. Sixty subjects were assigned to one of four conditions in a 2 X 2 factorial design (two motivational states crossed with innocence or guilt). The motivation manipulation failed to produce any significant effects. The control question test performed reasonably well. The original examiners' outcomes with guilty subjects were 80% correct, 7% incorrect and 13% were inconclusive. With the innocent subjects the original examiners' outcomes were 50% correct, 17% incorrect and 33% were inconclusive. Electrodermal measures provided the greatest discriminability between innocent and guilty subjects followed by respiratory and cardiovascular measures. The results add to the already complex set of motivation results in the literature, and they were discussed within the context of Steller's systems theory, previous research, and research needed in the future.

14. SUBJECT TERMS
detection of deception,
motivation, control question test

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT

Report No. DoDPI90-R-0003

The Effects of Incentives on the Detection of Deception

Charles R. Honts Ph.D. and Barbara L. Carlton, Ph.D.

August 1990

Department of Defense Polygraph Institute
Fort McClellan, AL 36205

Director's Foreword

This report describes a study designed to determine if psychophysiological detection of deception (PDD) accuracy is affected by offering examinees an incentive to be diagnosed non-deceptive. This is one of several published studies designed to address the influence of motivation on PDD examination decision accuracy. The authors interpret the analysis results as indicating that the incentive used in the study did not influence PDD examination decision accuracy. Results of some studies support this conclusion while others do not. The authors are careful to suggest that the incentive used may not have been of sufficient strength to influence responding. Additional analyses are provided regarding the accuracy of PDD examination decisions made by the original examiner, a blind examiner, and an automated scoring system.

While the results of this study suggest that examinee motivation does not effect the accuracy of PDD examination decisions, this conclusion should be interpreted with caution. The relationship between incentives, motivation, and behavior has not been precisely defined, nor is it clearly understood. In addition, the relatively small number of observations per group (15) could have provided insufficient data to adequately test the hypothesis. The question posed is, however, of great interest and importance to the forensic psychophysiology discipline and should be further investigated.

Michael H Capps

Michael H. Capps
Director

Acknowledgments

The authors would like to acknowledge and thank David Marine Schuck for his assistance in running the subjects in this experiment. The authors would also like to thank Donald Dutton, Ronald Farris, Robert Fritzche, Robert Manners, and Don Weinstein for conducting the polygraph examinations, and Larry Broadwell for conducting the independent scoring of the polygraph charts. The authors would like to express their appreciation to John Kircher and David Raskin for providing the ARCHIVE software that supported our physiological data analysis. This project was funded by the Department of Defense Polygraph Institute as DoDPI90-P-0001. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

Abstract

HONTS, C. R. and CARLTON, B. L. The effects of incentives on the detection of deception, May 1990, Report No. DoDPI90-R-0003, Department of Defense Polygraph Institute, Ft. McClellan, AL 36205.--A mock crime experiment was conducted to explore the effects of manipulating motivation to deceive on the physiological detection of deception using the control question test. Sixty subjects were assigned to one of four conditions in a 2 X 2 factorial design (two motivational states crossed with innocence or guilt). The motivation manipulation failed to produce any significant effects. The control question test performed reasonably well. The original examiners' outcomes with guilty subjects were 80% correct, 7% incorrect and 13% were inconclusive. With the innocent subjects the original examiners' outcomes were 50% correct, 17% incorrect and 33% were inconclusive. Electrodermal measures provided the greatest discriminability between innocent and guilty subjects followed by respiratory and cardiovascular measures. The results add to the already complex set of motivation results in the literature, and they were discussed within the context of Steller's systems theory, previous research, and research needed in the future.

Key Words: detection of deception, motivation, control question test

Executive Summary

HONTS, C. R. and CARLTON, B. L. The effects of incentives on the detection of deception. May 1990, Report No. DoDPI 90-R-0003, Department of Defense Polygraph Institute, Fort McClellan, AL 36205.

Although detection of deception tests are widely applied in criminal justice and national security, little research has been devoted to understanding their underlying psychophysiological processes. Motivation associated with the examination outcome has been assumed by some researchers to be a prerequisite for successful detection of deception in the field. Other researchers have suggested that explicit incentives associated with the test outcome are necessary to establish an appropriate deceptive context in laboratory simulation experiments. However, the effects of incentives on the control question test have not been explored explicitly, and the question of the importance of incentives for the detection of deception remains unanswered.

We examined the effects of incentives in a mock crime experiment of the control question test. The subjects were 34 male and 26 female volunteers from the training companies of an Army facility. Their ages ranged from 17 to 34, median = 19. Their years of education ranged from 10 to 17, $\underline{M}$ = 12.3. Half of the subjects committed a mock theft of a pistol from a truck in a busy parking lot. Those programmed guilty subjects were cautioned not to be discovered while stealing the pistol, and they were told not to confess their involvement to anyone. Programmed innocent subjects were asked to retrieve a backpack from the same truck as a favor to the experimenter. All subjects were told to tell their examiners that they had volunteered to take part in a polygraph experiment but that they had nothing to do with the theft of a pistol.

