This is to certify that the

thesis entitled

THE EFFECT OF PLACEBOS AND FEEDBACK
ON THE DETECTION OF DECEPTION

presented by

Howard William Timm

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Social Science/
Interdisciplinary

Frank Horvath
Major professor

Date June, 1979

THE EFFECT OF PLACEBOS AND FEEDBACK

ON THE DETECTION OF DECEPTION

By

Howard William Timm

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

College of Social Science/Interdisciplinary

1979

ABSTRACT

THE EFFECT OF PLACEBOS AND FEEDBACK ON THE DETECTION OF DECEPTION

By

Howard William Timm

The purpose of this study was to examine the effect of placebos and feedback on the detection of deception. The subjects consisted of 270 volunteers enrolled in undergraduate Criminal Justice courses at Michigan State University, Fall (1977) term. Each of the subjects committed a mock contract murder after which the investigator administered a series of five lie detection tests in an attempt to ascertain the specific facts involved in the simulated murders. Subjects were awarded additional extra credit if they could successfully mislead the examiner on three out of the five tests.

Prior to the actual testing, 15 male and 15 female subjects were randomly assigned to each of the following groups: (1) placebo pass, feedback pass; (2) placebo pass, feedback fail; (3) placebo pass, feedback control; (4) placebo fail, feedback pass; (5) placebo fail, feedback fail; (6) placebo fail, feedback control; (7) placebo control, feedback pass; (8) placebo control, feedback fail; and (9) placebo control, feedback control.

Subjects assigned to the placebo pass and placebo fail subgroups were given a lactose placebo coupled with the suggestion that

the "medication" would either help or hinder them in their endeavor to mislead the examiner, depending on the group to which they were assigned. Similarly, subjects assigned to the feedback pass and feedback fail subgroups were given arbitrary feedback concerning a "demonstration" card test in which they were led to believe they either "beat" the test or were correctly detected, depending on their respective subgroup. The placebo control and the feedback control subgroup did not receive those respective treatments.

The dissemination of placebos and the supervision of the mock murders were performed by research assistants, who worked independently from the polygraph examiner. The research assistants were informed that they were dispensing active medication to the subjects, and the examiner had no knowledge prior to testing regarding the specific facts involved with the subjects' mock murders.

A standard field polygraph was used to record the subjects' respiration and skin resistance responses (SRR). After the testing, the subjects' SRR and respiration patterns were scored using various objective procedures. Biographical, performance expectancy, and attitudinal data were also collected from the subjects during the experimental sessions. The significance level for all statistical tests in the study was .05.

Generally, the placebo and feedback conditions did not have a significant effect on the detection efficiency of the polygraph. Female subjects, however, did exhibit significantly less electrodermal activity than the males during the polygraph testing.

Several other findings were noted during the course of the analysis. Contrary to the results reported in some detection of deception studies, respiration was found to be as valid an indicator of deception as galvanic skin response. It was concluded that the relatively high level of detection efficiency associated with respiration in this study may have resulted from the manner in which it was quantified. The levels of detection for the four physiological measures examined (respiration and three measures derived from SRR) were all significantly greater than chance levels. A low correlation ($r > |.1|$) between respiration and the SRR measures supported the rationale for recording multiple physiological indices during deception testing. It should be noted, however, that due to the nature of this study any inferences from it to field polygraph situations must be drawn with extreme caution.

DEDICATION

To my parents, teachers, and professors who encouraged and supported my research endeavors from my childhood years to the present.

ACKNOWLEDGMENTS

This study represents the culmination of a project I have longed to conduct since I was an undergraduate in college. I am deeply indebted to many individuals, without whom this project would still be nothing more than an elusive dream.

I appreciate the guidance and cooperation I received from the members of the University Committee on Research Involving Human Subjects at Michigan State University. I would be remiss if I did not mention Dr. Arthur Seagull, who selflessly donated several hours training the research assistants how to handle potential emotional problems related to this study. Dr. Thomas Adams graciously provided technical assistance concerning electrodermal physiology and measurement. Russell Carlson was of invaluable assistance in building the numerous mechanical devices used in this study. Dr. George Felkenes helped provide the necessary testing rooms, when space was at a premium. W. Frank Pont, Neal Schmitt, and Bill Brown graciously provided computer and statistical assistance. My close friend, Tom Austin, portrayed the priest in the slides used for the mock murders and provided both a place for me to stay and encouragement during much of the writing phase. The police officer, fireman, Army colonel, and hospital orderly who also volunteered to pose for the mock murder slides were extremely considerate and accommodating. The same is true of the professors who agreed to provide extra credit to their students who participated in this study.

Sue Cooley deserves special recognition for her expert typing and editing of the manuscript. Vera Kean also deserves recognition for her conscientious handling of the administrative and clerical aspects associated with the project's financial matters.

I am particularly indebted to the numerous research assistants and the hundreds of subjects who participated in the study. The research assistants performed their tasks flawlessly and were punctual, dedicated, and cheerful throughout the entire project. The subjects who participated also helped to make the long hours of testing a very enjoyable aspect of the study.

I will be forever indebted to the members of my dissertation committee. Frank Horvath, E. James Potchen, Ralph Turner, Carl Frost, Charles Press, and John Hudzik will always represent more to me than just my dissertation committee. They were also among the faculty members I respected and admired the most throughout my doctoral program. Each of them provided guidance and knowledge during my course of study and most generously took time from their busy schedules to help during all facets of this research.

Special thanks are reserved for Frank Horvath, who served as the committee chairperson. Dr. Horvath provided a wealth of information concerning detection of deception testing throughout the project. He also served as my mentor and graduate assistantship supervisor for three years prior to the study, vastly improving my research and teaching skills. If it were not for Dr. Horvath, I would not have achieved the degree of success I currently enjoy, nor would I have internalized many of his rigid standards that I will always strive for but probably never attain.

I wish to thank Vern Miller, Ed McGowan, and the Stoelting Company for providing the polygraph used in this study. I am also deeply indebted to the Office of Criminal Justice Education and Training within the U.S. Department of Justice's Law Enforcement Assistance Administration for providing the necessary funding to conduct this study.

Finally, I wish to thank my wife, Judee, who endured many sacrifices during the course of this project. If it were not for the love, assistance, and encouragement she provided, it is doubtful that the completion of this project would every have been realized.

The material in this project was prepared under a research grant under Grant Numbers 77-NI-99-0078 and 78-NI-AX-0028 from the Office of Criminal Justice Education and Training, Law Enforcement Assistance Administration, U.S. Department of Justice. Researchers engaging in such projects under government sponsorship are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Justice.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table | Page

Table | Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

Introduction

Lie detectors, as the public commonly refers to them, are instruments that measure various physiological responses. When using these instruments for deception analysis, the examiner is interested in monitoring the physiological changes that are either directly or indirectly affected by the subject's autonomic nervous system.

The "lie detection" examiner attempts to control the subject's external environment so he can examine the relationship between the statements the subject makes in response to stimulus questions, and his involuntary physiological responses. Most professional polygraph examiners (Reid & Inbau, 1966) believe that the physiological changes associated with deception stem from the subject's fear of detection.

The use of polygraphs and other similar devices is widespread in the United States. "In 1972 the American Polygraph Association estimated that between 200,000 and 300,000 polygraph tests would be given during that year alone" (Barefoot, 1974, p. 179). These tests are given for a wide variety of reasons, including such sensitive areas as industrial security, police corruption and brutality, criminal investigations, and national security. Depending on the circumstances, the examiner's findings usually play a key role in deciding whether an applicant is hired, an investigation is continued, or an employee is fired. Since such important decisions are based on the results of these tests, it is essential to determine if there are any systematic ways of "beating" the test. In addition to pointing out the possible limitations of deception testing, this information could prove invaluable in developing procedures that are less susceptible to such measures.

The major purpose of the present study was to examine the effects of placebos and feedback on the detection of deception. Field polygraphists have reported incidents in which guilty individuals successfully avoided detection (Klump, 1965; Reid & Inbau, 1966; Barland & Raskin, 1973). In some of the cases described, the only explanation for the success appeared to have been a placebo effect produced by such seemingly innocuous measures as putting soap under the arms or putting bullets under the cuff used to record cardiovascular activity. Despite reports that indicate the placebo effect might substantially reduce the accuracy of deception tests, this hypothesis had not been scientifically tested.

Another phenomenon that might also affect the accuracy of deception tests is feedback. One study that directly examined the relationship between feedback and detection of deception was conducted by Gustafson and Orne (1965). Those experimenters reported that the accuracy of deception testing was reduced when subjects who wanted to deceive the examiner were arbitrarily told that they had successfully done so on a previous test.

A discussion of additional research that is pertinent to the current study is found in the next section--the review of selected literature.

Review of Selected Literature

The literature review presented in this chapter is structured to inform the reader about the issues and studies perceived to be most closely associated with the present research. Since several excellent general reviews of the literature in this area already exist (Abrams, 1973; Barland & Raskin, 1973; Horvath, 1974; Orne, Thackray, & Paskewitz, 1972; Timm, 1975), only the research that had a major influence on the direction and format of this experiment is presented. First, the review focuses on the considerations common to all laboratory detection-of-deception experiments. After laying this foundation, the review's orientation is changed to those studies that are most pertinent to the effects of placebos and feedback on detectability. Despite the limited research in the area, these studies illustrate the conceptual foundation from which the hypotheses to be tested in this experiment were derived.

Detection-of-Deception Experimental Designs

To examine the accuracy of detection of deception in an experimental context or the multitude of factors affecting its accuracy, it is necessary to design a situation in which some or all of the subjects will attempt to deceive the examiner. In some of these situations, the researcher attempts to differentiate between "innocent" and "guilty" subjects or to determine the nature of the subject's involvement (i.e., innocent, lookout, or perpetrator). In other detection-of-deception experiments, the researcher knows in advance that all of the subjects will attempt to deceive him/her, but is interested in differentiating between the truthful and nontruthful statements made by each of these subjects. The former type of situation has been called the "guilty-person paradigm," whereas the latter has been referred to as a "guilty-information paradigm" (Gustafson & Orne, 1964).

Generally, the experimental designs structured to create the types of paradigms mentioned above can be classified in one of three different categories. The first category is most commonly referred to as a "card test" design. Usually the card test employs a guilty-information paradigm in which the subject is asked to select a card from a small deck (Alpert, Kurtzberg, & Friedhoff, 1963; Block, 1957; Block, Rouke, Salpeter, Tobach, Kubis, & Welch, 1952; Burtt, 1921; Geldreich, 1941; Horvath, 1978; Kubis, 1962; Kugelmass, 1967; Landis & Wiley, 1926; Langfeld, 1921; Obermann, 1939; Van Buskirk & Marcuse, 1954; Violante & Ross, 1964). The subject is then instructed to respond "no" each time the examiner asks if the subject selected a certain card, regardless of whether or not it was the card actually drawn. Obviously, by the time the researcher has asked about all of the cards contained in the original deck, the subject will have been forced to lie once during the test.

The card test design can also be transformed into a guilty-person paradigm by including blank cards in the deck. Gustafson and

Orne (1964) examined the effect of differential subject perceptions emanating from these two paradigms in a card test detection-of-deception experiment. They reported that in their study the guilty-information paradigm was significantly less effective in detecting deception than the guilty-person paradigm.

The second most frequently used experimental model is one in which some or all of the subjects either observe or participate in a mock crime and then are given a detection-of-deception test regarding that scenario (Orne et al., 1972). This model can also involve either the guilty-information or guilty-person paradigm. Under the guilty-person paradigm, the researcher generally attempts to differentiate between subjects who have committed the mock crime and those who have not (Berrien, 1942; Berrien & Huntington, 1943; Chappell, 1929; Landis & Wiley, 1926; Marston, 1917; Obermann, 1939; Podlesny & Raskin, 1978; Raskin & Hare, 1978; Runkel, 1936). However, in some cases the examiner is also interested in determining the subject's degree of participation (i.e., participant, observer, lookout, those who had planned or attempted the crime, or those totally innocent [Baesen, Chung, & Yang, 1949; Davidson, 1969; Kubis, 1962, 1973]).

When the guilty-information paradigm is used in a mock crime context, the examiner generally knows in advance that all subjects were required to commit one of several different mock crimes; however, his/her task is to determine which particular crime the subject is guilty of committing. Burtt's (1939) study clearly illustrates the mock crime guilty-information paradigm. In that study, all subjects were required to open and examine the contents of one of two boxes containing miscellaneous objects. Each subject was guilty of peeking into one of the boxes, and the experimenter's task was to determine which one the subject had opened.

The third category of experimental detection-of-deception designs comprises studies in which some or all of the subjects are required to lie about certain information they possess that was not acquired through participating in a card test or mock crime. Lykken's (1960) study provides an excellent example of this type of design. In that study, subjects were motivated to deceive the examiner about personal information (i.e., father's name, name of their former high school, etc.). The examiner's role was to determine which of five personal histories belonged to the subject being tested. Hence, the study represented a guilty-information paradigm. This design can also be modified to form a guilty-person paradigm by including subjects whose histories have not been given to the examiner.

Regardless of which type of laboratory design the researcher employs, the experimental situation is always structured so the experimenter will be able to discover ground truth at the conclusion of the study. The researcher also has complete control over the number and types of possible alternatives the subject could be "guilty" of committing. For example, in a guilty-information card test experiment, the researcher knows that the subject had to select one of the cards from the deck. Generally, the researcher also knows in advance the numbers on all of these cards. Thus the researcher can structure the questions to pertain only to the possible alternatives the subject may have selected. The same is true in mock crime situations. The researcher usually knows in advance all of the possible alternative elements of the crime with which the subject could have been involved and can structure the questioning procedure accordingly.

Field and Experimental Questioning Procedures

Since the researcher has absolute control over the number and types of possible alternatives the subject could be guilty of committing, it is not surprising that the types of questioning procedures used in laboratory studies differ to some extent from those generally used by field polygraph examiners. Rarely do field examiners have such clear-cut alternatives or know in advance the precise range of the suspect's possible involvement in the criminal act they are investigating.

The three most common types of questioning techniques used by field polygraph examiners are relevant-irrelevant, control question, and peak-of-tension tests. In using the relevant-irrelevant technique, the examiner asks a series of questions, some of which pertain to the matter under investigation (relevant) and others that do not (irrelevant). The first two questions are usually irrelevant (e.g., "Is today Friday?"), followed by a "Do you know who . . ." rather than a "Did you . . ." question (Barland & Raskin, 1973). After the third question is asked, any other relevant questions can be asked, with irrelevant questions inserted whenever the examiner wants the response to return to the basal level or after a fixed number of relevant questions has been asked (Harrelson, 1964; USAMPS, 1970).

The second major questioning procedure used in the field is the control-question technique. This technique differs from the relevant-irrelevant format in that the order in which the questions are presented is predetermined and control questions are incorporated into the series. A control question is one designed to capture the psychological set of the innocent subject (Barland & Raskin, 1973). Suspects are led to believe that the control questions are important to the resolution of the matter under investigation. These questions are also formulated in a fashion that would make it difficult for anyone to answer confidently and completely truthfully. For example, in a theft case a possible control question might be: "Other than what you mentioned [during the pretest interview], did you ever steal anything while you were in high school?" It is believed that innocent individuals will react more strongly to the control than they do to the relevant questions, whereas the opposite would hold for guilty suspects. Naturally, formulating the control questions and establishing the psychological set that makes these questions more threatening than relevant ones to an innocent suspect is a skill that the examiner must develop (Raskin, 1978).

The peak-of-tension test is the third major technique used by polygraph examiners in the field. This test usually comprises between five and seven questions that are mutually exclusive and worded similarly (Harelson, 1964; Reid & Inbau, 1966). The critical question, the one actually corresponding to the facts known about the crime, is placed approximately in the middle of the series. For example, if an individual took four dollars during an armed robbery, an appropriate peak-of-tension test sequence might be:

1. Regarding the amount of money taken, do you know if it was one dollar?
2. Do you know if it was two dollars?
3. Do you know if it was three dollars?
4. Do you know if it was four dollars?
5. Do you know if it was five dollars?
6. Do you know if it was six dollars?
7. Do you know if it was seven dollars?

To increase the guilty person's apprehension about the critical question, the suspect is usually either told in advance the order of the questions (Barland & Raskin, 1973) or the same question series is repeated so the suspect knows the questions and their order on the subsequent tests (Reid & Inbau, 1977). It is believed that a guilty person's physiological responses to the questions will peak at the critical question, and then return to normal after it has passed. Obviously, the use of this technique is limited to situations in which, of those to be tested, the guilty person, and only the guilty person, knows the correct response. The examiner also has to use extreme care in formulating these questions, so that the alternatives are equally plausible and of approximately equal emotional value to innocent suspects.

The conditions that make it feasible to use the peak-of-tension test are those most closely resembling the conditions present in laboratory detection-of-deception situations. However, it would be fairly rare for an examiner to possess information he/she is certain that the perpetrator knows, but no innocent people who are tested would have found or figured out. The examiner must also be wary that the victim did not intentionally or unintentionally provide false information.

One of the more common questioning procedures employed in laboratory detection-of-deception situations is called the guilty-knowledge technique. This method was developed by Lykken (1959) for use in a mock-crime lie detection experiment. The guilty-knowledge method of deception analysis assumes that a guilty person knows certain facts pertaining to his crime that an innocent person does not know, and that the guilty individual's physiological responses during testing will differentiate between the relevant and irrelevant stimuli (questions) presented by the examiner. The technique is very similar to the peak-of-tension test. However, it differs in that the placement of the critical question is generally random and the subjects usually do not know the questions or their order in advance. The fact that the researcher normally does not know in advance which of the questions was the critical item also helps insure that the experimenter does not in some way bias the subject's response to it.

Although both the relevant-irrelevant and control-question techniques can be and have been used in laboratory studies (Barland, 1972; Barland & Raskin, 1973; Orne et al., 1972), their use is generally more appropriate in guilty-person-paradigm situations, whereas the guilty-knowledge and peak-of-tension methods are appropriate for either the guilty-person or guilty-information designs.