Half of the subjects were told that if the examiner found them truthful on their examination they would get the afternoon off without duties. During this afternoon they would have the chance to watch first run movies and have access to soft drinks, candy, and coffee: items they normally would not have access to during training. If the examiner found them deceptive or the test was inconclusive, they would have to return to their normal basic training duties in the afternoon. Questionnaires given before this experiment indicated that trainees would prefer the afternoon off rather than a reward of $25. The examinations were conducted by experienced Federal polygraph examiners who used the control question test developed at the University of Utah. The charts were numerically evaluated by a discriminant classification program.

The motivation manipulation had no significant effects on either the decisions or the scores produced by any of the three evaluations. Overall, the independent evaluator decisions were 64% correct, 13% incorrect, and 23% were inconclusive. Excluding inconclusives, 85% of the independent evaluator's decisions were correct. The independent evaluator's false positive rate was 13% and the false negative rate was 22%.

The results of this study suggest that explicit incentives and/or strong motivations are not necessary to achieve a significant level of detection of deception accuracy. These results go against much of the current thinking in the detection of deception profession and in detection of deception research. They suggest a stronger role for non-affective information processing rather than affective processes as the underlying psychophysiological nature of the phenomenon. Additional research on this topic is needed to extend the theoretical implications of these results.

Table of Contents

Title Page i
Director's Foreword ii
Acknowledgments iii
Abstract iv
Executive Summary v
Introduction 1
Method 4
  Subjects 4
  Apparatus 5
  Examiners 5
  Procedure 5
  Polygraph Tests 7
  Numerical Scoring 8
  CAPS Decisions 8
  Feature Extraction 9
Results 10
  Examiner Decisions 10
  Numerical Scores 11
  CAPS Discriminant Scores 11
  Physiological Features 11
Discussion 13
References 17

The detection of deception has been one of the major applications of psychophysiology. However, despite roughly 30 years of scientific research, acerbic debate continues unabated about the value of detection of deception techniques in the real world (for two reviews with differing perspectives and conclusions see Iacono & Patrick, 1988 and Raskin, 1989). As Dawson has recently noted, "no other question in contemporary psychophysiology has generated a higher ratio of heated debate to illuminating light than has this one" (1990, p. 120). Perhaps one of the reasons there has been so little light, is the relative dearth of theoretically oriented research and discussion in the area. Most research has been devoted to applied questions as science has attempted to catch up with the burgeoning applications of detection of deception in the real world. Most research has also been atheoretical and non programmatic. One noticeable exception to this focus on application has been a concern with the effects of motivation to deceive[1].

Most of the research on the effects of motivation has focused on motivation's effects on the detection of information, and that literature has presented a rather puzzling picture. Two studies have indicated that increased motivation enhances detection of information (Eladd & Ben-Shakhar, 1989; Gustafson & Orne, 1963). However, a number of other studies have found no effect for increased motivation on the detection of information (Davidson, 1968; Horvath, 1979; Lieblich, Naftali, Shmueli, & Kugelmass, 1974). The effects of motivation on the detection of deception have received less study. Podlesny and Raskin (1977) suggest that motivation associated with the outcome of an examination is a critical part of simulating the deceptive context of the real world in the laboratory. Others have stated that for the control question test to work, not only must the subjects be motivated to deceive, but they must also experience actual fear of detection (Saxe, 1984). The authors of a recent report on the validity of national security screening tests suggested that their extremely poor results in detecting deception may have been due to the lack of incentives associated with the outcomes of their examinations (Barland, Honts, & Barger, 1989).

Some support for the notion that motivation is associated with the accuracy of the control question test has been provided in a recent meta analysis of 14 laboratory studies of the most commonly used test in the field, the control question test (Kircher,

---

[1]We have used the term motivation to refer to manipulations that are designed to make it desirable for the subjects to produce truthful outcomes in their detection of deception tasks. We have not used the term to specifically refer to the hypothetical construct, motivation. Whether the hypothetical construct, motivation, was actually manipulated in this or any other study cannot be directly determined.

Horowitz & Raskin, 1988). Kircher et al. found three significant predictors of increased accuracy of the control question test in those studies. First, they reported that stronger incentives were associated with higher accuracy of control question tests, $\underline{r}$ = .73. Second, they reported that the more closely the characteristics of the subject sample matched those of the criminal population the more accurate the outcomes of control question tests, r=.61. Finally, they reported that control question test were more accurate when scored by experienced evaluators using field techniques (they referred to this as Decision Policy), $\underline{r}$ = .67. Unfortunately, the independent relationship of incentives to the accuracy of test outcomes is somewhat difficult to interpret since incentives were also strongly correlated with the other two predictors, subject sample, $\underline{r}$ = .83, and decision policy, $\underline{r}$ = .55. Unfortunately it appears that researchers who used nonrepresentative subject samples and decision policies also tended to not use strong incentives. It could be that incentives are not related to accuracy, but that subject sample and decision policy are related to accuracy.

Bradley and Janisse (1981) have conducted the only study to date that has directly addressed the effects of motivation on the control question test. In a laboratory experiment Bradley and Janisse threatened half of their subjects with a "painful but not permanently damaging electric shock if judged guilty" (p. 309). The other half of the subjects received no motivation. Bradley and Janisse reported that this manipulation had no effects on detectability.

The effects of motivation are an important issue for understanding the phenomenon of detection of deception. Any theory of detection of deception will have to reconcile the differences between the studies of the effects of motivation on the detection of information and extend those results to detection of deception tests. Resolution of this issue is also critical for conducting additional laboratory studies of the detection of deception. If detection is not possible or is greatly reduced under conditions lacking incentives, then it is critical that laboratory studies include at least a minimal motivational setting to assure the possibility of detection. On the other hand, if detection of deception situations are so intrinsically motivating that explicit incentives are not important for detection, then researchers should consider dropping this costly factor from their research designs.

Steller (1987) has provided a systems theory as an attempt to explain the phenomenon of detection of information and the detection of deception. Steller's theory ascribes important roles to both information processing and affective information evaluation. Briefly, differential physiological responses can result through two mechanisms. First, after sensing and encoding the stimulus, the information is processed in terms of its significance to the individual. According to Stellar, it is at this stage that most detection of information takes place through the attentional processes described by Ben-Shakhar and his colleagues (Ben-Shakhar, 1977, Ben-Shakhar & Lieblich, 1982; Elaad & Ben-Shakhar, 1989).

However, at the same time information is being evaluated for its informational significance a parallel affective information evaluation channel is also at work. This affective information evaluation channel assesses the relationship of the presented stimulus to the individual's goals and it also assesses the ability of the individual to cope with possible negative outcomes. Items that further or retard the individual's progress toward their goals are likely to result in increased physiological responding. Similarly, items that are evaluated as posing a threat of injury to the individual will also result in increased physiological responding as the examinee chooses options for coping.

In real world cases it seems likely that the affective information evaluation channel will be very active if not predominant for detection of deception and also possibly for detection of information tests. Real world cases involve obviously powerful goals and potentially negative consequences that the individual is respectively pursuing or trying to avoid. Some support for this notion comes from Raskin (1979) who has reported that physiological responding by criminal suspects in control question tests takes the form of defensive responses, suggesting that the tests were perceived as aversive.

In the laboratory, incentives associated with the examination outcome would seems to be an important factor in determining the involvement of the affective information evaluation channel. In situations of low motivation, differentiation and differential physiological responding between stimuli would depend primarily on the information processing channel. Recognition of concealed information would be sufficient for differential responses even in the absence of a detection of deception context or any explicit motivational structure, and detection has been demonstrated in both of those condition (Day & Rourke, 1974; Elaad & Ben-Shakhar, 1989). As the stimulus situation becomes more ego involving, either through the enhancement of the deceptive context or as incentives are associated with test outcomes, the affective information evaluation channel should become more active. As the affective information processing channel becomes more active differential physiological reactivity should increase and detection accuracy should increase.

In the present study, we explored the effects of motivation on the control question test with a motivation manipulation that used an incentive structure that had both positive and some negative aspects. Our subjects were U. S. Army recruits undergoing basic training who had been assigned to a polygraph examiner training facility for a day. Subjects receiving the motivation manipulation were offered the afternoon off with no supervision and with nothing to do except watch movies and take it easy, if they could pass their polygraph examinations. An afternoon off was a most unusual event for these individuals and one they appeared to desire very much. Questionnaires given to basic trainees before this study indicated that most troops would prefer an afternoon off to a monetary reward of 25 dollars. If subjects in the motivation condition failed their examinations, they were told that they would be taken back to their units and would have to resume basic training that afternoon. This was a somewhat unexpected event since troops assigned to the training facility were usually relieved of basic training duties for the entire work day. thus, both positive and negative events were associated with the outcome of the examination, although the positive aspects of the manipulation seem to predominate. Non motivated troops were simply taken back to their basic training units after their examinations.

It was predicted that the motivation manipulation would enhance the differential reactivity expected in the control question test. From Steller's theory, we expected that non motivated subjects would be detected at better than chance rates since the acts of their mock crime would provide for an information processing differentiation of relevant and control stimuli. However, non motivated subjects should have little involvement of the affective information evaluation channel, since they were neither threatened by negative consequences nor offered positive rewards associated with the examination's outcome. We predicted that the activity of the affective information evaluation channel would be enhanced by the motivation manipulation, since subjects were offered a valued goal, the afternoon off, and they were also threatened with a negative event, having to return to their units early. We expected that the increased activity of the affective information evaluation channel would result in increased differential reactivity and detection accuracy.

Method

Subjects

The subjects were 26 females and 37 male volunteers from the basic training companies of a large U. S. Army post. One subject was disqualified after he confessed during the pretest to having committed the mock crime. Two male subjects were run as the result of a miscalculation and they were subsequently dropped form all analyses. The remaining 60 subjects ranged in age from 17 to 34 years of age. Their median age was 19. Years of education ranged from 10 to 17, M = 12.3. None of the subjects reported any significant health problems. The subjects reported that they had an average 6 hours of sleep the evening prior to their participation.

Apparatus

All of the examinations were conducted with Scientific Assessment Technologies Computer Assisted Polygraph Systems (CAPS) in small sparsely furnished rooms that had been specifically designed for polygraph testing. The CAPS consists of a four channel Lafayette Instruments field polygraph and an PC/XT clone computer with analog to digital conversion capability. Four channels of physiological data and an event marker were collected and digitized with this system. Measures of abdominal and thoracic respiration were taken from two pneumatic tubes placed around the subject's abdomen and upper chest. Skin resistance response was measured from dry stainless steel electrodes placed on the palmar surfaces of the index and middle fingers of the subject's right hand. Finally, a continuous measure of cardiovascular reactivity was obtained from a partially inflated cuff placed on the subject's upper left arm. This cuff was inflated to a minimum of 60 mm Hg.