Other Important Methodological Considerations Pertaining to Detection-of-Deception Experiments

Thus far, different types of experimental designs and questioning procedures pertaining to the detection of deception have been discussed. Although the aforementioned classification systems provide a basis from which the reader can differentiate and interpret experimental detection-of-deception studies, it does so on only two dimensions. Other factors are also extremely important in considering the nature of the study. Using Easton's (1965) model, one can break down these factors into those associated with inputs, process, and outputs.

The input factors in laboratory detection-of-deception experiments encompass a wide array of variables. The following examples are intended to illustrate the diversity and vastness of these factors. However, it should be noted that this list is not meant to be comprehensive; rather, it is provided to stimulate the reader's thoughts about and awareness of these issues.

1. The setting and layout of the experimental station
2. The physical appearance of the experimenter(s)
3. The intentional and unintentional verbal and nonverbal cues emanating from the experimenter(s)
4. The length of time for the subject, from the prospect of involvement with the experiment through its completion
5. Comments from former participants, prospective participants, and nonparticipants regarding the experiment

6. The degree and nature of the subject's involvement in the experiment
7. Explicit statements to motivate the subject
8. The nature of the treatment variables
9. The questioning procedure used by the examiner

Unfortunately, the variables affecting how the input variables are processed are far more complex and are those that the experimenter cannot easily manipulate unless specific measures are taken to control for them. Included among the variables that might affect how the subject perceives and processes input variables are prior experiences the subject associates with the context of the experiment, as well as genetic, personality, and cultural differences. Therefore, it is highly unlikely that any of the subjects will perceive their role and status in the experiment exactly the same, or that they will react identically to the stimuli presented.

The output variables also add to the complexity of understanding differences among laboratory detection-of-deception studies. Frequently, different physiological indices are measured, using different equipment or different procedures. Even when the same physiological responses are monitored, they are not always measured and/or scored consistently from study to study.

These comments concerning the differences between laboratory detection-of-deception studies were not intended to make the reader believe it is futile to attempt to comprehend the relationship between different studies, nor were they necessarily intended to plead for more consistency among future studies. Rather, their purpose was to sensitize the reader to these differences and to convey the message that our current knowledge regarding the factors affecting the detection of deception is analogous to a massive jigsaw puzzle in which a vast majority of the pieces are missing. However, on a more optimistic note, each new study in this area unveils another piece of the puzzle, which may bring us one step closer to seeing the total picture.

Studies Having a Major Influence on the Issues Examined and the Procedures Employed in This Study

In this section of the literature review, the research having the greatest influence on the direction of this study is presented. Despite the limited amount of research pertaining to the effects of placebos and feedback on detectability, the studies illustrate the conceptual foundations relevant to both the issues examined and the procedures employed in this study.

The following studies examined the validity of the guilty-knowledge technique in detecting deception in mock crime situations.

Lykken (1959) reported a study in which he examined the effectiveness of the guilty-knowledge technique in detecting deception. His sample consisted of 49 male subjects of whom 13 were randomly assigned to commit both a mock murder and mock theft; 24 were randomly assigned to commit only one of these two crimes; and 12 were randomly assigned to commit neither of these two crimes. Each subject was blindfolded and informed that if his physiological responses during the polygraph test were indicative of guilt, he would be given an electric shock. Each subject was actually given electric shocks on an a priori schedule irrespective of their physiological responses to increase their motivation to escape detection. The subjects were also instructed to remain silent during the polygraph testing.

Twelve question series were administered to each subject. Six pertained to the mock murder, and six pertained to the mock theft. The number of questions contained in each series varied, having a mean of 4.67 alternatives on the six murder question sets and a mean of 5.0 alternatives on the six theft question sets. Only one of the alternatives within each question set conformed directly to some apsect of one of the crimes; the other alternatives were structured to be equally plausible to an innocent person. For example, Lykken used the following alternatives regarding the desk drawer in which the mock murderer hid the weapon: (a) upper left, (b) lower right, (c) lower left, (d) upper right, (e) middle. The guilty-knowledge technique is based on the premise that a guilty person will recognize the relevant alternative and have a different physiological response to it than to the other alternatives, whereas an innocent person who is unaware which alternative is the relevant stimulus should respond greater to the relevant alternatives only at chance levels.

The only physiological response monitored during the testing was skin conductance. The skin conductance amplitudes associated with the various alternatives within each of the 12 question series were ranked. If the largest amplitude occurring during a question set was to the relevant alternative, it was given a score of 2. If the second largest amplitude was to the relevant alternative, it was given the score of 1. Thus, a perfect guilty score for each of the two crimes would be 12. Lykken analyzed each of the crimes independently, categorizing scores of 6 and less as indicative of innocence. Using that system, 88 percent (44 out of 50) of the guilt classifications were correct and 100 percent (all 48) of the innocent classifications were correct.

Davidson (1968) also used Lykken's (1959) scoring procedure in a detection-of-deception experiment. Davidson randomly assigned a total of 48 subjects into 12 groups, each containing 4 subjects. Three of the four subjects in each group were motivated to commit a self-planned mock murder; however, the experiment was designed so that of those three subjects in each group one succeeded, one attempted but failed, and one did not make an attempt. The fourth subject in each group had no knowledge of the mock murder. Thus, only 12 out of the 48 subjects were actually guilty of committing the mock murder.

Davidson examined the effects of motivation stemming from monetary reward on the detection of deception by providing each of the 12 mock murder victims with an envelope containing a pay voucher. The subjects were told about the voucher and informed that if they successfully committed their mock murder and successfully "beat" a polygraph test regarding their guilt in that incident, they could keep the pay voucher. Six of the vouchers ranged in value from $25 to $50, whereas the other six ranged in value from 1¢ to 10¢.

The questioning and scoring procedures used by Davidson in this study were essentially the same ones developed by Lykken (1959). Davidson did, however, report monitoring cardiovascular activity and respiration (these indices were not scored). Davidson reported that all 36 of the innocent subjects and 11 out of 12 (91.7 percent) of the guilty subjects were correctly classified using the same criteria reported by Lykken. The only misclassified guilty subject was in the low-amount pay voucher group; however, the detection rates between the two monetary motivation groups were not significantly different.

Another detection-of-deception study that incorporated Lykken's (1959) procedure was conducted by Podlesney and Raskin (1978). Their study also examined the effectiveness of the control-question technique in a laboratory situation; however, only the guilty-knowledge portion of that study will be addressed in this section. Twenty subjects took part in the guilty-knowledge experiment, of whom half were assigned to the guilty condition and half to the innocent condition. Guilty subjects committed a highly ego involving mock theft, whereas the innocent subjects were told about the theft but they neither enacted the crime nor were told any of the details pertaining to the theft. Subjects in both groups were informed that they would receive a $10 bonus if they appeared innocent on the lie detector test.

Five question series were administered to each subject during the polygraph test, each consisting of one relevant and four irrelevant questions. The subject's respiration, skin conductance, and cardiovascular activity were monitored during the polygraph test. Each of the three physiological indices was objectively quantified in several ways, then scored independently using the Lykken (1959) procedure. Only skin conductance and plethysmograph scores significantly discriminated between guilty and innocent subjects. However, guilt/innocence classifications based solely on the objective quantification of skin conductance responses were 90 percent correct, and all errors were false negatives.

The following studies attempted to determine whether the lie detection procedure was vulnerable to certain methods employed by subjects to deceive the examiner.

Lykken (1960) conducted a study to determine if the GSR could effectively be used to detect deception if the subjects attempted to "beat" the test. He trained 20 college students in the theory of using the GSR to detect guilty knowledge. He also allowed his subjects to practice inhibiting or producing false GSR's and informed them of the interrogation and scoring procedure that would be used. To ensure that the subjects were motivated to "beat the machine," they were offered ten dollars if they were successful.

Instead of using a mock crime situation, the subjects were tested on personally relevant information (i.e., father's name, name of high school, etc.). The subjects were divided into subsets consisting of five subjects. During the polygraph test the subjects were asked a series of questions containing one relevant and four irrelevant questions pertaining to each category of personally relevant information (i.e., "What is your father's name?" followed by the alternatives). The experimenter's task was to independently match the responses to the stimuli presented during the polygraph test for each subset of five subjects with the information contained on questionnaires those subjects completed prior to testing. Lykken reported obtaining a 100 percent correct classification using objective scoring of the GSR protocol alone.

Kubis (1962) also examined the effectiveness of certain methods of trying to "beat the polygraph" in a laboratory study. Twenty subjects used muscle tension, exciting imagery, and yoga during different tests. Subjects using the exciting imagery were instructed to think of something exciting or upsetting at the appropriate times during the interrogation procedure. The subjects using the muscle-tension method attempted to induce reactions symptomatic of deception by pressing their toes against the floor. The yoga group tried to avoid detection by maintaining an abstract frame of mind that would allow them to separate themselves mentally from the outside stimuli. The examiner used a standard three-channel polygraph (GSR, Respiration, and Cardiovascular Activity) in an attempt to determine which number a subject picked during a card test. It was found that the yoga method was not very successful; however, both the muscle-tension and the exciting-imagery methods reduced the examiner's effectiveness from the high-statistical-significance to the chance level.

Weinstein, Abrams, and Gibbons (1970) conducted a study in which they examined the effect of hypnotically induced repression and guilt. They selected six college students on the basis of their ability to enter deep hypnotic states. The subjects were divided into two groups. The three members of the first group were told to enter an office and take one of three bills ($1, $5, or $20). Afterwards, they were hypnotized and told that they would not recall taking the money. The second group of students did not take any money; however, they were told under hypnosis that they had stolen one of the bills and that they would experience considerable guilt because of this. The examiner was completely misled by the three innocent students. In fact, he stated with certainty that each had taken the hypnotically suggested amount. The examiner was only partly convinced that the members of the guilty group had taken the money and only correctly identified the amount taken by one of them.

As illustrated by the preceding studies, attempts to "beat" the lie detector normally consist in the individual trying either to reduce his responses to the critical items incorporated in the test or to create accentuated responses to the noncritical items. The subject usually attempts to accomplish this chemically, mentally, or through some form of movement. The studies conducted by Kubis (1962) and Weinstein et al. (1970) suggested that it might be easier for the subject to create responses to noncritical items than to suppress responses to the critical items. However, in certain field-testing procedures (e.g., pre-employment screening, in which most of the questions could be considered "critical items"), the subject's ability to suppress his/her responses to the critical items is more crucial. In addition, a well-trained examiner would probably notice if the subject was engaging in certain methods of attempting to create accentuated responses to noncritical items during field polygraph tests.

An alert examiner should also be able to detect physical signs during the pretest interview if the individual has consumed a sufficient quantity of drugs (alcohol included) to markedly alter his/her responses. The pretest interview is a standard field procedure, during which the examiner conditions the subject for the test and tailors the test questions to the individual and the information he/she provides. If a highly drugged person was able to get through the pretest interview undetected, his/her pattern would probably be either so erratic or so flat (depending on the type of drug) that the examiner would probably request another test date or judge the test inconclusive.

Ferguson and Miller (1973) reported that an individual's GSR pattern can be used to differentiate between responses caused by physical movement as opposed to those caused by emotion. The examiner may also see that individual making obvious physical movements.

Since the responses indicative of deception are theoretically affected by the subject's degree of concern over the possibility of detection, a more productive method of attempting to deceive the examiner might be by reducing this concern. Three factors associated with the subject's degree of concern are: (1) his/her involvement in the matter being tested; (2) the magnitude of the sanctions contingent on the testing, as perceived by the subject; and (3) the subject's degree of certainty that the true status of his/her involvement in the offense will be correctly or incorrectly diagnosed by the examiner.

The importance of the perceived sanctions contingent on testing was demonstrated in a study by Gustafson and Orne (1963), who examined the effects of the subject's level of motivation to escape detection on the accuracy of detection of deception. Thirty-six college students were divided into two groups; one group was motivated to deceive the examiner, and the other was not. Subjects who were motivated to deceive the examiner listened to a recording that contained the following information: (1) the experiment was designed to see how well the subject could keep information away from the experimenter; (2) that this was extremely difficult to do, and that only people of superior intelligence and great emotional control were able to do it; (3) they were to try as hard as they could to beat the experimenter and the equipment; and (4) if they were successful, they would receive an extra dollar.

Subjects who were motivated to deceive produced larger skin responses more frequently than did the other group. The objective scoring procedure successfully detected the information processed by members of the motivated group at a much greater than chance level, whereas detection occurred only at a chance level in the other group. The researchers concluded that the degree of autonomic response to significant stimuli appeared to be a function of motivation.

In field situations, however, the subject's degree of confidence in the outcome he/she expects seems to be the factor associated with the subject's degree of concern that is the most independent from the context of the testing situation. This appears true because both the subject's involvement in the matter being tested and the magnitude of the sanctions contingent upon the testing are to a large extent dictated by the actual circumstances.

One factor that might affect the subject's degree of certainty that his/her involvement in the offense will be correctly or incorrectly diagnosed is feedback from prior polygraph testing.

Gustafson and Orne (1965) examined the effect of perceived role and role success on GSR for deception analysis. Sixty-four college students were divided into two groups: need to deceive and need to be detected. The need-to-deceive members listened to a tape that attributed positive qualities to those who could "beat the machine." The other group listened to a different tape, which, conversely, gave positive attributes to those normally detected by the "lie detector." After the subjects completed the initial test, and regardless of the actual results, half of each group were told that they had been detected; the other half were told that they had not been detected. The investigators found that if the subjects received information that was consistent with their perceived roles, they were detected significantly less frequently than were subjects who received information not consistent with their roles.

The following comments by Reid and Inbau (1966) illustrate that similar factors may affect polygraph examinations in the field:

> A subject's concern over the possibility of detection appears to be the principal factor accounting for the physiological changes that are recorded and interpreted as symptoms of deception. . . . Conversely, a lack of concern over the possibility of detection may prevent a diagnosis of deception. . . . There is the rare subject who, because of the positive evidence against him, has developed an attitude of hopelessness; in other words, he has "given up" and abandoned any expectation of ultimate clearance of suspicion or accusation. As to him, too, the test results may be inconclusive (p. 168).

Another measure that might affect the subject's degree of certainty that the examiner will correctly or incorrectly diagnose his/her involvement in the offense is the placebo effect. This hypothesis appears to be supported by the following comments made by Reid and Inbau (1966):

> An unwitting type of psychological evasion may result from a subject's belief, however unfounded it may be, that something he has done of a physical or medical nature will prevent a display of deception criteria during the test. For instance, if he has taken a sedative or some other drug which he fully believes to be effective in permitting him to evade detection, he may thereby be relieved of the necessary concern over possible detection and either avoid deception reactions or produce a polygraph record that will not permit a definite diagnosis one way or the other. Another example--and an actual one in our own experience--is that of a police officer of limited intelligence, who, immediately prior to the test, was observed placing bullets under the pneumograph tube and under the blood pressure-pulse cuff. He apparently believed that by doing so he would suppress whatever deception indications he would otherwise display during the test. His polygraph records (either because of this belief or for some other reason) were devoid of deception criteria when asked the relevant as well as the control questions, and his deception would have remained undetected had the bullet stuffing efforts not been observed from the adjoining observation room (p. 167).

In an article on individuals who are nonreactors during polygraph examinations, Arther (1977) stated that phenomenon could result from the person's taking drugs before the examination. His comments further supported the premise that in certain cases the placebo effect might seriously jeopardize the outcome of polygraph examinations:

> There are two aspects of this problem--the physiological and the psychological.
>
> The physiological aspect deals with what drug(s) is involved, in what amounts it has been taken, what is the person's tolerance to that particular drug, when he last had it, what was already in his stomach when he took it, his overall physical condition.
>
> Just as important is the psychological aspect. That is: Does the person really believe that whatever he has taken will really result in his "beating the lie detector?"
>
> Of course, the more he is convinced that the drug will "beat the lie detector", the more likely he will be a non-reactor. This is even true when the drug supposedly should not cause the test to be affected. In fact, even a placebo can result in a non-reactor (p. 3).

Despite the belief held by many polygraph examiners that the placebo effect could seriously affect an individual's responses during a polygraph examination, this hypothesis had never been scientifically examined. However, numerous studies have dealt with the placebo effect in other situations. Several of these studies will now be discussed, since they indicate the strength of this effect and the factors affecting its magnitude.

The term placebo is by no means new, nor has its definition remained consistent throughout the years. Shapiro (1968) traced the semantic changes associated with the word, from the Hebrew Bible to the present. Today the term placebo is generally used to connote a nonactive tablet or capsule, normally lactose, for use as a control measure in pharmacological research or as a therapeutic agent administered by a physician to promote or reinforce the patient's favorable expectancies. The term placebo effect is used in this study to refer to the change in the outcome of a given situation attributable solely to the psychological effect produced by some form of intervention taken by that person, which altered his/her set of expectancies regarding the probable outcome of that situation.

In this section of the literature review, selected studies that treated the placebo effect as their principal independent variable are presented. The first studies demonstrate the strength of the placebo effect on a variety of dependent measures. After laying that foundation, the studies presented focus on the independent variables that affect the magnitude of the placebo effect.

Evans (1969) examined the relationship between the placebo response and hypnotic susceptibility. Although he did not find a strong relationship between those two variables, he did report significant reductions in both subjective and objective measures of ischemic muscle pain after subjects ingested a placebo. The objective measures used in that experiment were the volume of water a subject could pump from one flask to another with a sphygmomanometer cuff around his/her arm inflated to 200 mm mercury above his/her systolic pressure, and the length of time it took to do this.

In another study examining the effectiveness of placebos in combating pain, Beecher (1965) reported placebos were far more effective in reducing pathological pain than pain generated experimentally. He found that placebos' average effectiveness in cases pertaining to pathological pain of the tissue was 35%, whereas their effectiveness was only 3.2% with experimentally contrived pain produced by heat, tourniquet, etc. This difference suggests that when anxiety and stress are severe, placebos are more effective than when stress is of a lesser degree or absent (Stroebel, 1972).