Data from these four data channels and a time mark channel were digitized and stored on floppy disk by the CAPS. CAPS stored timing events indicating each question's onset, offset and answer. CAPS stored the physiological data as an interrupted time series of 20 s intervals following each question onset. CAPS averaged the cardiovascular waveforms so that potentially misleading fluctuations in pulse amplitude were hidden (see Geddes & Newberg, 1977, for a discussion of the potentially misleading nature of pulse amplitude in this signal). CAPS software also provided editing capability so that artifacts due to subject movement and examiner centering adjustments could be removed from the time series. Finally, CAPS software provided for a discriminant analysis classification (Kircher & Raskin, 1988; Raskin, Horowitz, & Kircher, 1989; Raskin, Kircher, Honts, & Horowitz, 1988).

Examiners

Six experienced field polygraph examiners conducted the examinations reported in this experiment. Five of the examiners were federal special agents who were also instructors at a federal polygraph training facility. The sixth examiner was a psychophysiologist who was also an experienced field polygraph examiner and instructor at the training facility.

Procedure

Subjects were run in groups of five or six per day. They received an introductory briefing where they were asked to read and sign two consent forms. One form requested consent for the polygraph examination and participation in the research project and the second form requested consent to be video-taped during the examination.

After consent was obtained the subjects were given a brief description of what they would be doing. They were told that a crime was to be committed sometime during the morning, and that they would all be given a polygraph examinations to determine their involvement in that crime. They were told that the crime would be the theft and hiding of a gun. It was explained to them that only the thief would know the exact kind of weapon, where it was taken from and where it was hidden. However, everyone would asked to maintain the same cover story with the polygraph examiner.

Subjects were told they would be taken to another building and separated so that the identity of the thief would be known only to the thief. The group was also told that they would be visited by a major who was acting as the case agent. They were told that the major would talk to each of the subjects individually, and that he would have them run an errand prior to taking them to the polygraph suite for the examination.

All subjects were instructed to use the following cover story. They arrived at the polygraph training facility after breakfast; were initially located in a office trailer; signed some consent forms; and they were then taken to another building where they were separated. They waited in separate rooms until a major came into the room and talked to them a while. They ran an errand for the major prior to taking the polygraph examination. Subjects were told that if they were asked about the nature of the errand they were to say that they picked up something for the major and that's all that would be necessary.

Subjects were informed that it was the job of the examiner to use his "charts" to determine their innocence or guilt. The last instruction given to the group of subjects was that their task was to appear innocent on the polygraph examination and at no time should they admit to the examiner that they were the thief. If they did admit to being the thief, the subjects were told that the polygraph examination would be stopped, that their mission would be aborted, and they would be returned to duty in their unit immediately. It was explained to the subjects that after their examinations the examiners would score the charts and tell them the results of their test. Subjects were told that the examiner would conclude one of three things. The examiner could conclude that they were being truthful and that they were not involved in the theft. The examiner could also conclude that they were being deceptive and that they were involved in the theft or the examiner could conclude that the test was inconclusive and he was not sure about the subjects' involvement in the theft. Subjects were told that their goal was for the examiner to tell them that they were believed to be truthful and not involved in the theft. Subjects were told that it should be very easy for the person who was not involved in the theft to appear truthful because they were in fact telling the truth. However, appearing truthful might be more difficult for the thief as they would be lying to the examiner.

At this point, subjects in the no motivation condition were told that after they completed their examination, they would be taken to another building where they would have lunch and then they would be returned to their unit. It was stressed that they should try as best they could to maintain their innocence as this was an important part of their mission. Subjects in the motivation group were told that if the examiner found them to be truthful they would have a "stand down day" for the rest of the afternoon. This meant that the rest of the afternoon was theirs to use as they chose. They would have access to candy and soft drinks, as well as the use of a VCR with a 35" screen on which they could view several recently released movies.

After the initial briefing, subjects were taken from the office trailer to another building where they were placed in individual rooms. They waited in these rooms until the case agent joined them. The case agent entered the room, greeted each of the subjects, and then proceeded to program the subjects according to the condition to which they had previously been assigned.

Subjects were randomly assigned to conditions on two factors. One factor was guilt. Programmed guilty subjects were instructed to leave the building and go into the parking lot, to enter a camouflaged truck, to take a gun from the cab of a truck, to hide the gun and the knapsack, to bring the gun back into the building, and to hide the gun and the knapsack in a small attic. Once they completed this mission, they returned to their small room.

Programmed innocent subjects were given similar instructions by the case agent except that instead of stealing a gun from the cab of the truck, programmed innocent subjects were simply asked to go to the same truck and get an empty knapsack from the cab as an errand for the major. Upon their return the case agent had all of the subjects repeat the cover story they were to maintain. The subjects also had to repeat the instructions about appearing innocent during the polygraph exam. The case agent then escorted the subject to the polygraph suite where the polygraph examination took place.

The second factor on which subjects were assigned was motivation. Subjects were randomly assigned on this factor by groups. On any given day all subjects would either be offered the afternoon off or not. This was done for convenience and to avoid possible problems of cross subject rivalry and contamination through discussion of the manipulations.

<u>Polygraph Tests</u>

Each subject was given a single issue forensic control question test using the technique developed at the University of Utah (Kircher & Raskin, 1988). A brief pretest interview was conducted, and it included a stimulation test (Raskin & Hare, 1978). The control question test contained three relevant, three control, and five unscored buffer questions. All questions were reviewed with the subjects prior to the examination. Subjects were unaware of the order of question presentation, and the questions series was presented three times. Question order was systematically varied with each presentation.

Numerical Scoring

The physiological data were subjected to two numerical scorings. The first scoring was by the original examiners at the end of each examination. The second scoring was by an independent evaluator who was only provided with the charts and enough information to determine which questions were relevant and which were control. All of the analyses used the numerical scoring rules taught at the Defense Polygraph Institute. In this system the evaluator compared the physiological response to a relevant question with the physiological response to an adjacent control question. If the physiological response to the control question was the larger, a positive score was assigned. If the response to the relevant question was the larger a negative score was assigned. Responses of equal magnitude to both relevant and control questions resulted in zero scores. Numerical scores ranged from -3 to +3 with the size of the score reflecting the magnitude of the perceived difference between the physiological responses. The Defense Polygraph Institute rules for criteria of reaction are too complex and lengthy to describe here in detail (for detail, see Weaver, 1980).

Decisions of truthful or deceptive were developed from the numerical scores using the standard federal polygraph decision rules taught at the Defense Polygraph Institute. The numerical scores for each of the three relevant questions were summed. If any of the three relevant questions received a total score of -3 or less, or if the grand total of the scores was -6 or less, the subject was reported as deceptive. To be reported as truthful the subject must have had a total score of +6 or more and the scores for each of the relevant questions must have been greater than zero. Any combination of scores other than those described above resulted in an inconclusive opinion.

CAPS Decisions

The CAPS software also performed an evaluation of the charts. The CAPS evaluation was the discriminant analysis classification procedure described in detail by Kircher and Raskin (1988). This procedure extracted three physiological features: amplitude of the skin resistance response; amplitude of the cardiovascular response; and the average abdominal and thoracic respiration length (Timm, 1982) for each relevant and control question. These features were converted to z scores within each feature. These z scores were then averaged so that a single relevant and control question score was developed for each of the three features. The average of the z scores for the relevant question was then subtracted from the average of the z scores for the control questions. Then the three resulting differences scores were multiplied by discriminant weights and were summed to give a single discriminant score. Negative discriminant scores indicated that relatively greater responses were given to relevant questions while positive discriminant scores indicated that relatively greater responses were given to control questions.

Finally, CAPS calculated probabilities of obtaining the discriminant score if the subject were either guilty or deceptive. These two probabilities were entered into Bayes Theorem and the final result was a posterior probability of truthfulness. The guilty and innocent populations used for the development of the discriminant classifications and the probability calculations were derived from both laboratory and field data (Kircher & Raskin, 1988; Raskin et al., 1988). Empirical analysis (Raskin et al., 1988) has suggested that the accuracy of decisions could be maximized, and the number of inconclusive results minimized, when decisions about truthfulness were made at posterior probabilities of truthfulness of .25 and .75. We employed that rule in this analysis, thus posterior probabilities of truthfulness of .25 or less were considered deceptive outcomes and posterior probabilities of truthfulness of .75 or more were considered truthful outcomes. Posterior probabilities of truthfulness between .25 and .75 were reported as inconclusive.

Feature Extraction

The following 12 physiological features were extracted from the digitized data for each relevant and control question: skin resistance amplitude, skin resistance half recovery time, skin resistance rise time, skin resistance length, cardiovascular response amplitude, cardiovascular response half recovery time, cardiovascular response rise time, cardiovascular response length, abdominal respiration length, and thoracic respiration length. This extraction was done with the Archive program developed by Kircher and Raskin (1990). During extraction the Archive program performed a within subjects transformation of the raw scores into percent range scores for each feature with the following formula:

$$\text{Percent Range} = \frac{(X - \text{minimum value})}{\text{Maximum Value - Minimum Value})*100}$$

Where, $\underline{X}$ = an individual score, and the minimum and maximum value are the extreme scores for relevant and control question values for that measure for that subject.

The percent range scores were averaged to produce a single value for control and relevant questions for each of the 12 features. The values for relevant questions were then subtracted from the values for the control questions. The resultant 12 difference scores were used as the data base for analysis.

Results

Examiner Decisions

Possible effects of motivation in the decisions of the original examiners, the independent evaluator, and the Computer Assisted Polygraph System (CAPS) were tested with Kruskal/Wallis oneway ANOVAs. None of those analyses indicated a significant effect for the motivation manipulation on decisions. Therefore, the decision matrix was collapsed across the motivation factor for additional analyses. The decisions for the three evaluations are shown in Table 1. The CAPS noticeably outperformed the human evaluators with the innocent subjects by classifying 80% of them correctly while the original examiner and the independent evaluator correctly identified the innocent in only 50% and 37% of the cases, respectively. With the guilty subjects, the original examiners and the independent evaluator outperformed the CAPS by correctly classifying 80% and 70% of the guilty subjects correctly, while the CAPS correctly classified only 50% of the guilty subjects.