In an experimental study, Gottschalk and Gleser (1969) examined the influence of placebos on achievement strivings. They reported that subjects who read a written statement saying that they would feel "more peppy and energetic" after ingesting their medication, showed significant increases in achievement strivings (as measured by content analysis). This relationship was consistent, whether or not the subjects received a placebo, secobarbitol (100 mg), or dextroamphetamine (10 mg). However, the group receiving the dextroamphetamine did have significantly higher achievement strivings than did the other two groups.

Weiner and Sierad (1975) conducted an experiment on the relationship between the placebo effect and achievement needs. In that study, 200 male subjects were classified as having either high or low achievement needs. The subjects in the treatment condition were given a placebo, coupled with the suggestion that the drug would interfere with their hand-eye coordination. All subjects were then given four trials at a digit-symbol substitution task that was structured in such a fashion that no one was able to complete it successfully; however, the number of digit-symbol substitutions was recorded for each trial. Compared with subjects in the control groups, ascription of failure to the pill augmented the performance of subjects low in achievement needs, whereas it decreased the performance of subjects high in achievement needs.

Sternbach (1964) examined the effects of placebos on stomach motility. Each of six subjects was tested under the following three conditions: (1) stimulant, (2) relaxant, and (3) placebo. In each instance, the drug administered was a placebo containing only a small magnet, which was used to measure gastric peristaltic rate. The subjects were informed the "relaxant" would decrease stomach motility and the stimulant increase it, whereas the placebo would have no effect. For four of the six subjects, stomach motility was highest after they took the "stimulant" placebo and lowest after they ingested the "relaxant" placebo. In one case the reverse order was found; in the other case the order was mixed. Overall, the placebo effects conformed to the suggestions with which they were paired and their mean differences were statistically significant ($p < .05$).

The studies presented above demonstrate that the placebo effect can have a dramatic influence on a wide range of dependent variables. Sternbach's (1964) article is of particular importance because the dependent measure (stomach motility) was primarily controlled by the autonomic nervous system. This suggests the placebo effect might also affect the autonomic functions monitored during polygraph examinations. The next articles reviewed depict the independent variables affecting the magnitude of the placebo effect.

Lasagna, Mosteller, Von Fesinger, and Beecher (1954) conducted a study in which they examined differences between placebo reactors and nonreactors. Their sample comprised 162 postoperative patients, who were observed for their ability to receive pain relief from subcutaneous injections of both saline solution (placebo) and morphine. The researchers reported that reactors were more likely to (1) like "everyone," (2) report the hospital care was "wonderful," (3) exhibit somatic symptoms under stress, (4) be "talkers," (5) have less education, (6) be regular churchgoers, (7) be slightly older, and (8) have different Rorschach scores than the nonreactors. No significant differences were found on the basis of sex or intelligence (as measured by the Wechsler-Belleveue scale).

Rickels, Hesbacher, Weise, Gray, and Feldman (1970) studied the placebo response in psychoneurotic outpatients. They reported clinical improvement was significantly correlated with the number of placebos taken daily for clinic and general practice patients; however, this relationship did not hold for private psychiatric practice patients. It was also found that the patient's social class seemed to influence the results. In comparison to other SES groups, patients in the lowest SES group were most likely to report improvement, regardless of the number of placebos taken. Patients in both the lowest and the highest SES groups also reported significantly greater improvement as the placebo dosage increased, whereas this relationship was not significant for patients in the middle SES group.

Buckalew (1972) conducted a study to analyze the experimental components in a placebo effect. Fifty subjects were required to complete both a pretest and posttest measuring their motor skill reaction time to a visual stimulus. The subjects were randomly assigned to the following conditions: (1) control, (2) placebo only, (3) placebo plus reinforcement, (4) placebo plus suggestion, and (5) placebo plus both reinforcement and suggestion. Subjects receiving only the placebo were not told how it would affect their performance, whereas those receiving the suggestion were told that the placebo would reduce their reaction time. The reinforcement consisted in the experimenter informing those subjects that their reaction time was improving. The order of the mean reaction times for the groups, from slowest to fastest, was: (1) placebo only, (2) control, (3) placebo plus reinforcement, (4) placebo plus suggestion, and (5) placebo plus both suggestion and reinforcement. The last group performed the test significantly ($p < .01$) faster than all other groups except the placebo-plus-suggestion group, based on the results derived from Duncan's new multiple range test.

CHAPTER II

METHOD

This chapter contains a discussion of the selection of subjects for the study, their assignment to treatment groups, the apparatus employed, the procedures of the study, and the scoring methods used.

Selection of Subjects

The subjects were volunteers enrolled in selected Criminal Justice classes as Michigan State University during Fall term 1977. The courses from which subjects were drawn were: two sections of Introduction to Criminal Justice, Criminology, Police Process, and Juvenile Delinquency. These courses were selected on the basis of the following criteria: They were all large, undergraduate classes in which the instructor agreed to permit his/her students to participate in the experiment for extra credit. To maintain consistency, the extra credit was standardized for all classes, based on a set percentage of the total points for each class.

During the first week of the Fall (1977) term, the investigator visited each of the aforementioned classes. Students attending these classes were told the following information:

1. The purpose of this study is to examine the effects of certain drugs on the accuracy of lie detection tests.

2. The experiment itself will involve a mock contract murder. After the volunteer has been briefed about his/her intended victim, the volunteer will be required to shoot at an image of the victim shown on a movie screen. The volunteer will then be given a lie detection test concerning the mock murder. The person administering the lie detection test will attempt to ascertain the name of the victim, what the victim's occupation was, how many times the volunteer shot at him, where the Mafia family that hired the volunteer was located, and how much the volunteer was paid for the killing.

3. One percent of the total possible points in the student's respective class will be given to each subject who completes the experiment without breaking any of the rules. However, subjects who are able to "beat" the lie detector will be awarded a total of 5 percent of the total possible points. The number of points each subject will be awarded will be determined by objectively scoring his lie detection charts. If the individual who analyzes the charts is able to correctly identify the information pertaining to the subject on three or more of the five tests, the subject will receive only 1 percent extra credit. However, if the subject successfully deceives the examiner on three or more of the tests, he/she will be awarded the 5 percent extra credit.

4. The effects of two "drugs" on lie detection will be studied in this experiment. One of the drugs is believed to make it more difficult for a person to "beat" the lie detector, whereas the ingredients in the other "medication" should make it much easier. Volunteers will be randomly assigned to one of three treatment groups.

One of the groups will receive the "drug" that should make it harder for them to successfully deceive the examiner. Another group will receive the "medication" that should help them "beat" the lie detector, whereas the third group will receive no medication at all.

5. For a student to participate in this study, all of the following conditions must be met:

a. The student has to volunteer for the experiment.

b. The student must sign an informed consent form.

c. The student must sign a medical release form, which will permit physicians at the Olin Health Center to examine the subject's medical history.

d. A physician must verify that there is nothing in the subject's medical records that indicates that either of the "drugs" would have a harmful effect on him/her.

e. The volunteer has to agree not to discuss this experiment with any other volunteers until the study has been completed.

It is important for the reader to note that in order to realize the research objectives of this study, the subjects had to believe the above information was correct. Once again, however, the actual purpose of this study was to examine effects of placebos and feedback on the detection of deception. The two "drugs" that were administered were in reality two differently colored pharmaceutical placebos containing only lactose. Since the "drugs" were only placebos, no actual check was made of the volunteers' medical records. The medical release form was included to give the placebo additional credibility and to prevent any possible complaints of unauthorized disclosure of university medical records.

Immediately after the study was explained to the students, any questions they had regarding the experiment were answered. Students interested in participating in the study were given a copy of both the informed consent form (Appendix A) and the medical release form (Appendix B) and were asked to read them carefully. Students who still wanted to take part in the study were then asked to sign and return the forms.

Each of the volunteers was given a subject number and was randomly assigned to one of nine treatment groups. Since the possibility existed that the subject's sex might be associated with autonomic responsivity, the ratio between males and females was held constant for all groups. It may be of interest that although the investigator did not anticipate finding significant differences between males and females, the almost even split between male and female volunteers made examining this hypothesis too attractive to ignore. In this experiment the two principal independent variables (placebos and prior feedback) were both manipulated to produce three levels of treatment for each variable, yielding the treatment matrix presented in Figure 1.

A more detailed description of these groups is presented in the procedure section. Each of the nine groups included 15 males and 15 females, producing an _N_ of 270. The ages of the subjects ranged from 17 to 42 ($\bar{X}$ = 19.8; _SD_ = 2.34). The breakdown of the subjects' year in school was as follows: freshmen, 91; sophomores, 78; juniors, 67; seniors, 38; graduates, 1.

		PLACEBO CONDITION					
		PASS		CONTROL		FAIL	
		Male	Female	Male	Female	Male	Female
FEEDBACK CONDITION	PASS	n=15	n=15	n=15	n=15	n=15	n=15
	CONTROL	n=15	n=15	n=15	n=15	n=15	n=15
	FAIL	n=15	n=15	n=15	n=15	n=15	n=15

Figure 1.--Treatment matrix.

Apparatus

A Stoelting field polygraph (model #22642) was used to record both the respiration and the skin resistance responses (SRR) of the subjects. Respiration was recorded by a pneumatic tube positioned around the subject's thoracic area. The SRR was recorded from two stainless steel electrodes attached to the volar surfaces of the first and third fingers of the subject's right hand. All SRR recordings were made with the instrument in the manual centering model.

The instrument used to objectively score both respiration responses and one of the measures of SRR examined in this study was a modified map-distance measurer. The instrument was designed to measure curvilinear distances between two points on a sheet of paper.[1] The instrument's original 1/4"-diameter circular wheel which came into contact with the line on the paper being measured was replaced with a 10-tooth gear having an outer circumference diameter of 2 mm. This modification made it easier to keep from deviating from the paths on the charts formed by the polygraph's ink pens. The gear was also less susceptible than the original wheel to sliding on the paper as opposed to turning, which was necessary to achieve an accurate measurement of the distances. The original map distance measurer was also filed in certain places to permit the aforementioned gear to have free contact with the surface of the polygraph charts.

Procedure

All subjects reported individually to the room where they were to commit their mock murder. When each subject arrived, he/she first met with a research assistant who worked independently of the polygraph examiner. Each of the research assistants who worked in this capacity had received training from the Director of the Psychology Clinic at Michigan State University on how to identify and interact with subjects who might have an adverse psychological reaction to the experiment. When each subject arrived for the experiment, the research assistant greeted the subject at the door and shook his/her hand. The handshake served two purposes. First, it was intended to

---

[1]The writer gratefully acknowledges the suggestion of Dr. Frank Horvath to measure physiological responses with this type of instrument.

facilitate an open interaction between the subject and the research assistant. Second, it helped the research assistant ascertain whether the subject was overly anxious about participating in the study. Subjects who had cool, clammy, sweaty hands or who exhibited other behavioral signs symptomatic of excessive anxiety (i.e., fast talking, hyperactivity, or other overt signs of nervousness) received special attention from the research assistants. Basically, this entailed the research assistant spending additional time conversing with the subject on matters not related to the study, such as courses, sports, or weather.

After the subject appeared relatively calm, the research assistant summarized what the experiment would entail. The subject was then shown his/her mock murder contract (Appendix C). This document specified the following information: the name of the individual the subject was to simulate killing, the victim's occupation, the amount of play money the subject was to receive, the number of shots the subject had to fire, the location of the Mafia family that was purchasing his/her services, and a picture of the intended victim. The contract also specified that the subject was required to say, "(victim's name), I am shooting you for betraying the (city where the Mafia family was located) branch of the Mafia," before he/she fired the pistol.

Each subject was asked to read the mock murder contract silently, while the research assistant read the document aloud. The research assistant also answered any questions the subject raised. After reviewing the contract, each subject was given the option of withdrawing from the study and still receiving 1 percent extra credit in his/her respective class. Only two subjects withdrew from the study at that point.

The research assistant was also responsible for administering the placebos to subjects assigned to certain groups. The assistant gave all subjects assigned to the placebo-fail group a yellow placebo and informed them that the drug should make it more difficult to "beat" the lie detector. Subjects in the placebo-pass condition were given an orange placebo and told that the "drug" should make it easier for them to "beat" the lie detector. Subjects assigned to the placebo-control group were not given a placebo. It is important for the reader to note that all of the research assistants were led to believe the "medication" they were dispensing was active medication, not placebos.

The research assistant removed a capsule for subjects assigned to the placebo-pass and placebo-fail groups from the appropriate university prescription-medication vial containing the respective placebos. The label on the vial containing the placebos for the placebo-pass group included the following information: "A Tranquilizing Agent to decrease emotional responses," and the medication number 1139. Conversely, the label on the vial containing the placebo-fail capsules contained the following: "An Adrenergic Agent to increase emotional responses," and the medication number 1134. The following additional information appeared on the outside of both prescription vials: the prescribing physician's name, that the medication was for research purposes, and a warning that only one pill was to be taken.

The research assistant verified that all subjects who were to receive the placebo (1) did in fact take the placebo, (2) were told that the placebo should make it either easier or harder to beat the test (depending on their respective group), and (3) were told that the medication would take effect in approximately 15 minutes. The research assistant also warned all subjects that if they gave the examiner any indication of whether they had or had not received a "drug," they would be immediately disqualified from the study.

After administering the appropriate placebos to subjects assigned to either the placebo-pass or placebo-fail group, the research assistant showed the subject one of five sets of slides based on the occupation of his/her intended victim. An equal number of subjects within each of the nine placebo-feedback treatment conditions were randomly assigned to shoot at the image of either a fireman, policeman, soldier, priest, or a surgeon. Each subject was shown a total of six slides portraying one of those occupational options. The first slide in each of the occupational sets of slides depicted a building in the city in which the experiment took place that conformed to that particular occupational group (i.e., police station, church, ROTC headquarters, fire station, and University Health Service). Each of the building slides was photographed from a position that made it possible for the subjects to read the sign identifying the building that was situated in front of the structure from the projected image of the slide. The second, third, and fourth slides showed an individual working in various capacities congruous with the particular occupation he was portraying. The fifth slide projected a full-length image of that same individual. In order to give the subject the impression that the image on the fifth slide was looking at him/her, each of the individuals photographed looked at the camera when that slide was taken. The sixth slide depicted the person lying on the ground simulating death. Each of the individuals who appeared in the aforementioned slides wore a distinctive uniform appropriate for someone working in the particular occupation they were selected to portray.

Although a different individual was portrayed in each of the five occupational sets of slides, the same individual appeared in all of the slides for each of the given occupational options. In order to maintain some consistency between the individuals photographed, all five were males and had similar physical characteristics.

Photographs were reproduced from the fifth slide of each of the five possible occupational options. Of those photographs the one that corresponded to the occupational option to which the subject was randomly assigned was attached to that subject's mock murder contract. Therefore, if the subject's mock murder contract specified that his/her victim was a fireman, the subject was shown the set of slides taken of the same individual dressed in the same fireman's apparel that was displayed on his/her mock murder contract.

An equal number of subjects in each of the nine placebo/feedback treatment combinations were also randomly assigned to one of five different options for each of the following categories: (1) victim's name, (2) number of shots to be fired, (3) Mafia family location, and (4) price of the contract. The specific options corresponding to each of the categories of information used in the mock murders are listed below:

Category I Target's Occupation	Category II Target's Name	Category III No. of Shots to Be Fired
1. fireman	1. John Martin	1. 2 times
2. policeman	2. Michael Brown	2. 3 times
3. soldier	3. Edward Johnson	3. 4 times
4. priest	4. Henry Clark	4. 5 times
5. surgeon	5. Peter Miller	5. 6 times

Category IV Location of Mafia Family	Category V Contract's Price
1. Kansas City	1. $20,000
2. Miami	2. $30,000
3. Chicago	3. $40,000
4. New York	4. $50,000
5. Boston	5. $60,000

Therefore, of the 30 subjects in each of the nine placebo/feedback treatment combinations, 6 in each condition were randomly assigned to one of the five options in each category. The information corresponding to the options to which the subject was assigned was filled in on that subject's mock murder contract. The same information was then used as the specifications for that subject's mock murder.

The six slides of the subject's intended victim were shown on a white paper screen situated directly in front of a pellet backstop consisting of a wall of boxes filled with paper. The equipment was designed to enable the research assistant to provide an unpunctured screen for each subject, which was supplied by a roll of wide paper fixed atop the frame of the screen. The slide projector was placed back far enough to project life-size figures onto the screen.

The research assistant gave each subject a loaded pellet gun, closely resembling a 38-caliber revolver. Every subject was required to stand on a spot to the left side of the screen, which was close enough to make relatively certain that each shot would strike the intended victim's image. After the subject was shown the fifth slide, which was structured to give the subject the impression the victim was looking directly at him/her, the subject was required to say, "________, I am shooting you for betraying the __________ branch of the Mafia." The subject then fired at the victim the required number of times. After the subject was through firing at the image, the research assistant switched to the 1st slide, which portrayed the victim lying on the ground simulating death. Finally, the research assistant counted out the appropriate amount of play money and handed it to the subject, who was then also required to count it.

After the subject had finished counting the money, the research assistant warned the subject again that if he/she informed the lie detection examiner whether or not he had received one of the pills, or any details about the mock contract killing, he/she would immediately be disqualified from the experiment and would not receive any credit.

The subject was then sent to see the polygraph examiner, who was located in another office down the hall. The examiner spent approximately 30 minutes administering the Biographical Data Sheet (Appendix D), explaining to the subject the theory behind lie detection, and informing him/her how the equipment worked. After the explanation, the examiner gave the subject a "demonstration" of the instrument.