Table 1
Decision for the Three Evaluations

	Detection efficiency	Decision		
		Truthful	Inconclusive	Deceptive
Original examiner	.63			
Innocent		15	10	5
Guilty		2	4	24
Blind evaluator	.62			
Innocent		11	16	3
Guilty		1	8	21
CAPS	.65			
Innocent		24	4	2
Guilty		4	11	15

Decisions were also evaluated by calculating detection efficiency coefficients (Kircher, Horowitz, & Raskin, 1988). The detection efficiency coefficient, a measure of discriminatory power, is a correlation between the dichotomous criterion and the trichotomous decision, and can vary between -1 and 1. Inconclusive outcomes reduce the value of the detection efficiency coefficient, but not to the extent that errors do. The detection efficiencies for the three evaluations are also shown in Table 1. There were no statistically significant differences between the three evaluations in the discriminatory power of their decisions.

Numerical Scores

The numerical scores of the original examiners and the independent evaluator were subjected to a repeated measures ANOVA. Evaluator (Original, Independent) was entered in this analysis as a within subjects factor, while Guilt (Innocent, Guilty) and Motivation (Motivated, Non Motivated) were entered into the analysis as between subjects factors. The main effect of Guilt was significant, F (1, 56) = 40.17, p <.001, indicating that Innocent subjects produced positive numerical scores, m = 9.15, SD = 9.99, while guilty subjects produced negative scores, M = -7.85, SD = 11.22. There was also a significant main effect of Evaluator, F (1, 56) = 5.39, p =.024, indicating that overall the independent evaluator gave more positive scores, M = 1.5, SD = 12.53, than did the original examiners, M = -0.2, SD = 14.65. The only other significant effect was a small, unexpected and difficult to interpret interaction of Motivation and Evaluator, F (1, 56) = 4.01, p = .05.

The total numerical scores of the Original Examiners and the Independent Evaluator were correlated with the Guilt/Innocence criterion. The resultant validity coefficients were significant, r = .63, p <.001, and r = .63, p <.001, respectively.

CAPS Discriminant Scores

The discriminant scores produced by the CAPS analysis were tested with a Guilt (Innocent, guilty) by Motivation (Motivated, Not Motivated) ANOVA. the only significant effect of that analysis was a main effect for Guilt, F (1,56) = 37.96, p <.001. That effect indicated that Innocent subjects produced positive discriminant scores, M = 1.34, SD = 1.36, while Guilty subjects produced negative discriminant scores, M = 0.65, SD = 1.60. The discriminant scores were correlated with the guilt innocence criterion and produced a significant validity coefficient, r = .63, p < .001.

Physiological Features

Possible motivation effects on the 12 physiological feature difference scores were explored with a repeated measures ANOVA. The 12 physiological feature difference scores were entered into this analysis as a within subjects factor, called Feature. Motivation (Motivated, Not Motivated) and Guilt (Guilty, Innocent) were entered as between subjects factors. Two subjects were dropped from this analysis because of missing data. The only between subjects effect that was significant was the main effect for Guilt, F (1, 54) = 21.45, p <.001. Prior to evaluating the between subjects effect that was significant was the main effect for Guilt, F (1, 54) = 21.45, p <.001. Prior to evaluating the within subjects effects, a Mauchly sphericity test was conduced. That test was significant, W = .00285, p < .001, indicating that the sphericity assumption of the repeated measures ANOVA had been violated. The degrees of freedom for the subsequent within subjects tests were adjusted using the Greenhouse-Geisser correction, Epsilon = .54.

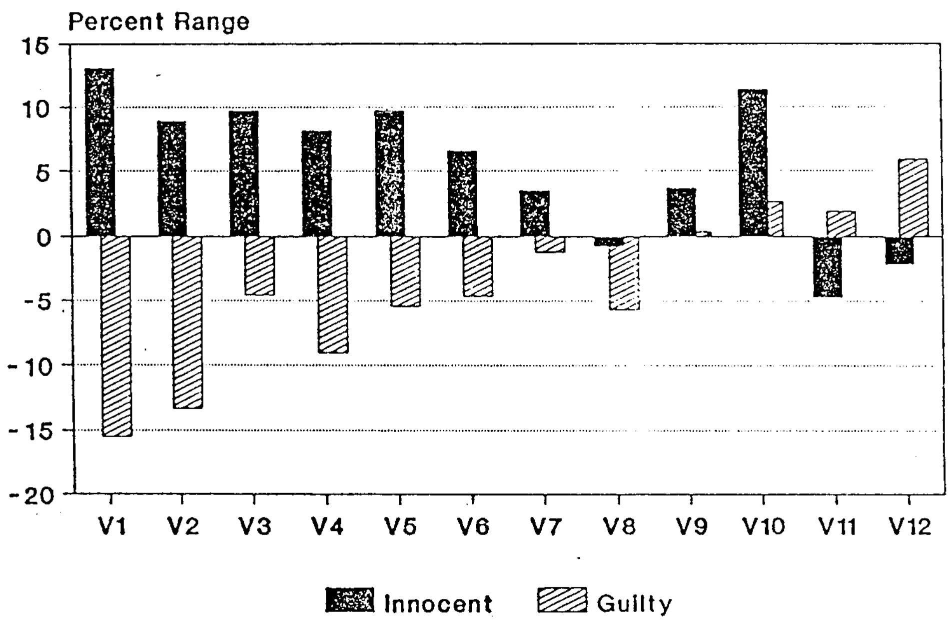

Figure 1. Mean percent range difference scores for the 12 physiological features. Ordering of features is arbitrary. (Legend, V1 = SR amplitude, V2 = SR Length, V3 = Abdominal Respiration Length, V4 = SR Rise Time, V5 = CR Amplitude, V6 = Thoracic Respiration Length, V7 = CR Rise Time, V8 = CR Length, V9 = CR Half Recovery Time, V10 = SR Half Recovery Time, V11 = Number of CR Responses, V12 = Number of SR Responses.)