It is important to note that the actual purpose of this demonstration was to manipulate the nature of the feedback the subjects received from it. Subjects randomly assigned to the feedback-pass condition were led to believe they successfully deceived the examiner on the demonstration test, whereas subjects in the feedback-fail condition were led to believe they failed in that endeavor. Subjects in the feedback-control condition were not given any indication of how well they did on the demonstration test.

During the demonstration test, each subject was shown five cards that were placed face down, and was asked to shuffle them without turning any of the cards over. All of the cards given to subjects in the feedback-fail group had the number 15 on their face. Subjects assigned to the feedback-pass group were given a deck composed of the numbers 2, 4, 10, 10, and 19. Note that the number 10 appeared twice and the number 15 was omitted from that deck. Members of the feedback-control group were given a normal deck composed of the numbers 2, 4, 10, 15, and 19.

After the subject was satisfied that the cards had been adequately shuffled, he/she was asked to pull one of the cards aside, still keeping it face down. The examiner then removed the remaining four cards without looking at them. After removing the four cards, the examiner turned his back and asked the subject to turn over the card he/she had selected and to memorize the number. The subject was also required to write the number down on a pad of paper that was placed directly in front of him/her, and then to turn over the pad of paper, placing it on top of the card so the examiner could not see which card had been selected.

Before beginning the "demonstration," the subject was told that he/she would be asked a series of questions regarding the possible number chosen and that he/she should respond "no" to each question, regardless of whether it mentioned the number actually drawn. Next, the examiner wiped off the subject's fingers with a tissue to remove any excess dirt and perspiration. He then placed the GSR electrodes on the first and third fingers of the subject's right hand and adjusted the instrument. The subject was asked to close his/her eyes and face straight ahead without moving during the test.

After the examiner made certain the subject was following these directions, he asked the following questions:

1. Did you select card number 2?
2. Did you select card number 4?
3. Did you select card number 10?
4. Did you select card number 15?
5. Did you select card number 19?

During the "demonstration," the examiner increased the GSR sensitivity for subjects in the feedback-fail and feedback-pass groups immediately after he asked them if they had selected card number 15. The sensitivity was returned to its previous level before the subjects were asked the last question. This resulted in an increase in the GSR amplitude corresponding to the number 15.

Individuals in the feedback-pass and feedback-fail groups were shown their charts. They were also informed about the method the examiner used to interpret the GSR patterns. However, the examiner did not turn the pad of paper over to confirm to disprove his interpretation of a subject's GSR responses until he was completely finished with that subject. Therefore, subjects in the feedback-fail group were intentionally led to believe that they had been detected, whereas those in the feedback-pass group were led to believe they had successfully deceived the examiner. Subjects in the feedback-control group were not shown the results of their "demonstration" test.

After discussing the "demonstration" test with the subjects, the examiner informed them that he was interested in determining whether or not individuals were cognizant of how well they did on polygraph tests. The subjects were also told that in order to resolve this problem the examiner would have to control for the subjects' preconception of how well they thought they would perform on the actual test. Each subject was questioned to make certain he/she understood that the examiner was interested in determining how well they thought they would do on the actual test, not how they had done on the "demonstration" test. The examiner also stressed that this aspect of the study would have no bearing on the number of extra-credit points the subject would be awarded.

After the subject appeared to understand the above information, he/she was asked to place an "X" next to the one statement contained in a performance expectancy self-report that most applied to him/her. The performance expectancy self-report contained the following series of statements:

1. I am almost positive I will "beat" three out of the five tests.
2. I am pretty sure I will "beat" three out of the five tests.
3. I have absolutely no idea how well I will do.
4. I am pretty sure I will not "beat" three out of the five tests.
5. I am almost positive I will not "beat" three out of the five tests.

To minimize the level of additional contamination that could have stemmed from this procedure, the examiner turned his back while the subject checked the appropriate number. The subject was also asked to turn the pad upside down after marking it so the examiner could not see how he/she had responded.

After subjects completed that procedure, the examiner administered the actual test. He asked each of the subjects the sets of questions listed below.

A. During the following series of questions you will be asked about the victim's occupation. Are you ready for me to begin?

    1. Was the person you shot a doorman?
    2. Was the person you shot a fireman?
    3. Was the person you shot a soldier?
    4. Was the person you shot a surgeon?
    5. Was the person you shot a priest?
    6. Was the person you shot a policeman?

B. During the following series of questions you will be asked about the victim's name. Are you ready for me to begin?

1. Was the person you shot named Thomas Wilson?

2. Was the person you shot named John Martin?

3. Was the person you shot named Michael Brown?

4. Was the person you shot named Edward Johnson?

5. Was the person you shot named Henry Clark?

6. Was the person you shot named Peter Miller?

C. During the following series of questions you will be asked about the number of times you shot the victim. Are you ready for me to begin?

1. Did you fire one shot at the victim?

2. Did you fire two shots at the victim?

3. Did you fire three shots at the victim?

4. Did you fire four shots at the victim?

5. Did you fire five shots at the victim?

6. Did you fire six shots at the victim?

D. During the following series of questions you will be asked about the location of the Mafia organization that hired you. Are you ready for me to begin?

1. Were you hired by the Los Angeles branch of the Mafia?

2. Were you hired by the New York branch of the Mafia?

3. Were you hired by the Miami branch of the Mafia?

4. Were you hired by the Chicago branch of the Mafia?

5. Were you hired by the Boston branch of the Mafia?

6. Were you hired by the Kansas City branch of the Mafia?

E. During the following series of questions you will be asked about how much you were paid for murdering the victim. Are you ready for me to begin?

1. Were you paid $10,000?

2. Were you paid $20,000?

3. Were you paid $30,000?

4. Were you paid $40,000?

5. Were you paid $50,000?

6. Were you paid $60,000?

It is important to note that question 1 in each of the above series of questions did not represent one of the possible options to which the subject could have been randomly assigned. These questions were included to serve as a buffer for the subject's initial physiological responses associated with the introduction of a new question series.

To avoid some of the possible experimenter contamination due to the examiner knowing the subject's feedback classification, the questions were tape-recorded. The questions asked on the tape were presented at 20-second intervals; there were 30-second intervals between the different test series.

Before testing, the subject's fingers were wiped off with tissue and he/she was asked to close his/her eyes and face forward without moving while responding to the taped questions. The subject was also instructed to respond "no" to each question he/she was asked during the test, except the questions asking the subject if he/she was ready to begin the new test series. These questions were included to help make sure the subject paid attention to the content of the questions.

After the testing was completed, the attachments were removed and the subject was asked to think about how well he/she had done in his/her attempt to deceive the polygraph examiner. Once again, after being told that this information would have no bearing on the number of extra-credit points he/she would receive, the subject was asked to place an "X" beside the one statement contained in the second performance expectancy self-report that most applied to him/her. The second performance expectancy self-report instrument consisted of the following statements:

1. I am almost positive I "beat" three out of the five tests.
2. I am pretty sure I "beat" three out of the five tests.
3. I have absolutely no idea how well I did.
4. I am pretty sure I did not "beat" three out of the five tests.
5. I am almost positive I did not "beat" three out of the five tests.

To reduce the effect of the experimenter's presence on this aspect of the experiment, he turned his back while the subject checked the appropriate number. The subject was asked to turn the pad upside down after marking it so the examiner would not see the response until he turned the pad over.

Subjects were then thanked for participating in the study and were told that they would be informed later in the term how many extra-credit points they would receive. It is important to note that no subjects were permitted to see their charts or to find out how many points they had received until all of the subjects had been tested, since their feedback to other volunteers might have contaminated the study.

Immediately before the subjects were told how many extra-credit points they had received, they were asked to complete a brief questionnaire (Appendix E). This instrument was designed to query the subjects about their overall perceptions of the experiment, as well as about any methods they might have employed to assist them in their endeavor to beat the lie detector.

Objective Scoring Procedures

Two physiological parameters were recorded continuously during the polygraph examinations: thoracic respiration and skin resistance. Described below are the methods used to objectively score these physiological patterns.

Respiration Total Length

The respiration patterns were objectively scored by measuring the total length of the pattern produced by the polygraph respiration pen from the instant the stimulus question was asked until 15 seconds had transpired. This distance was measured using the modified map distance measurer described in the Apparatus section. Since the nature of the respiration pattern is affected by many variables, such as the subject's degree of obesity, the tightness of the pneumograph tube, and individual breathing differences, comparisons for detection-of-deception purposes were restricted to differences across the same person.

The respiration patterns corresponding to the five questions associated with each test were ranked from 1 to 5, using the method described above. Since the suppression of breathing is generally associated with deception, the shortest pattern was assigned the rank of 1. The other four responses were then ranked from 2 to 5, using the same procedure.

<u>GSR Total Length</u>

One of the objective procedures used to score the subjects' electrodermal patterns was to measure the total length of the pattern formed by the polygraph's GSR pen from the instant the stimulus question was asked until 15 seconds had transpired. This distance was also measured using the aforementioned modified map distance measurer. The recorded measures of electrodermal activity, like those for respiration, are affected by many other variables in addition to those associated with deception. In this case such factors as the level of sensitivity at which the instrument was set, the humidity in the room, and individual differences make direct comparisons across subjects meaningless. Therefore, all of the objective methods used to score skin resistance were eventually ranked from 1 to 5 based on a comparison of the patterns corresponding to the five questions for each test.

A relatively large decrease in skin resistance is generally associated with deception. The polygraph used in this experiment was designed to show decreases in skin resistance as upward movements by the pen used to record electrodermal activity. Therefore, of the five electrodermal responses associated with the questions on each test, the largest response was assigned the rank of 1 (most indicative of deception), whereas the smallest was given the rank of 5 (least indicative of deception).

GSR Amplitude

Another objective procedure used to score the subjects' electrodermal patterns was to measure the vertical rise of the largest wave occurring from onset of the stimulus question until 15 seconds had transpired. The length of the vertical rise was measured from the lowest point prior to the wave's assuming a positive slope to the highest point it reached within the 15-second period. (See Figure 2.)

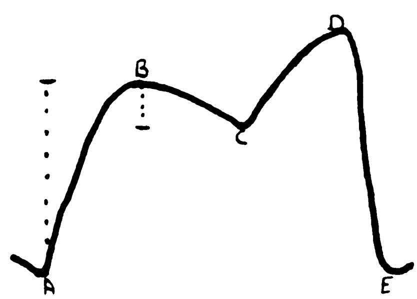


Figure 2.--Example of GSR amplitude.

To determine if more than one wave was present, the following method was employed: If the vertical rise from A to B was more than twice the vertical decline from B to C, then ABC was not treated as a separate wave. Therefore, the vertical increase from A to D would constitute the "GSR amplitude" in this type of situation.

If there was absolutely no positive rise during the 15-second interval, the response was called a GSR amplitude falling pattern. It was impossible to assign "falling patterns" a numerical value because the instrument used did not specifically indicate ohm levels and did not reflect a consistent decrease in ohms for equal mm of vertical decline (unless the starting points of the GSR pen were exactly the same and the pen was not mechanically raised). Unfortunately, this problem persisted even if the tangent error due to the curved path of the polygraph's ink pens was taken into account.

The values for the GSR amplitude for the five questions associated with each test were ranked from 1 (largest value) to 5 (smallest value). If "falling patterns" were included among the five, they were assigned equal ranks, which denoted the smallest measurements. Therefore, if only one "falling pattern" occurred among the five, it was assigned a rank of 5; if two occurred, they were both given the rank of 4.5; if three occurred, all three were ranked 4; and so on.

GSR Maximum Height

The last procedure used to objectively score the electrodermal patterns was to measure the highest point reached during the 15-second interval commencing the instant each stimulus question was asked. This was accomplished by measuring the length in mm of a vertical line drawn from the highest point reached by the pen (during each time interval) to the bottom of the chart paper. If it was necessary to mechanically adjust the position of the GSR pen during one of the tests, the amount of increase or decrease was subtracted or added, respectively, to all the responses in that series of questions that followed the pen adjustment on that test.

If the response to the buffer question was higher than the response to the first actual question and the response to the first actual question was higher than the responses to all four of the other questions on that test, the entire test was said to have exhibited a GSR maximum height "downward drift pattern." Essentially, a "downward drift pattern" indicated that the subject's GSR pattern was falling, which implied that GSR maximum height was not an appropriate measure for detection-of-deception purposes in the manner in which it was scored in this study. This phenomenon is often referred to as either a "falling galvo" or a "plunging galvo" by field polygraph examiners. When a "downward drift pattern" was present on a test, all five responses for that test were assigned a rank of 3 for their GSR maximum height value. Otherwise, the values for the GSR maximum height for the five questions associated with each test were ranked from 1 (largest value) to 5 (smallest value).

Summary

This chapter contained a detailed explanation of the methodology used in conducting the study. Included was an account of how subjects were selected for the project, as well as their assignment to the various treatment groups. Following this, the procedures involved in carrying out the "demonstration" as well as the actual test were discussed. The final section of the chapter was an account of the objective scoring procedures used during the polygraph examinations. Chapter III contains the results of those tests.

CHAPTER III

RESULTS

Introduction

In this section of the dissertation, the methods used to analyze the data and their results are presented. The section commences with a brief description of analysis of variance and multivariate analysis of variance. These two statistical techniques were the principal methods used to analyze the data collected in this study. After laying this foundation, a more comprehensive examination of the relationships between the specified variables ensues, drawing from the data collected on the subjects' polygraph charts, performance expectancy self-reports, and questionnaires.

As previously mentioned, the principal objective of this study was to examine the effects certain placebo, sex, and feedback conditions had on the detection of deception. Other related issues that were examined include: the reliability of certain objective procedures used to quantify the data, the effect of the treatment conditions on the performance expectancy scores, the correlations between the different dependent variables, the degree of electrodermal nonresponsiveness, the detection efficiency of the different physiological indices in differentiating between critical and noncritical items, and the association between selected variables contained on both the biographical data sheet (Appendix D) and the follow-up questionnaire

(Appendix E) and both polygraph detection efficiency and electrodermal responsiveness. All statistical inferences presented in this chapter are treated as statistically significant when $p \leq .05$. Two-tailed tests of significance were used whenever Z-tests or t-tests were conducted.

Principal Statistical Techniques Employed

The principal methods used to analyze the data collected in this study were analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA). The major independent variables examined were placebo treatments, feedback treatments, and sex differences. As noted in the preceding chapter, four dependent variables were extracted from the polygraph charts. They were GSR maximum height, GSR amplitude, GSR total height, and respiration total length.

Since the investigator was interested in testing the statistical significance of each independent variable by itself as well as the interaction effect between variables, it was necessary to employ a factorial design. Analysis of variance is a statistical technique used to analyze data obtained in a factorial design when only one dependent variable is being considered. As will be shown, the scores the different subjects obtained on the various dependent measures examined in this study reflected a certain degree of variability. Part of the score variance was attributed to the fact that subjects received different placebo and feedback treatments. Other parts of the score variance were attributed to sex differences and to the interaction effects between these variables. Finally, a residual or error variance was produced, resulting from differences between subjects that were not accounted for by either the treatment or interaction effects.

Analysis of variance permitted the investigator to determine the degree of the total score variance attributable to each of these sources. It accomplishes this task by establishing a set of ratios by using the mean square for the residual component as the denominator and the mean squares of the other sources of variation as the numerators.[1] The number produced by each of these ratios is referred to as an "F" ratio. If the F score is sufficiently large, taking into account the number of factors and their levels (degrees of freedom, to be more precise) that are associated with those two sources of variation, it is said to be statistically significant at the particular level selected by the investigator. For example, if the F score for the main effect for sex is found to be significant at the .01 level, it indicates that the investigator can be 99 percent sure that the differences between males' and females' scores were not produced by chance alone. As with all inferential statistical tests, significant differences may not represent substantive differences, which are left to individual interpretation.

The specific type of analysis of variance that was employed in this study is called a three-factor, fixed-effect ANOVA. The three factors were sex, placebo treatments, and feedback treatments. Since either all levels of a factor were included in the study (i.e.,

---

[1]For a more detailed explanation, almost any intermediate statistics text can be consulted.

male-female) or only the levels of particular interest were selected (i.e., placebo pass, fail, and control; feedback pass, fail, and control), the factors were all considered fixed. The fact that the factors were fixed (as opposed to some or all being randomly sampled from a population of levels) affected the generalizability of the experiment by limiting it to the treatments actually tested, and also affected the manner in which the ANOVAs were calculated.

A posteriori contrasts were employed when statistically significant ANOVA main effects were found. These contrasts were used to pinpoint which of the treatment means were statistically significantly diffferent from the other means that were included in that main effect analysis. It should be noted that one-way ANOVA and tests for ANOVA main effects only determine whether significant differences between the various means exist. Without a posteriori contrasts one would have to rely on educated guesses when attempting to determine which means were actually statistically significantly different from the others. Tukey post hoc comparisons were used when the means being compared were based on equal cell sizes, whereas both Duncan's multiple-range test (Duncan) and Scheffe's test were utilized when the cell sizes were unequal. Of the three tests used, Duncan is the most liberal (most likely to indicate significant differences) and Scheffe's test is the most conservative (least likely to indicate significant differences).

As with all inferential statistics, the appropriateness of the results derived from analysis of variance in a factorial experiment is contingent on several assumptions. It is generally assumed that independent observations will be drawn from populations in which the dependent variable is normally distributed and that the dependent variable for these populations will have equal variances. It is fortunate that ANOVA is relatively "robust" to certain violations of these assumptions (Glass & Stanley, 1970). This is an important feature, since the ordinal-level data that were used in calculating the ANOVAs in this study might jeopardize the normality assumption. It is also possible to use certain techniques to test whether some of the assumptions have been violated. For example, in this study Cochran's C and Bartlett's Box F are reported; they examine homogeneity of variance.

The second principal technique employed to analyze the data was multivariate analysis of variance (MANOVA). Essentially, MANOVA is a generalization of ANOVA that is used to test the effects of several dependent variables simultaneously. Instead of a single measurement on each experimental unit, MANOVA integrates the individual measurements of all the dependent variables into a single vector of responses. In ANOVA the analysis is based primarily on the means of the individual variables; however, in MANOVA the analysis is based on vectors of means in which each element of the vector is a group's mean for a particular variable. The purpose of MANOVA is to test whether there are statistically significant differences among the means vectors, which are also referred to as the group's centroids. This concept is graphically illustrated in Figure 3, which depicts a simple bivariate dependent variable MANOVA situation in which the differences would be statistically significant.

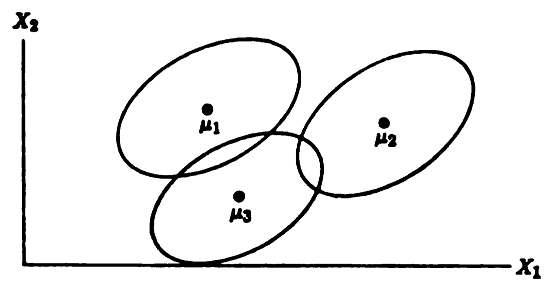


Figure 3.--A simple bivariate dependent variable MANOVA situation, in which the differences among the three populations are "real" (Cooley & Lohnes, 1971, p. 224).

As in univariate analysis of variance, the appropriateness of the inferences drawn from MANOVA is contingent upon a similar set of assumptions. The dependent variable vector is assumed to be multivariate normal in distribution with the same dispersion, or variance-covariance matrix, for each population. Equality of dispersions is the MANOVA extension of the assumption of homogeneity of variances in ANOVA designs (Cooley & Lohnes, 1971, p. 224). This assumption was tested in this study using Boxes M.[1] If a significant difference in dispersions is obtained, the F test for the differences between group centroids may be inflated.

The differences between the group centroids were examined using four different criteria to derive levels of significance. They were

[1]For a more complete description of Boxes M, see Amick and Walberg (1975).

Wilk's Lambda, Hotelling's trace criterion, Roy's largest root, and Pillai's criterion. All four criteria are a function of the eigenvalues (characteristic roots) of the ratio between the determinants of treatment sum-of-squares and cross-products matrix.[1] However, for the sake of brevity, only the value of the most liberal criterion on each of the multivariate tests will be presented in this chapter. When significant multivariate results are reported, readers interested in knowing the values for all four criteria will find them listed in Appendix F.

The MANOVAs were computed using Northwestern's version 7.0 SPSS MANOVA program at the Computer Center of Michigan State University. This program automatically calculated the approximate F-ratio and its level of significance for Wilk's Lambda, Hotelling's trace criteria, and Pillai's criteria. The level of significance associated with Roy's largest root was derived from Heck percentage-point charts contained in Timm (1975). In addition to calculating the MANOVAs, the version 7.00 MANOVA program calculated the univariate F test of significance for each response separately.

MANOVA Results

The complete results of the statistical analysis derived from the multivariate analysis of variance procedures are presented in Appendix F. Six MANOVA "runs" were made--five pertaining to the ranks of the critical items (CI) on each of the

---

[1]The formulas for calculating all four criteria are presented in Cohen and Burns (1977).

polygraph tests and one in which the CI ranks from the five polygraph tests were added together for each of the four dependent variables.

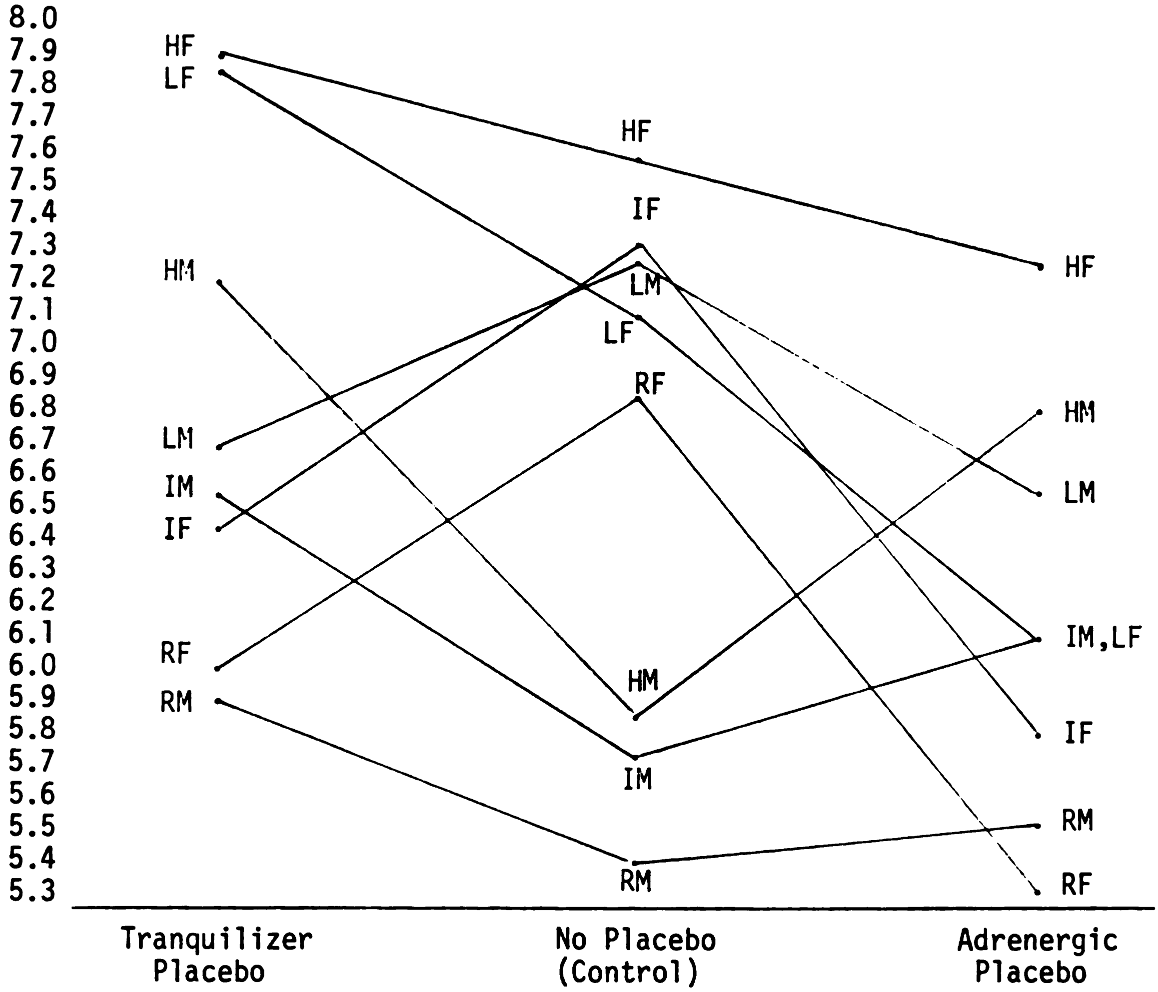
As stated in the preceding section, Boxes M is often employed to test whether the equality-of-dispersion assumption has been violated. This test was conducted on all six MANOVAs. A significant difference in dispersions was noted, corresponding to the third, fourth, and fifth polygraph tests (p = .04, .009, and .03, respectively). This finding suggests that the F value for the differences between group centroids may be inflated, especially for the values relating to the fourth polygraph test. Therefore, the validity of the multivariate significance values reported for those three polygraph tests should be viewed with a certain degree of skepticism.

None of the multivariate tests of significance for the main effects of feedback or placebo with respect to the critical item scores were significant for any of the polygraph tests or for their combined ranks on the four dependent measures. However, the multivariate tests for the main effect of sex were significant on three out of the following six tests. The F value associated with the main effect of sex on the test analyzing the combined CI ranks of the five polygraph tests for the four dependent variables was significant at p = .0009. Its values associated with polygraph tests one through five were p = .02, .11, .07, .07, and .04, respectively.

There was also a significant sex x placebo multivariate interaction (p = .001) corresponding to the fourth polygraph test; however, this relationship was not found for any of the other polygraph tests or on the analysis conducted on their combined dependent values.

The aforementioned significant interaction effect is illustrated in Figure 4. Since feedback was not part of the significant interaction, the three levels of feedback were collapsed into their corresponding drug and sex categories, yielding a possible range of CI mean rank scores from 3 to 15. The higher the mean composite rank score, the less detection efficiency the polygraph had in correctly differentiating the critical items. The CI mean scores for males on GSR maximum height, GSR amplitude, and respiration indicated that the detection efficiency was the lowest in the tranquilizer placebo treatments, almost as low in the adrenergic placebo groups, and highest in the placebo control groups. However, males' GSR total length CI mean score indicated the lowest detection efficiency for the placebo control groups ($\overline{X}$ = 7.3) and higher detection efficiency in the tranquilizer placebo and adrenergic placebo groups ($\overline{X}$ = 6.7 and 6.6, respectively). The CI mean scores for females on respiration and GSR amplitude suggested a relationship opposite to that found for males. Their CI mean scores for those dependent variables indicated the lowest detection efficiency in the placebo control groups and considerably higher efficiency levels in the tranquilizer and adrenergic placebo conditions. The detection efficiency for females on GSR total length and GSR maximum height was the lowest in the tranquilizer placebo conditions ($\overline{X}$ = 7.9 and 7.9), higher in the placebo control groups ($\overline{X}$ = 7.0 and 7.6), and highest in the adrenergic placebo conditions ($\overline{X}$ = 6.1 and 7.3).

None of the other multivariate tests of significance for the sex x feedback interaction, the drug x feedback interaction, or the sex



Key: RM = Respiration mean values for males
RF = Respiration mean values for females
IM = GSR amplitude mean values for males
IF = GSR amplitude mean values for females
HM = GSR maximum height mean values for males
HF = GSR maximum height mean values for females
LM = GSR total length mean values for males
LF = GSR total length mean values for females

Note: The higher the mean rank score, the less detection efficiency the polygraph had in correctly differentiating the critical items from the noncritical items. Possible range: 3-15.

Figure 4.--The effect of sex and three placebo conditions on the respiration, GSR amplitude, GSR maximum height, and GSR total length responses during the fourth polygraph test.

x drug x feedback interaction were significant for any of the polygraph tests.

ANOVA Results

The analysis of variance procedures examined the same relationships as discussed in the MANOVA section. However, instead of analyzing the treatment and interaction effects on all dependent variables simultaneously, ANOVA was used to examine these relationships completely separately for each dependent variable. The results from these procedures are presented in their entirety in Appendix F.

The most dramatic and consistent relationship found was the main effect for sex on GSR maximum height. On the total summation of CI ranks for GSR maximum height from the five polygraph tests, the main effect for sex was highly significant at $\underline{p} = .0002$.

The main effect for sex on GSR maximum height was also significant on four out of the five polygraph tests with $\underline{p} = .02$, .15, .02, .01, and .004 for tests one through five, respectively. On all of these tests, the detection efficiency for females with respect to that dependent variable was consistently lower than that found for males. (See Table 1.)

The only other dependent variable that had a significant main effect for sex was GSR amplitude, which was only significant on the fifth polygraph test ($\underline{p} = .05$). On that test, the CI mean score for females with respect to GSR amplitude was 2.76, whereas the CI mean score for males on that particular dependent variable was 2.45.

Thus the detection efficiency on the fifth tests was higher for males than for females with respect to GSR amplitude.

Table 1.--CI mean ranks for males and females on GSR maximum height for the five polygraph tests.

Sex	Polygraph Test					
	Test One	Test Two	Test Three	Test Four	Test Five	Combined Tests
Males (N = 135)	1.86	1.99	2.19	2.21	2.23	2.10
Females (N = 135)	2.19	2.19	2.49	2.53	2.60	2.40

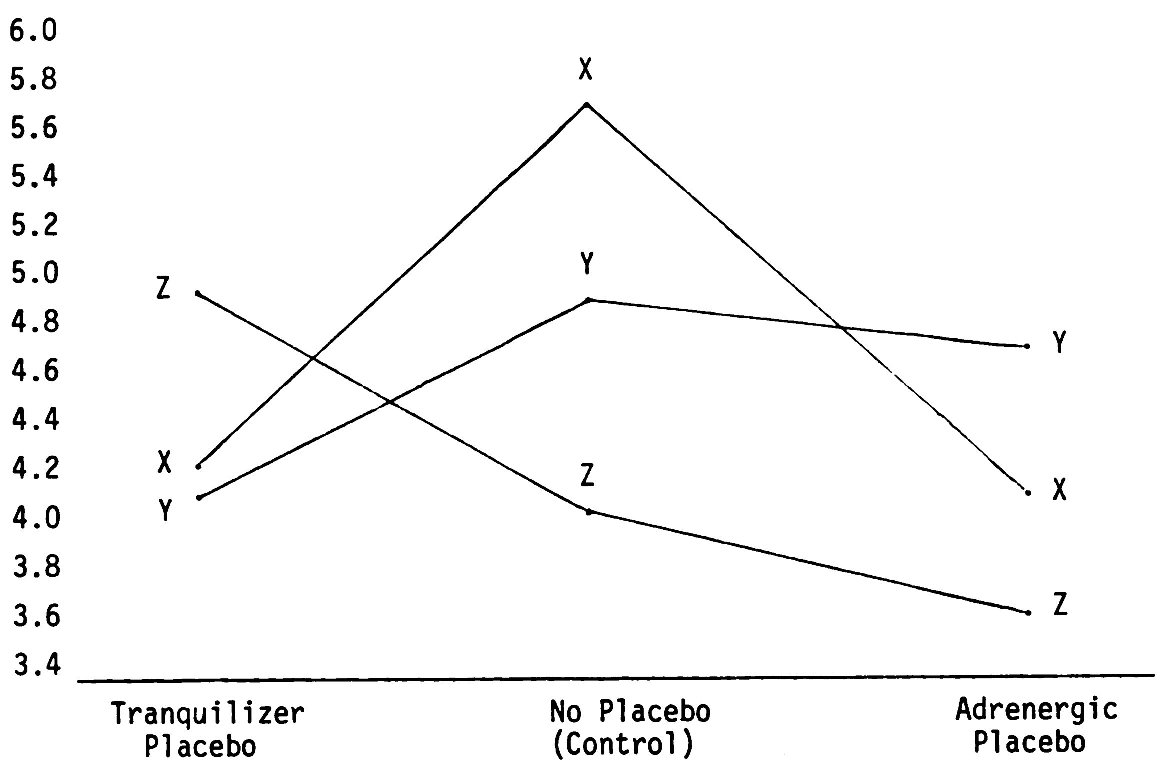
Note: The lower the mean ranks, the higher the detection efficiency.

No significant main effects for feedback were found for any of the dependent variables on any of the five polygraph tests or on the tests conducted on the summed CI ranks of all five polygraph tests for each dependent variable.

The only significant main effects found for the placebo conditions appeared on test five. On that test and only on that test, significant placebo main effects were indicated for both GSR total length ($\underline{p}$ = .04) and GSR amplitude ($\underline{p}$ = .03). The CI mean scores for GSR total length on that polygraph test were 2.49, 2.42, and 2.89 for the adrenergic, control, and tranquilizer placebo conditions, respectively. Once again, the higher the number, the lower the detection efficiency. A Tukey post hoc comparison of those means indicated a significant difference between the placebo control and the tranquilizer placebo conditions. The CI mean scores for GSR amplitude on the fifth polygraph test also suggest a similar pattern, with the following values for the adrenergic, control, and tranquilizer placebo conditions: $\overline{X}$ = 2.58, 2.37, and 2.88, respectively. A Tukey post hoc comparison of the means indicated a significant difference between the placebo control and the tranquilizer placebo conditions.

Two significant drug x feedback interactions were found, both of which were on GSR maximum height. These interaction effects appeared on polygraph tests one and four (p = .02 and .01, respectively). Figure 5 depicts the drug x feedback interaction on the first polygraph test and Figure 6 illustrates that interaction on the fourth polygraph test. Since sex was not an important element in these interaction effects, the two levels of sex were collapsed into their corresponding feedback and placebo conditions, producing a possible range from 2 to 10 on both interactions.

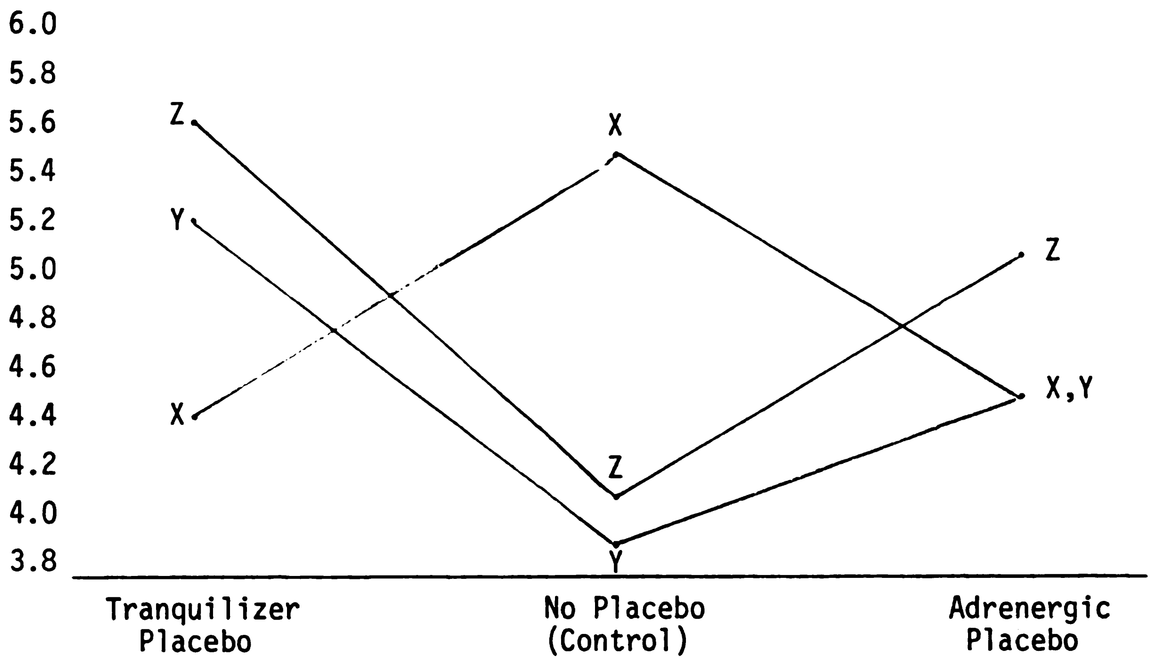
On polygraph test one, the detection efficiency for GSR maximum height on the three feedback-pass conditions was the lowest in the placebo control conditions ($\overline{X}$ = 5.7) and considerably higher in the tranquilizer and adrenergic placebo conditions ($\overline{X}$ = 4.2 and 4.1, respectively). These efficiency levels for the feedback control groups were also lowest in the placebo control condition ($\overline{X}$ = 4.9); however, they were perceptibly higher in the tranquilizer placebo condition ($\overline{X}$ = 4.1) than in the adrenergic placebo condition ($\overline{X}$ = 4.7). The detection efficiency levels for GSR maximum height in the feedback-fail groups were lowest in the tranquilizer placebo groups ($\overline{X}$ = 4.9), surpassed by the placebo control and adrenergic placebo conditions ($\overline{X}$ = 4.0 and 3.6, respectively).