The only significant within subjects effect was the interaction of Condition and Feature, F (6, 321) = 7.37, p <.01. This interaction is illustrated in Figure 1. Visual inspection of Figure 1 suggests that some features were powerful discriminators of guilt and innocence while other features were not.

The twelve physiological feature difference scores were then subjected to a series of oneway ANOVAs with Guilt (Guilty, Innocent) as the grouping factor. They wee also correlated with the Guilt Innocence criterion. The results of those analyses are shown in Table 2.

Table 2
Univariate F Ratios and Correlations with the Criterion for the 12 Physiological Feature Difference Scores

Feature	*F* (1, 56)	*r*
Skin resistance		
Amplitude	32.64***	.61***
.5 Recovery time	4.00*	.26*
Rise time	10.43**	.40**
Length	20.62***	.52***
# of responses	4.34*	.27*
Cardiovascular response		
Amplitude	10.02**	.39**
.5 Recovery time	.60	.10
Rise time	1.64	.17
Length	2.19	.19
# of responses	2.37	.20
Respiration		
Thoracic length	8.29**	.36**
Abdominal	11.28**	.41**

* p <.05
** p <.01
** p <.001

All of the Skin Resistance and Respiration variables were significant discriminators between Guilty and Innocent subjects, but only the amplitude of the cardiovascular response was significant.

Discussion

The results of this experiment provide no support for the predicted effects of the motivation manipulation. No differences were found between the motivated subjects and the not motivated subjects in either the test outcomes, the numerical scores, the discriminant scores, or in the objective measurements of the physiological data. Discrimination of the guilty from the innocent was significant for both motivation conditions. Overall 85% of the decisions of the original examiners were correct.

The computer based discriminant analysis classification system performed as well as the original examiners and the blind evaluator. However, the two types of decision makers showed different biases. The computer based system produced more false negative errors than false positive errors, while the human evaluators produced more false positive than false negative errors. This is an unexpected finding and deserves additional research. However, questions of decision bias aside, these results provide additional support for the use of statistical classifiers in the detection of deception. The discriminative power of the discriminant analysis classifier was as good as the original examiners and the blind evaluator. This is now a well replicated finding (Honts, 1986; Horowitz, 1989; Kircher & Raskin, 1988; Raskin et al., 1988, 1989). Given equivalent discriminate power, statistical decision makers are generally preferable to human evaluators because they are reliable, and they are unaffected by extraneous information and pressures that can mislead human evaluators (Meehl, 1954; Kircher & Raskin, 1988).

Significant discrimination between guilty and innocent was provided by 8 of the 12 physiological features tested. The skin resistance response amplitude measure was found to be the single most discriminating feature, as it has been in most studies of the detection of deception (i.e., Bradley & Janisse, 1981; Dawson, 1980; Honts, 1986; Kircher & Raskin, 1988; Raskin, et al., 1988). The respiration measures proved to be the next most powerful discriminators and the cardiovascular features proved to be the least powerful of the predictors tested. These results suggest that polygraph examiners may be giving too much weight to the respiratory and cardiovascular measures since their scoring systems weight all of the physiological measures equally.

With regard to motivation, the results of the present study agree with those of a number of studies of the psychophysiological detection of information that have failed to find significant effects of motivation (Davidson, 1968; Horvath, 1979; Lieblich et al., 1974). They also agree with Bradley and Janisse's (1981) finding that the threat of electric shock had no effects on the accuracy of the control question test. However, the results of this study contradict the findings of Elaad & Ben-Shakhar (1989) and Gustafson & Orne (1963) who have found significant effects of motivation. Elaad & Ben-Shakhar have suggested that realistic mock crimes may be intrinsically motivating. They suggest that this intrinsic motivation may be substantial enough to result in maximal detection in all conditions. Effects of motivation are then hidden by ceiling effects. This does not appear to be the case in this experiment. Even though detection rates in this study were relatively high, other studies have performed at noticeably higher levels of discrimination (Dawson, 1980; Kircher & Raskin, 1988; Raskin & Hare, 1978; Rovner, 1986). This suggests that ceiling levels of detection efficiency were not obtained in this study, and that there should have been room for the expression of motivation manipulation effects, if they had been powerful.

The results of this study failed to support our predictions about the effects of motivation based on Steller's (1987) systems model. It appears that the deceptive context of our mock crime was sufficient to enable a substantial amount of differential reactivity even without explicit incentives. It may be that motivation associated with the test outcome is a relatively trivial variable in the detection of deception and information. The very nature of the detection of deception tests may be intrinsically powerful enough to give sufficient signal value to the various stimuli to ensure differential reactivity. Motivation effects might only be expressed in relatively artificial detection of information settings that lack intrinsic interest to the subject. Once some threshold of motivation or personal interest is crossed additional increments in motivation may have no effects.