Key: X = Feedback pass condition
Y = Feedback control condition
Z = Feedback fail condition

Note: The higher the mean, the less detection efficiency GSR maximum height had in properly differentiating the critical items. Possible range: 2-10.

Figure 5.--The effect of the three placebo conditions and the three feedback conditions on GSR maximum height during polygraph test one.



Key: X = Feedback pass condition
Y = Feedback control condition
Z = Feedback fail condition

Note: The higher the mean, the less detection efficiency GSR maximum height had in correctly differentiating the critical items. Possible range: 2-10.

Figure 6.--The effect of the three placebo conditions and the three feedback conditions on GSR maximum height during polygraph test four.

On polygraph test four, the drug x feedback interaction for GSR maximum height was even more dramatic. The highest detection efficiency levels for both the feedback control ($\overline{X}$ = 3.9) and feedback fail ($\overline{X}$ = 4.1) groups were found in the placebo control conditions, whereas their lowest efficiency levels appeared in the tranquilizer placebo conditions ($\overline{X}$ = 5.2 and 5.6, respectively), followed by the adrenergic placebo conditions ($\overline{X}$ = 4.5 and 5.1, respectively). Conversely, the detection efficiency levels for GSR maximum height for the feedback pass groups were lowest in the placebo control conditions ($\overline{X}$ = 5.5) and much higher in the tranquilizer and adrenergic groups ($\overline{X}$ = 4.4 and 4.5, respectively).

No significant sex x feedback, sex x placebo, or sex x placebo feedback interaction effects were found for any of the dependent variables on any of the five polygraph tests or on the tests conducted on the summed values for all five polygraph tests for each dependent variable.

Reliability of the Procedures Used to Measure GSR Total Length and Respiration Total Length

As stated in the Apparatus section, the investigator developed a modified map distance measurer to measure GSR total length and respiration total length. The reliability of the measurements obtained using the instrument was determined by randomly selecting 30 polygraph charts from the 270 charts produced in the study. The original measurements of GSR total length and respiration total length for all critical and noncritical items and their corresponding ranks on the selected charts were recorded. These numbers were then masked with black tape, making it impossible to see the original values. A different research assistant than the one who originally measured the responses remeasured the total lengths of the GSR and respiration patterns, following the same procedures as had been used to derive the original measurements. These new measurements were then ranked from one to five (see Method chapter) for each of the five tests included on the polygraph charts.

The reliability of the measurements for the total length of the GSR and respiration responses was calculated by comparing the original measurements with their corresponding values compiled from the second measurement. The absolute value frequencies of the difference between the GSR total length measurement-remeasurements are presented in Table 2. Twenty-three percent of the values were exactly the same and 80.8 percent of the values were ± 1.0 unit (inclusive; 1 unit = 3.896 mm). The mean length of the GSR patterns found on the 30 randomly selected polygraph charts was originally 10.11 units, whereas the mean length computed on the remeasured values was 9.86 units. The Pearson correlation coefficient calculated on the two sets of values was $\underline{r} = .94$.

The same procedure was used to compare the ranks assigned to the GSR total length measurements. The absolute value frequencies of the differences between the GSR total length rank and rerank determinations are presented in Table 3. Approximately 61 percent of the ranks based on the remeasured values corresponded exactly to their original ranks. Almost 87 percent of the corresponding ranks from the two data sets were $\leq |1|$ rank from each other. The Pearson

Table 2.--The absolute value frequency distribution of the differences between the original and subsequent measurements of GSR total length.

$\lvert X_1 - X_2 \rvert$	Percentage of Comparisons	Cumulative Percentage
.0	23.5	23.5
.1	10.2	33.7
.2	5.1	33.8
.3	4.5	43.3
.4	4.8	48.1
.5	9.2	57.3
.6	4.3	61.6
.7	3.5	65.1
.8	3.2	68.3
.9	5.3	73.6
1.0	7.2	80.8
1.1	3.9	84.7
1.2	1.2	85.9
1.3	.7	86.5
1.4	2.3	88.8
1.5	2.8	91.6
1.6-2.0	4.4	96.0
2.1-2.5	1.2	97.2
over 2.5	2.8	100.0

$\bar{X} = .649$; $N = 750$

Table 3.--The absolute value frequency distribution of the differences between the original and subsequent determinations of GSR total length ranks.

$\lvert X_1 - X_2 \rvert$	Percentage of Comparisons	Cumulative Percentage
0.0	60.9	60.9
1.0	26.0	86.9
1.5	.3	87.2
2.0	8.9	96.1
2.5	.3	96.4
3.0	2.8	99.2
4.0	.8	100.0

$\bar{X} = .565$; $N = 750$

correlation coefficient comparing the original ranks of the critical and noncritical items to their corresponding second values was $\underline{r}$ = .73.

The reliability of the respiration total length measurements and their corresponding ranks was calculated in the same manner as used for the GSR total length values. The absolute value frequencies of the differences between the original respiration total length measurements and their corresponding remeasured values are presented in Table 4. Approximately 6.8 percent of the original measurements were exactly the same as their corresponding remeasured values, 54.3 percent were within ± 1.0 unit (inclusive), and 94.9 percent were within ± 3.5 units (inclusive). The average length of the respiration patterns was 26.56 units originally and 26.81 units after they were remeasured. The Pearson correlation coefficient comparing the two data sets was $\underline{r}$ = .95.

The absolute value frequencies of the differences between the respiration total length rank and rerank determinations are presented in Table 5. Almost 62 percent of the ranks based on the remeasured values corresponded exactly to their original ranks and 89.1 percent were within ± 1 rank (inclusive) from each other. The Pearson correlation coefficient comparing the original ranks to their corresponding reranked values was $\underline{r}$ = .78.

Table 4.--The absolute value frequency distribution of the differences between the original and the subsequent respiration measurements.

$\lvert X_1 - X_2 \rvert$	Percentage of Comparisons	Cumulative Percentage
.0	6.8	6.8
.1	6.7	13.5
.2	2.0	15.5
.3	2.5	18.0
.4	3.7	21.7
.5	7.6	29.3
.6	4.3	33.6
.7	1.1	34.7
.8	2.4	37.1
.9	6.8	43.9
1.0	10.4	54.3
1.1	5.2	59.1
1.2	1.3	60.4
1.3	1.9	62.3
1.4	2.1	64.4
1.5	5.1	69.5
1.6-2.0	7.4	76.9
2.1-2.5	11.8	88.7
2.6-3.0	3.7	92.4
3.1-3.5	2.5	94.9
3.6-4.0	2.2	97.1
4.1-4.5	.5	97.6
4.6-5.0	.4	98.0
over 5.0	2.0	100.0

$\overline{X} = 1.35$; $N = 750$

Table 5.--The absolute value frequency distribution of the differences between the original and subsequent determinations of respiration ranks.

$\lvert X_1 - X_2 \rvert$	Percentage of Comparisons	Cumulative Percentage
0.0	61.6	61.6
1.0	27.5	89.1
2.0	8.1	97.2
3.0	2.4	99.6
4.0	.4	100.0

$\bar{X}$ = .525; N = 750

The Effect of the Treatment Conditions on Performance Expectancy Scores

The subjects' mean performance expectancy scores obtained immediately after the "demonstration" tests are reported in Table 6. The means ranged from 2.00 for males in the placebo pass-feedback pass group to 4.0 for females in the placebo fail-feedback fail group. The higher the mean, the less certain the subjects were that they could "beat" three out of the five polygraph tests they were to take after marking their predictions. The males' performance expectancies indicated they were more confident that they could beat three out of the five tests than the females in each of the treatment subgroups. The performance expectancies for the feedback conditions were hierarchically ordered, with the subjects in the feedback pass group the most optimistic in each case, the no feedback group in the middle, and the feedback fail group with the most pessimistic outlook in all cases when controlling for drug condition and sex. The effect of

Table 6.--The effect of sex, three placebo conditions, and three feedback conditions on the mean performance expectancy scores acquired immediately after the "demonstration" test.

Feedback Condition	Sex of Respondent					
	Male			Female		
	Adrenergic Placebo	No Placebo	Tranquilizer Placebo	Adrenergic Placebo	No Placebo	Tranquilizer Placebo
Feedback Pass	2.40	2.13	2.00	2.80	2.60	2.27
No Feedback	2.73	2.60	2.73	3.13	3.10	3.00
Feedback Fail	3.20	3.33	3.07	4.00	3.40	3.60

$N = 270$; $n = 15$; $\overline{X} = 2.89$

Note: The higher the mean, the less certain the subjects were that they would "beat" three out of the five tests.

the drug conditions on performance expectancy was less stable. Generally, subjects taking the adrenergic placebo had the least optimistic predictions; their level of optimism was surpassed by the no placebo and the tranquilizer placebo conditions, respectively. However, this was not the case for males in the no feedback and feedback fail conditions or for females in the feedback fail condition. In those treatment categories, one of the three mean scores for the placebo treatments did not follow the hypothesized order.

The performance expectancy scores the subjects marked immediately after the demonstration test were analyzed using analysis of variance. The results derived from that technique are reported in Table 7. The main effects for both sex and feedback were highly significant (p = .00003 and p = .00001, respectively), whereas the main effect for placebo was not significant (p = .069). A Tukey post hoc comparison of the three means for the feedback conditions indicated that all three groups were significantly different from each other. Hence, males were significantly more certain they would beat three out of the five actual polygraph tests than females, and subjects taking the tranquilizer placebos were significantly more optimistic than members of the control group, who were in turn significantly more optimistic concerning their ability to beat the polygraph than the subjects in the adrenergic group. None of the interaction effects stemming from those conditions was significant.

Table 7.--Analysis of variance: the effect of sex, three placebo conditions, and three feedback conditions on the mean performance expectancy scores acquired immediately after the "demonstration" test.

Source of Variation	Sum of Squares	D.F.	Mean Square	F Value	Signif. of F
Sex	11.20	1	11.20	17.84	.00003
Drug	3.39	2	1.69	2.70	.06949
Feedback	51.23	2	25.61	40.79	.00001
Sex x drug	.54	2	.27	.43	.65066
Sex x feedback	.12	2	.06	.09	.90999
Drug x feedback	1.31	4	.33	.52	.72187
Sex x drug x feedback	1.84	4	.46	.73	.57136
Within cells	158.27	252	.63		

The mean performance expectancy scores obtained from the subjects after they had completed the five actual polygraph tests are reported in Table 8. In this situation, the lower scores indicate that the subjects were more certain they had successfully deceived the examiner on three out of the five polygraph tests. Once again, the females generally had more pessimistic predictions than the males within the levels of the other various treatment conditions. However, the mean scores for females were lower (more optimistic) in the feedback fail-no placebo and no feedback-tranquilizer placebo groups and the same as the mean scores for males in the feedback pass-adrenergic placebo group. The mean scores associated with the various feedback treatments continued the same hierarchical trend, with the feedback pass groups having the scores indicative of the highest degree of optimism and the feedback fail groups generally having the lowest when sex and placebo treatment were controlled.

Table 8.--The effect of sex, three placebo conditions, and three feedback conditions on the mean performance expectancy scores acquired immediately after the actual test.

Feedback Condition	Sex of Respondent					
	Male			Female		
	Adrenergic Placebo	No Placebo	Tranquilizer Placebo	Adrenergic Placebo	No Placebo	Tranquilizer Placebo
Feedback Pass	2.33	2.53	2.20	2.53	2.53	2.53
No Feedback	3.07	2.80	3.00	3.13	3.40	2.60
Feedback Fail	3.27	3.13	3.00	3.80	2.93	3.60

$N = 270$; $n = 15$; $\overline{X} = 2.91$

Note: The higher the mean, the less certain the subjects were that they "beat" three out of the five tests.

The only exceptions were the tie between the no feedback-tranquilizer placebo and the feedback fail-tranquilizer placebo groups for males and the lower mean score ($\overline{X}$ = 2.93) for females in the feedback fail-no placebo group compared to their feedback fail-no placebo condition ($\overline{X}$ = 3.40). Although there was a trend for the mean scores in the adrenergic placebo groups to represent less optimistic predictions than those in the other two placebo groups and the tranquilizer groups to be the most optimistic when controlling for sex and feedback conditions, there were several exceptions to this order.

Once again, the performance expectancy scores the subjects marked after taking the five actual tests were analyzed using analysis of variance. The results of that procedure are reported in Table 9. This time only the main effect for feedback was significant ($\underline{p}$ = .00001). A Tukey post hoc comparison of the feedback means indicated that the subjects in the feedback pass group were more optimistic about the outcome of the polygraph test than subjects in the other two feedback conditions. The F value for the main effect for sex and the sex x drug x feedback interaction were the only other sources of variation that approached significance ($\underline{p}$ = .086 and $\underline{p}$ = .088, respectively).

Table 10 depicts the net change in the subjects' mean predictions from how well they thought they would do after taking the "demonstration" test to how well they thought they had done after taking the five polygraph tests. The negative numbers indicate that the subjects thought their chances of "beating" three out of the five

Table 9.--Analysis of variance: the effect of sex, three placebo conditions, and three feedback conditions on the mean performance expectancy scores provided by the subjects immediately after the actual test.

Source of Variation	Sum of Squares	D.F.	Mean Square	F Value	Signif. of F
Sex	2.50	1	2.50	2.97	.08612
Drug	1.87	2	.93	1.11	.33227
Feedback	33.16	2	16.58	19.66	.00001
Sex x drug	.21	2	.10	.12	.88435
Sex x feedback	.56	2	.28	.33	.71655
Drug x feedback	4.11	4	1.03	1.22	.30335
Sex x drug x feedback	6.93	4	1.73	2.05	.08753
Within cells	212.53	252	.84		

tests had improved, whereas the positive numbers indicate their second prediction was less optimistic. Overall, the females' scores tended to decrease, indicating they thought their chances improved, whereas the males' scores increased. The changes in the two predictions represented an interesting pattern when analyzed by feedback conditions. All of the feedback fail groups' differences were negative or zero, indicating that they thought their chances of beating three out of the five tests had either remained the same as when they completed their first prediction or had improved after taking the actual tests. Conversely, all of the no feedback groups' differences were positive or zero with the sole exception of the no feedback-tranquilizer placebo for females, which decreased by .40. The differences in the feedback pass groups suggest an interaction with sex, since the females' scores were all negative, whereas all of the group differences for males were positive, excluding the

Table 10.--Net changes in performance expectancy mean scores comparing predictions made immediately before and after the actual polygraph tests for all three placebo conditions and for both sexes.

Feedback Condition	Sex of Respondent					
	Male			Female		
	Adrenergic Placebo	No Placebo	Tranquilizer Placebo	Adrenergic Placebo	No Placebo	Tranquilizer Placebo
Feedback Pass	-.07	+.40	+.20	-.27	-.07	-.26
No Feedback	+.34	+.20	+.27	0	+.30	-.40
Feedback Fail	-.07	-.20	-.07	-.20	-.47	0

N = 270; n = 15; $\overline{X}$ = -.056

Note: The negative numbers indicate that the subjects thought their chances of "beating" three out of the five tests had improved, whereas the positive numbers indicate their second predictions were less optimistic.

feedback pass-adrenergic placebo category, which resulted in a slight decrease ($X_1 - X_2 = -.07$).

Correlations Between the Dependent Variables

The correlations between the dependent variables for each of the five polygraph tests are reported in Table 11. In each case, the ranks of the critical item on the two specified dependent measures were compared to each other for the 270 subjects. A high correlation between the two dependent variables indicated that those two ranks on the critical item for each subject tended to be the same or very close to it. A high correlation also suggested that the discrimination value of the two variables might be limited, since to a large extent they would both be accounting for the same variance.

Table 11.--Pearson correlation coefficients comparing the ranks of the dependent variables on the critical items for each polygraph test.

Dependent Variables Correlated	Test One	Test Two	Test Three	Test Four	Test Five
Respiration x GSR maximum height	.096	.049	.039	-.074	.019
Respiration x GSR total length	.024	.028	-.076	-.014	-.005
Respiration x GSR amplitude	.023	.046	.057	-.006	-.035
GSR total length x GSR maximum height	.151	.295	.382	.377	.487
GSR maximum height x GSR amplitude	.333	.362	.382	.367	.460
GSR total length x GSR amplitude	.680	.709	.718	.784	.734

The correlation coefficients between respiration and all other dependent measures were very low for all five polygraph tests. They ranged from a maximum positive value of .096 to a maximum negative value of -.074. GSR maximum height was fairly highly correlated with both GSR total length and GSR maximum height, with correlations ranging from .151 for the GSR total length x GSR maximum height comparison on test one to .487 for those same dependent variables on test five. The variables showing the highest degree of correlation were GSR total length and GSR amplitude. Their coefficients from test one to test five were .680, .709, .718, .784, and .734, respectively.

The correlations for the ranks of the critical items between the different polygraph tests for each of the four dependent variables are reported in Table 12. A high correlation between two tests for a certain dependent variable would indicate that the subjects tended to have the same rank for the critical items on both tests, with respect to that particular physiological variable. The correlation coefficients for the various polygraph test combinations across all of the dependent measures were relatively low. The lowest correlation was .018 for the test one x test four comparison using GSR amplitude as the dependent variable, whereas the highest correlation was only .371 for the same comparison using GSR maximum height as the dependent variable. GSR maximum height tended to have the highest correlations from test to test, followed by respiration, GSR amplitude, and GSR total length, respectively.

Table 12.--Pearson correlation coefficients for the ranks of the critical items comparing the different polygraph tests on each dependent measure.

Polygraph Tests Correlated	Respiration	GSR Amplitude	GSR Maximum Height	GSR Total Length
Test one x test two	.264	.073	.290	.049
Test one x test three	.144	.074	.184	.091
Test one x test four	.175	.018	.371	.041
Test one x test five	.111	.076	.218	.060
Test two x test three	.217	.140	.136	.170
Test two x test four	.196	.113	.168	.138
Test two x test five	.074	.103	.192	.069
Test three x test four	.142	.153	.243	.176
Test three x test five	.223	.185	.181	.127
Test four x test five	.048	.249	.282	.232

The Incidence of GSR Maximum Height Downward Drift and GSR Amplitude Falling Patterns

Table 13 depicts the number and percentage of GSR maximum height downward drift patterns that were found on each of the polygraph tests. Those figures are also presented for the "demonstration" test; however, only the three feedback control conditions were used in deriving the number and percentage of GSR maximum height downward drift patterns for that particular test. The various feedback pass and feedback fail conditions were excluded from these calculations since the experimenter intentionally manipulated their respective charts to yield the desired feedback effects.

The percentage of subjects producing GSR maximum height downward drift patterns remained relatively constant for each of the polygraph tests. The percentages ranged from 23.0 percent on the second test to 37.0 percent on the fourth test. Although the percentage of charts containing GSR maximum height downward drift patterns tended to increase with each successive polygraph test, this relationship was not consistent enough to constitute a strong trend.

Table 13.--The frequency of GSR maximum height downward drift patterns on the demonstration and five actual polygraph tests.

Test	Absolute Frequency	Relative Frequency (Percentage)
Demonstration test[a] (N = 90)	25	27.8
Test one (N = 270)	83	30.7
Test two (N = 270)	62	23.0
Test three (N = 270)	85	31.5
Test four (N = 270)	100	37.0
Test five (N = 270)	98	36.3

[a]The calculations for the demonstration test were based only on charts produced by the 90 subjects in the feedback control conditions.

Table 14 shows the percentage of GSR amplitude falling patterns that were present for the five items on each of the polygraph tests. Once again, those figures presented for the demonstration test were based solely on the charts produced by subjects in the three feedback control conditions. Excluding the demonstration test, the frequency of GSR amplitude falling responses increased with each subsequent polygraph test. The percentage of GSR amplitude falling patterns ranged from 8.6 percent on the first test to 23.26 percent on the fifth test.

Table 14.--The percentage of all GSR amplitude responses that were categorized as falling patterns on each polygraph test.

Polygraph Test	Percentage of Responses Categorized as Falling Patterns
Demonstration[a]	9.34
Test one	8.64
Test two	13.92
Test three	17.12
Test four	22.44
Test five	23.26

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

The percentage of GSR amplitude falling patterns occurring on the critical items for each of the polygraph tests is presented in Table 15. Once again, the percentages increased for each subsequent polygraph test. The percentage of critical items constituting a GSR amplitude falling pattern ranged from 0 percent on the demonstration test to 17.4 percent on the fifth polygraph test. These percentages, which were based solely on the critical items, were perceptibly lower than those presented in Table 14 for all five items on each polygraph test.

If there was no difference in the incidence of GSR amplitude falling patterns between the critical and noncritical items, one would expect 20 percent of the total number of falling patterns to have occurred on the critical items. The 20 percent figure represents the 1:5 ratio between the number of critical items to the total number of questions. However, significantly less than 20 percent of the total number of falling patterns were associated with the critical items ($\underline{z}$ = -9.21). This demonstrates that the subjects were less likely to have GSR amplitude falling patterns on critical items than on noncritical items during the polygraph tests.

Table 15.--The percentage of all critical items that were categorized as GSR amplitude falling patterns on each test.

Polygraph Test	Percentage of Critical Items Categorized as Falling Patterns
Demonstration[a]	0
Test one	2.6
Test two	4.1
Test three	11.1
Test four	13.3
Test five	17.4

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

The Accuracy of the Different Physiological Indices in Differentiating Between Critical and Noncritical Items

Table 16 shows the percentage of critical items for each polygraph test that were ranked "one" (the response most indicative of deception) with respect to respiration, GSR amplitude, GSR maximum height, and GSR total length. Since there were five items on each test, the chance expectancy that the critical item would be ranked "one" was 20 percent for each of the physiological indices monitored. All of the percentages, regardless of the test or physiological

Table 16.--The percentage of critical items ranked "one" (the most indicative of deception) with respect to respiration, GSR amplitude, GSR maximum height, and GSR total length for each polygraph test.

Polygraph Test	Physiological Index			
	Respiration	GSR Amplitude	GSR Maximum Height	GSR Total Length
Demonstration[a]	41.1	60.0	57.8	51.1
Test one	45.2	46.3	40.0	37.8
Test two	48.1	53.7	46.3	43.3
Test three	41.9	29.6	31.5	28.5
Test four	52.2	43.0	33.0	37.0
Test five	40.0	26.7	27.0	28.9

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

measure on which they were based, were significantly more accurate than this chance level ($\underline{z} \geq 2.74$). Excluding the demonstration test, respiration was generally the best or one of the best physiological indices in discriminating between the critical and noncritical items. However, on the demonstration test (based solely on the 90 subjects in the three feedback control conditions), the three GSR indices had higher detection efficiency levels than respiration. The same relationships are demonstrated in Table 17, which depicts the mean rank of the critical item with respect to the aforementioned four dependent variables on each of the polygraph tests. The chance level for each of the dependent variable mean ranks on all of the polygraph tests was $\bar{X} = 3.0$, $\underline{SD} = 1.4$ (based on their probability distributions); that mean is a significantly higher value than any of the actual mean rank scores attained ($|\underline{z}| \geq 4.49$). Thus the polygraph had a significantly higher detection efficiency than chance for each physiological parameter on all of the polygraph tests with regard to both the percentage of responses to the critical items that were scored as the most indicative of deception and for the mean score of the dependent variable ranks.

Table 17.--The mean rank of the critical items with respect to respiration, GSR amplitude, GSR maximum height, and GSR total length for each polygraph test.

Polygraph Test	Physiological Index			
	Respiration	GSR Amplitude	GSR Maximum Height	GSR Total Length
Demonstration[a]	2.33	1.76	1.79	1.97
Test one	2.18	2.12	2.23	2.30
Test two	2.10	1.86	2.09	2.00
Test three	2.14	2.45	2.34	2.49
Test four	1.95	2.11	2.37	2.31
Test five	2.25	2.61	2.42	2.60

Note: The smaller the mean, the higher the detection efficiency the variables had in identifying the critical items.

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

One of the factors reducing the detection efficiency of the three GSR indices was the lack of electrodermal responsiveness demonstrated by certain subjects. To control for this phenomenon, the GSR maximum height accuracy percentages and means were recalculated, excluding tests in which the subjects produced a GSR maximum height downward drift pattern. Similarly, these values for GSR amplitude and GSR total length were recalculated excluding charts in which all five of the responses on a particular test were GSR amplitude falling patterns. Since no related problems were associated with the respiration patterns, there was no need to recalculate their respective values. Table 18 shows the recalculated percentages of critical items that were ranked "one" for the four dependent variables, taking into account the exclusions mentioned above. Table 19 depicts the recalculated mean-ranks of the critical items for the polygraph tests with respect to each of the dependent variables.

When the tests containing a GSR maximum height downward drift pattern and/or five GSR amplitude falling patterns were excluded from the two measures of detected efficiency, GSR maximum height became the most accurate index of deception, as demonstrated by the highest percentage of critical items ranked "one" and the lowest critical item mean ranks for each polygraph test. Respiration, GSR amplitude, and GSR total length were fairly equivalent in their ability to discriminate between critical and noncritical items, when the charts with all five responses constituting GSR amplitude falling patterns were excluded from the accuracy calculations for GSR amplitude and GSR total length. However, of these three dependent variables GSR amplitude was the most valid measure on the demonstration and first two actual polygraph tests, whereas respiration was the more accurate on the last three polygraph tests.

Table 18.--The percentage of critical items ranked as the most indicative of deception with respect to the four principal dependent measures for each polygraph test.

Polygraph Test	Physiological Index			
	Respiration	GSR Amplitude	GSR Maximum Height	GSR Total Length
Demonstration[a]	41.1 (n=90)	60.0 (n=90)	80.0 (n=65)	51.1 (n=90)
Test one	45.2 (n=270)	47.4 (n=264)	57.8 (n=187)	38.6 (n=264)
Test two	48.1 (n=270)	54.9 (n=264)	60.1 (n=208)	44.3 (n=264)
Test three	41.9 (n=270)	31.4 (n=255)	46.0 (n=185)	30.2 (n=255)
Test four	52.2 (n=270)	45.1 (n=257)	52.4 (n=170)	38.9 (n=257)
Test five	40.0 (n=270)	28.4 (n=254)	42.4 (n=172)	30.7 (n=254)

Note: These figures were calculated excluding tests containing GSR maximum height downward drift patterns in calculating the values associated with GSR maximum height and excluding tests containing GSR amplitude falling patterns on all five responses on a given test in deriving the values associated with GSR amplitude and GSR total length.

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

Detection Rates Attained Using the Scoring Procedure Developed by Lykken (1959)

The accuracy of the polygraph in this experiment was also analyzed using the scoring procedure developed by Lykken (1959). If the dependent variable associated with the critical item was ranked "one" (most indicative of deception), it was given a score of 2 on

Table 19.--The mean rank of the critical items with respect to the four principal dependent variables for each polygraph test.

Polygraph Test	Physiological Index: Respiration	GSR Ampli-tude	GSR Maximum Height	GSR Total Length
Demonstration[a]	2.33 (n=90)	1.76 (n=90)	1.32 (n=65)	1.97 (n=90)
Test one	2.18 (n=270)	2.10 (n=264)	1.89 (n=187)	2.29 (n=264)
Test two	2.10 (n=270)	1.83 (n=264)	1.81 (n=208)	1.97 (n=264)
Test three	2.14 (n=270)	2.42 (n=255)	2.03 (n=185)	2.46 (n=255)
Test four	1.95 (n=270)	2.07 (n=257)	2.01 (n=170)	2.27 (n=257)
Test five	2.25 (n=270)	2.58 (n=254)	2.08 (n=172)	2.57 (n=254)

Note: These figures were calculated excluding the tests containing GSR maximum height downward drift patterns in calculating the values associated with GSR maximum height and excluding charts containing GSR amplitude falling patterns on all five responses on a given test in deriving the values associated with GSR amplitude and GSR total length.

[a]The calculations pertaining to the demonstration test were based only on the charts produced by the 90 subjects in the feedback control conditions.

that test. If the dependent variable associated with the critical item was ranked "two," it was given a score of 1. Thus by summing the scores on the five polygraph tests a perfect guilty score for each of the dependent variables was 10.

Since none of the subjects in this study was innocent, it was impossible to make a direct comparison between the actual scores of innocent and guilty subjects. However, it was possible to calculate the theoretical distribution to estimate the expected proportions of innocent subjects that would have achieved each of the various scores. For example, the probability that an innocent subject would have received a score of 10 would be $(.2)^5$ assuming that it was equally likely that the subject's largest response would have been to the critical item as it was to any of the four noncritical items on each of the five tests. Thus one would expect .032 percent of all innocent subjects to have a score of 10 if an infinite number of innocent subjects was tested.

The estimated proportions (probability distribution) of innocent subjects that would have obtained each of the scores possible in this study are presented in Table 20. The population mean of these scores based on their probability distribution is 3, with a standard deviation of 1.789. As indicated in Table 20, if a cut-off point of scores 5 or greater was selected as values indicative of guilt, theoretically approximately 20 percent of the innocent subjects would have been misclassified as guilty. It should be noted that the percentages depicted in Table 20 are actually probability values. Thus if innocent subjects had been included in this study, the actual percentage of innocent subjects misclassified would probably be either slightly higher or lower than indicated by the table. However, the more innocent subjects tested the less deviation should occur between the actual percentage of misclassification and the estimated values given the aforementioned chance model.

Table 20.--The estimated proportion (probability distribution) of innocent subjects attaining each of the possible scores for the testing model incorporated into this study using the Lykken (1959) scoring procedure.

Score	Estimated Relative Frequency (Percent)	Cumulative Relative Frequency (Percent)
10	.032	.032
9	.160	.192
8	.800	.992
7	2.240	3.232
6	5.920	9.152
5	10.592	19.744
4	17.760	37.504
3	20.160	57.664
2	21.600	79.264
1	12.960	92.224
0	7.776	100.000

The subjects' GSR amplitude values were scored using the aforementioned scoring procedure. The frequency distribution of those scores is presented in Table 21. The mean score for subjects with respect to that dependent variable was 5.07, with a standard deviation of 2.38. As indicated by Table 21, 57.8 percent of the subjects had a score of 5 or greater. Since several of the subjects demonstrated a low degree of electrodermal responsiveness on some of the tests, the frequency distribution was recalculated excluding subjects that had three or more GSR amplitude falling patterns on three or more of the five polygraph tests. This procedure eliminated 33 out of the 270 subjects included in this study. The frequency distribution of scores for the remaining 237 subjects is presented in Table 22. The mean score for those subjects was 5.25, with a standard deviation of 2.31. A comparison between the original mean with

Table 21.--The actual proportion of subjects attaining each of the possible scores derived from scoring the subjects' GSR amplitude values using the Lykken (1959) procedure.

Score	Absolute Frequency (n)	Relative Frequency (Percent)	Cumulative Relative Frequency (Percent)[a]
10	7	2.6	2.6
9	12	4.4	7.0
8	30	11.1	18.1
7	31	11.5	29.6
6	31	11.5	41.1
5	45	16.7	57.8
4	50	18.5	76.3
3	25	9.3	85.6
2	18	6.7	92.3
1	11	4.1	96.4
0	10	3.7	100.1

[a]Cumulative relative frequency column totals more than 100 percent due to rounding.

Table 22.--The actual proportion of subjects attaining each of the possible scores derived from scoring the subjects' GSR amplitude values using the Lykken (1959) procedure excluding subjects that had three or more GSR amplitude falling patterns on three or more of the five polygraph tests.

Score	Absolute Frequency (n)	Relative Frequency (Percent)	Cumulative Relative Frequency (Percent)[a]
10	7	3.0	3.0
9	11	4.6	7.6
8	27	11.4	19.0
7	31	13.1	32.1
6	28	11.8	43.9
5	40	16.9	60.8
4	42	17.7	78.5
3	22	9.3	87.8
2	16	6.8	94.6
1	8	3.4	98.0
0	5	2.1	100.1

[a]Cumulative relative frequency column totals more than 100 percent due to rounding.

respect to GSR amplitude and the recalculated mean after the 33 subjects with a low degree of electrodermal responsiveness were eliminated indicated that the two means were not significantly different ($|\underline{z}|$ = .87). Table 22 shows that 60.8 percent of the remaining 237 subjects had scores of 5 or greater when their GSR amplitude responses were scored using the Lykken (1959) procedure.

The ranks of the subjects' respiration responses to the critical items were also scored using the aforementioned scoring procedures. The frequency distribution for those scores is presented in Table 23. The mean score with respect to respiration was 5.58, with a standard deviation of 2.44. The mean score for respiration was significantly greater than the original mean score for GSR amplitude, but not significantly greater than the mean score for GSR amplitude with the 33 low responsiveness subjects eliminated ($\underline{z}$ = 2.46 and 1.56, respectively). Table 23 shows that 68.9 percent of the subjects had scores of 5 or greater with respect to respiration.

The ranks of the subjects' responses to each of the questions with respect to GSR amplitude, GSR maximum height, and respiration were added together for each of the five polygraph tests. These composite values were then ranked from 1 to 5 (a rank of 1 having the smallest composite value) for every item in each of the five polygraph tests. The rank of the critical item based on the composite values was then scored using the Lykken (1959) procedure. The frequency distribution of those scores is presented in Table 24. The mean score with respect to the aforementioned composite values was 6.64, with a standard deviation of 2.24. Of the 270 subjects included in this

Table 23.--The actual proportion of subjects attaining each of the possible scores derived from scoring the subjects' respiration values using the Lykken (1959) procedure.

Score	Absolute Frequency (n)	Relative Frequency (Percent)	Cumulative Relative Frequency (Percent)
10	14	5.2	5.2
9	17	6.3	11.5
8	33	12.2	23.7
7	37	13.7	37.4
6	41	15.2	52.6
5	44	16.3	68.9
4	28	10.4	79.3
3	22	8.1	87.4
2	18	6.7	94.1
1	10	3.7	97.8
0	6	2.2	100.0

Table 24.--The actual proportion of subjects attaining each of the possible scores derived from scoring the subjects' composite values with respect to GSR amplitude, GSR maximum height, and respiration using the Lykken (1959) procedure.