On the other hand, it may be that our manipulation was simply not strong enough to effectively engage the affective information evaluation channel and thereby increment differential reactivity beyond that level created by the mock crime. It is possible that our offer of the afternoon off was not really motivating to the subjects even though other similar subjects reported that they would prefer the afternoon to a cash reward of $25. Additional research with more improvished deceptive contexts and/or stronger motivations will be needed to explore the boundaries of the effects of motivation on detection of deception. In particular, additional research is needed to explore the effects of strong negative motivations. It may be that the information evaluation channel only becomes activated under strong threat conditions.

It is also possible that once the affective channel is activated, a qualitative change in physiological reactivity might result. This might explain the large difference in false positive rates between the Raskin and Hare (1978) study and its constructive replication by Patrick and Iacono (1989). These two studies addressed the possibility that Psychopathic prison inmates might be difficult to detect with the control questions test. Both studies produced similar results with guilty psychopaths, finding them relatively easy to detect. However, the false positive rate in the Patrick and Iacono study was many times that in the Raskin and Hare study. The primary difference between those two studies appears to be the nature of the motivational context. Raskin and Hare offered a monetary reward if subjects could pass the examinations. Patrick and Iacono offered approximately the same reward, but they also included a strong threat. Patrick and Iacono's subjects were told that if a certain number of the subjects failed the test then none of the subjects would receive the monetary reward. Further, the subjects were told that the results of the examinations would be made public so that the other prisoners would know who kept them from receiving their rewards. It may be that this type of serious threat triggered an effective process that instead of increasing differential reactivity in innocent individuals, reduced it. If qualitative changes occur at higher motivational levels, the generalizability of laboratory studies run at lower or moderate motivation levels may be seriously limited. These possibilities need additional research.

References

Barland, G H., Honts, C. R., & Barger, S. D. (1989). Studies of the accuracy of security screening polygraph examinations. Department of Defense Polygraph Institute, Fort McClellan, Alabama.

Ben-Shakhar, G. (1977). A further study of the dichotomization theory in detection of information. Psychophysiology, 14, 408-413.

Ben-Shakhar, G., Lieblich, I. (1982). The dichotomization theory for differential autonomic responsivity reconsidered. Psychophysiology, 19, 277-281.

Bradley, M. T., & Janisse, M. T. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. Psychophysiology, 18, 307-315.

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

Dawson, M. E., (1980). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. Psychophysiology, 17, 8-17.

Dawson, M. E., (1990). Where does the truth lie? A review of The polygraph test: Lies truth, and science. Psychophysiology, 27, 120-122.

Day, D. A., & Rourke, B. P. (1974). The role of attention in "lie detection". Canadian Journal of Behavioral Science, 6, 270-276.

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. Psychophysiology, 26, 442-451.

Geddes, L. A., & Newberg, D. C. (1977). Cuff pressure oscillations in the measurement of relative blood pressure. Psychophysiology, 14, 198-202.

Gustafson, L. A., & Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. Journal of Applied Psychology, 47, 408-411.

Honts, C. R. (1986). Countermeasures and the physiological detection of deception: A psychophysiological analysis. Dissertation Abstracts International, 47 1761B. (Order No. DA8616081)

Horowitz, S. W. (1989). The role of control questions in physiological detection of deception. Unpublished doctoral dissertation, University of Utah, Salt Lake City.

Horvath, F. (1979). Effect of different motivational instructions on detection of deception with the psychological stress evaluator and the galvanic skin response. Journal of Applied Psychology, 64, 323-330.

Iacono, W. G., & Patrick, C. J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), Clinical assessment of malingering and deception. New York: Guilford. (205-233).

Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-Analysis of mock crime studies of the control question polygraph technique. Law and Human Behavior, 12, 79-90.

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

Kircher, J. C., & Raskin, D. C. (1990). Archive Version 1.01. Salt Lake City, Ut: Scientific Assessment Technologies.

Lieblich, I., Naftali, G., Shmueli, J., & Kugelmass, S. (1974). Efficiency of GSR detection of information with repeated presentation of series of stimuli in two motivational states. Journal of Applied Psychology, 59, 113-115.

Meehl, R. 1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis: University of Minnesota Press.

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

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

Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), The orienting reflex in humans (pp. 587-605). Hillsdale, NJ: Erlbaum.

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), Psychological methods in criminal investigation and evidence. New York: Springer. (247-296).

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

Raskin, D. C., Horowitz, S. W., & Kircher, J. C. (1989). Computerized analysis of polygraph outcomes in criminal investigations. Report of research and results of Phase II of Contract No. TSS 86-18 to the United States Secret Service, Salt Lake City: University of Utah, Department of Psychology.

Raskin, D. C. Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). A study of the validity of polygraph examinations in criminal investigations. Final Report to the National Institute of Justice, Grant Number 85-IJ-CX-0400, Salt Lake City: University of Utah, Department of Psychology.

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

Saxe, L. (1984). Proposed use of polygraphs in the Department of Defense. Statement before the Committee on Armed Services, U. S. Senate. Unpublished manuscript, American Psychological Association, Washington, D. C.

Steller, M. (1987). Psychophysiologische aussagebeurteilung, Gottingen, FRG: Verlag fur Psychologie.

Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. Journal of Applied Psychology, 67, 391-400.

Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. Polygraph, 9, 94-108.Table 1. Decision for the three evaluations.