Score	Absolute Frequency (n)	Relative Frequency (Percent)	Cumulative Relative Frequency (Percent)[a]
10	25	9.3	9.3
9	35	13.0	22.3
8	46	17.0	39.3
7	48	17.8	57.1
6	39	14.4	71.5
5	28	10.4	81.9
4	23	8.5	90.4
3	15	5.6	96.0
2	4	1.5	97.5
1	4	1.5	99.0
0	3	1.1	100.1

[a]Cumulative relative frequency column totals more than 100 percent due to rounding.

study, 81.9 percent had a score of 5 or more based on the composite value ranks of the five critical items.

Each of the four means (GSR amplitude, GSR amplitude with nonreactors eliminated, respiration, and rank of the composite values) calculated using the Lykken (1959) scoring procedure was significantly higher than the population mean based on the probability distribution of scores for innocent subjects (each $\underline{z} > 19$). Therefore, the polygraph testing procedure used in this study detected the deceptive responses made by the subjects significantly more frequently than chance expectancy levels.

A separate three-factor analysis of variance was conducted to examine the effects of sex, placebo condition, feedback condition, and all of their possible interactions on each of the four sets of scores presented above. The results derived from these statistical tests are presented in Appendix G. No significant main or interaction effects were found for any of the four dependent measures that were scored using the aforementioned scoring procedure. Hence, none of the principal independent variables examined in this study had a significant effect on polygraph detection efficiency when the Lykken (1959) scoring procedure was employed.

The Results Derived From the Biographical Data Sheet and the Follow-Up Questionnaire

The results derived from the biographical data sheet (Appendix D) and the follow-up questionnaire (Appendix E) are presented in this section. To convey this material in a systematic and comprehensible manner, the findings are broken down into several subsections.

Each subsection commences with a report of the relative frequencies of subjects falling into the various categories described under that particular subheading. For example, one of the questions the subjects were asked related to their church attendance during the year preceding their involvement in the experiment. When that topic is presented, the various categories of church attendance on which the data were compiled (i.e., 0 times, 1-5 times, etc.) and the number of subjects falling into those particular categories are depicted. After providing that descriptive information, the subsection focuses on the relationship between the variable being discussed and three dependent variables: (1) a refined measure of detection efficiency, which is referred to as the sum of critical item composite ranks (SCICR); (2) the frequency of GSR amplitude falling patterns; and (3) the frequency of downward drift patterns with respect to GSR maximum height.

The sum of critical item composite ranks dependent measure was designed to take into account some of the findings already noted in this section. To compensate for GSR maximum height downward drift patterns and/or when all five responses on a given polygraph test were scored as GSR amplitude falling patterns (in both situations, all five responses would have been assigned ranks of three for those dependent measures), the ranks of the dependent measures on each item were added together. The sums of the original ranks for the five items on each test were then ranked from one (smallest composite) to five (largest composite). Finally, the new ranks associated with critical items on each of the five polygraph tests were added together, yielding a possible range from 5 to 25. A sum of critical item composite rank value of five would indicate that the critical item was correctly differentiated from the noncritical item on each of the five polygraph tests.

Only three of the four dependent measures derived from the polygraph charts were used in compiling the sum of critical item composite rank values. The value for GSR total length was eliminated from these calculations because it was highly correlated with GSR amplitude and it provided little discriminatory value. The latter point was demonstrated when discriminant analysis was conducted to determine the relative contribution of each of the four dependent variables in discriminating between the critical and noncritical items. The values of the Beta weights associated with GSR total length were negligible for all five polygraph tests, indicating that it was not necessary to incorporate it as a major dependent variable. However, the Beta weights for the other three dependent measures were much higher and relatively equivalent, suggesting that weighting them equally was acceptable.

Thus the first step in determining the sum of critical item composite rank value involved adding the relative ranks for respiration, GSR maximum height, and GSR amplitude together for each of the five items on the polygraph tests. Next, the five sums for each polygraph test were ranked. Finally, the new ranks assigned to the five critical items were added together. For example, if there were absolutely no positive GSR responses on a given test, all five ranks with respect to both GSR maximum height and GSR amplitude would have been originally ranked as three. Using this new system, the two ranks of three for each of the five items would cancel each other out and the sole determinant of the critical item's relative rank on that particular test would be respiration. However, if no GSR maximum height downward drift or GSR amplitude falling patterns occurred, all three measures would be weighted equally in computing the rank of the critical item on that particular test.

The other two dependent measures discussed in each of the remaining subsections are the frequency of GSR amplitude falling patterns and GSR maximum height downward drift patterns, which were both defined in the final section of the methodology chapter. These additional dependent measures are examined in each of the subsections for two reasons. First, since GSR maximum height is in effect excluded from the sum of critical item composite rank calculations on a test when downward drifting occurs and the same applies for GSR amplitude when falling patterns occur on all five responses for a given test, it is important to note the prevalence of these phenomena when considering the sum of critical item composite rank values. Second, because of the abnormally high frequency of subjects not showing any positive response on the GSR measures to many of the questions, it was of major importance to determine why this lack of electrodermal responsiveness occurred.

To accomplish the aforementioned tasks, three one-way ANOVAs were calculated, comparing the means of the different categories within each subsection with respect to each of the dependent variables. For example, in the subsection pertaining to church attendance, the subjects are classified as having either low, medium, or high church attendance. Then the three means for these groups are compared to each other with respect to each of the three dependent variables. This should indicate which of the variables derived from the biographical data sheet and the follow-up questionnaire appear to be related to the accuracy of the polygraph decisions and/or the lack of electrodermal responsiveness.

Age

The subjects ranged in age from 17 to 42 ($\overline{X}$ = 19.8, SD = 2.54). To facilitate analyzing the data, the age groupings of the subjects were collapsed into the following three categories: (1) ages 17 to 18, (2) ages 19 to 21, and (3) ages 22 to 42. Eighty-nine subjects fell into the 17 to 18 age group, 152 subjects into the 19 to 21 age category, and 29 into the last group. Table 25 depicts the means of those three age groups with respect to the sum of the critical item composite ranks, GSR maximum height downward drift, and GSR amplitude falling patterns. The table also reports the level of significance derived from the one-way analysis of variance calculations, which examined differences among the three age group means for each dependent variable. None of the three ANOVAs indicated there were any significant differences among the age group means for the three respective dependent variables.

Table 25.--A comparison of means for three age categories with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Age Categories: 17-18 Years (n=89)	19-21 Years (n=152)	22-42 Years (n=29)	One-Way ANOVA Results
Sum of critical item composite ranks[a]	$\bar{X}$=9.96	$\bar{X}$=9.16	$\bar{X}$=9.41	$F(2,267)=1.74$ $\underline{p}=.18$
GSR maximum height downward drift patterns[b]	$\bar{X}$=1.54	$\bar{X}$=1.70	$\bar{X}$=1.10	$F(2,267)=1.74$ $\underline{p}=.18$
GSR amplitude falling patterns[c]	$\bar{X}$=4.16	$\bar{X}$=4.51	$\bar{X}$=3.34	$F(2,267)=.58$ $\underline{p}=.56$

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Sex

As previously mentioned, 135 male and 135 female subjects participated in this study. Table 26 shows the differences between the means for males and females with regard to the sum of the critical item composite scores and the two measures of electrodermal responsiveness, as well as indicating their level of significance. It is interesting that there was not a significant difference between males and females when their sum of critical item composite scores were compared ($\underline{p}$ = .44). However, females had significantly more

GSR maximum height downward drift and GSR amplitude falling patterns than did males ($p < .0001$).

Table 26.--A comparison of means for males and females with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Sex Categories		Z-Test Results
	Male (n=135)	Female (n=135)	
Sum of critical item composite ranks[a]	$\bar{X}$=9.30	$\bar{X}$=9.60	z=-.773 p=.44
GSR maximum height downward drift patterns[b]	$\bar{X}$=1.01	$\bar{X}$=2.16	z=-6.23 p=.0000
GSR amplitude falling patterns[c]	$\bar{X}$=2.47	$\bar{X}$=6.07	z=-5.77 p=.0000

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Immediate Family Size

During the pretest interview, the subjects were asked how many individuals, including themselves, were in their immediate family. Their responses were collapsed into the following categories: 2 to 4 members, 5 to 7 members, and immediate families with more than 7 members. Table 27 presents the means of these categories with respect to the sum of the critical item composite score values and the numbers of both GSR maximum height downward drift and GSR amplitude falling patterns present on the polygraph tests for these respective groups. The table also shows the significance levels attained when the immediate family size category means were compared to each other for each of these three dependent variables. Although none of the means were significantly different on any of the three dependent variables, they did suggest an interesting pattern. Both the detection efficiency, as measured by sum of critical item composite ranks, and electrodermal responsivemess decreased as family size increased.

Table 27.--A comparison of means for three categories of family size with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Immediate Family Size: 2 to 4 Members (n=68)	5 to 7 Members (n=156)	Over 7 Members (n=46)	One-Way ANOVA Results
Sum of critical item composite ranks[a]	$\overline{X}$=9.92	$\overline{X}$=9.48	$\overline{X}$=8.66	F(2,267)=2.17 $\underline{p}$=.12
GSR max. height downward drift patterns[b]	$\overline{X}$=1.35	$\overline{X}$=1.60	$\overline{X}$=1.89	F(2,267)=1.53 $\underline{p}$=.22
GSR amplitude falling patterns[c]	$\overline{X}$=3.75	$\overline{X}$=4.21	$\overline{X}$=5.21	F(2,267)=1.01 $\underline{p}$=.37

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Combined Family Income

During their pretest interview, the subjects were also asked to indicate the combined income of their parents. For analysis purposes, responses were placed in one of the following categories: (1) less than \$15,000, (2) \$15,000-\$24,999, and (3) over \$24,999. The numbers of subjects falling into each of these categories were 34, 92, and 142, respectively. Table 28 compares the group means for these categories with respect to the sum of the critical item rank composite scores, GSR maximum height downward drift, and GSR amplitude falling patterns. None of the combined family income group means were significantly different from each other for any of the three dependent variables.

Table 28.--A comparison of means for three categories of combined family income with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Combined Family Income			One-Way ANOVA Results
	Less Than \$15,000 (n=34)	\$15,000-24,999 (n=92)	More Than \$24,999 (n=142)	
Sum of critical item composite ranks[a]	$\overline{X}$=9.34	$\overline{X}$=9.30	$\overline{X}$=9.51	F(2,265)=.12 <u>p</u>=.88
GSR max. height downward drift patterns[b]	$\overline{X}$=1.41	$\overline{X}$=1.53	$\overline{X}$=1.68	F(2,265)=.50 <u>p</u>=.61
GSR amplitude falling patterns[c]	$\overline{X}$=3.56	$\overline{X}$=3.85	$\overline{X}$=4.73	F(2,265)=1.07 <u>p</u>=.35

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Subject's Year in School

One hundred sixty-four subjects participating in this study were either freshmen or sophomores; 106 of the subjects had attained junior status or higher. A Z-test was conducted for each of the three dependent variables (sum of critical item composite ranks, GSR maximum height downward drift, and GSR amplitude falling patterns) to determine if the means for the two groups were significantly different. The means and their respective significance levels for those Z-tests are presented in Table 29. None of the three Z-tests indicated there were any significant differences between these two categories of subjects.

Table 29.--A comparison of group means for underclassmen and upperclassmen with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Subject's Year in School: Freshmen and Sophomores (n=164)	Subject's Year in School: Juniors or Above (n=106)	Z-Test Results
Sum of critical item composite ranks[a]	$\bar{X}$=9.67	$\bar{X}$=9.11	z=1.11 p=.27
GSR max. height downward drift patterns[b]	$\bar{X}$=1.61	$\bar{X}$=1.55	z=.34 p=.73
GSR amplitude falling patterns[c]	$\bar{X}$=4.27	$\bar{X}$=4.25	z=.04 p=.97

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Subject's Grade Point Average

Table 30 shows the levels of significance derived from three one-way ANOVAs, which compared different grade point average categories with respect to their means on the sum of critical item composite ranks, GSR maximum height downward drift, and GSR amplitude falling patterns. The grade point average categories selected were: less than 2.8, 2.8 to 3.3, and over 3.3. All of the subjects' grade point averages were based on a four-point scale (A = 4.0). Fifty-four subjects had a grade point average less than 2.8, 144 subjects' grade point averages were between 2.8 and 3.3, and 72 were 3.4 or above. The three ANOVAs indicated there were no significant differences among the subjects in the different grade point average categories for any of the three dependent variables.

Subject's Religious Preference

Each of the subjects was asked to indicate his/her religious preference. Fifty-two subjects responded none, 13 replied Jewish, 106 stated Catholic, 98 said Protestant, and one declined to state a religious preference. Three one-way ANOVAs were conducted to determine if there were any differences among the four religious preference means with respect to the sum of the critical item composite ranks, GSR maximum height downward drift, and GSR amplitude falling patterns. The different means and their respective levels of significance are presented in Table 31. Although none of the aforementioned mean comparisons indicated significant differences, the mean for the Jewish subjects on GSR amplitude falling patterns ($\overline{X}$ = 6.77) was higher than that for subjects indicating no religious preference ($\overline{X}$ = 3.90), a preference for Catholicism ($\overline{X}$ = 4.77), or preferring the Protestant denominations ($\overline{X}$ = 3.62). The relatively high probability value associated with that ANOVA (p = .16), despite the perceptibly higher mean for GSR amplitude falling patterns associated with Jewish subjects, was in part a function of the small number of subjects stating a Jewish preference.

Table 30.--A comparison of group means for three categories of school grade point average with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Subject's Grade Point Average: Less Than 2.8 (n=54)	2.8 to 3.3 (n=144)	Greater Than 3.3 (n=72)	One-Way ANOVA Results
Sum of critical item composite ranks[a]	$\overline{X}$=9.44	$\overline{X}$=9.33	$\overline{X}$=9.70	F(2,267)= .32 p=.73
GSR max. height downward drift patterns[b]	$\overline{X}$=1.80	$\overline{X}$=1.55	$\overline{X}$=1.50	F(2,267)= .60 p=.55
GSR amplitude falling patterns[c]	$\overline{X}$=4.31	$\overline{X}$=4.28	$\overline{X}$=4.19	F(2,267)= .01 p=.99

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Table 31.--A comparison of group means for four religious preference categories with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Subject's Religious Preference				One-Way ANOVA Results
	None (n=52)	Jewish (n=13)	Catholic (n=106)	Protestant (n=98)	
Sum of critical item composite ranks[a]	$\bar{X}=9.89$	$\bar{X}=10.46$	$\bar{X}=9.47$	$\bar{X}=9.10$	$F(3,265)=1.20$ $p=.31$
GSR max. height downward drift patterns[b]	$\bar{X}=1.52$	$\bar{X}= 1.62$	$\bar{X}=1.81$	$\bar{X}=1.38$	$F(3,265)=1.26$ $p=.29$
GSR amplitude falling patterns[c]	$\bar{X}=3.90$	$\bar{X}= 6.77$	$\bar{X}=4.77$	$\bar{X}=3.62$	$F(3,265)=1.75$ $p=.16$

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Church Attendance

Table 32 depicts the levels of significance calculated on three one-way ANOVAs, which compared different church attendance categories with respect to the means associated with them on the sum of critical item composite rank scores, GSR maximum height downward drift, and GSR amplitude falling patterns. The data collected on church attendance pertained to the 52 weeks preceding the subjects' participation in the experiment. The categories chosen for analysis purposes were: 0 times, 1 to 25 times, and over 26 times, which comprised 47 subjects, 133 subjects, and 89 subjects, respectively. Although the one-way ANOVA using the sum of critical item composite rank scores as the dependent variable was not significant ($p = .14$), individuals who did not attend church at all in the year preceding their involvement in the experiment had a perceptibly higher mean on that variable (indicating a lower detection efficiency rate; $\bar{X} = 10.27$) than those who attended church between 1 and 25 times ($\bar{X} = 9.29$) and those who attended more than 25 times ($\bar{X} = 9.22$). The ANOVAs conducted on both GSR maximum height downward drift and GSR amplitude falling patterns indicated highly significant differences between their respective means ($p = .005$ and $p = .004$, respectively). The GSR maximum height downward drift means were $\bar{X} = 1.60$, $\bar{X} = 1.30$, and $\bar{X} = 2.02$, respectively, for the 0 time, 1 to 25 times, and over 25 times church attendance groups. Both Duncan and Scheffe post hoc procedures indicated a significant difference between the 1 to 25 times and the over 25 times church attendance groups with respect to their mean score on GSR maximum height downward drift patterns. The means for the GSR amplitude falling patterns followed the same pattern, with values of $\overline{X} = 4.98$, $\overline{X} = 3.16$, and $\overline{X} = 5.52$, respectively, for the 0 time, 1 to 25 times, and over 25 times church attendance categories. A Duncan post hoc comparison indicated the 1 to 25 times group was significantly different than either of the other two church attendance groups; however, the more conservative Scheffe procedure indicated statistically significant differences only between the 1 to 25 times group and the over 25 times group.

Table 32.--A comparison of group means for three categories of church attendance with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Church Attendance in Last Year: 0 Times (n=47)	1 to 25 Times (n=133)	Over 25 Times (n=89)	One-Way ANOVA Results
Sum of critical item composite ranks[a]	$\overline{X}$=10.27	$\overline{X}$=9.29	$\overline{X}$=9.22	$F(2,266)=1.97$ $p=.14$
GSR max. height downward drift patterns[b]	$\overline{X}$= 1.60	$\overline{X}$=1.30	$\overline{X}$=2.02	$F(2,266)=5.49$ $p=.005$
GSR amplitude falling patterns[c]	$\overline{X}$= 4.98	$\overline{X}$=3.16	$\overline{X}$=5.52	$F(2,266)=5.64$ $p=.004$

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Subject-Generated Polygraph Countermeasures

One hundred sixty-six subjects out of the 261 who completed the follow-up questionnaire (Appendix E) indicated that they had attempted to use at least one type of countermeasure to alter their physiological responses during the polygraph tests. Each of these countermeasures was classified as belonging to one of the two following categories: (1) those intended to reduce the subject's physiological responses to the critical items and (2) those intended to increase emotional reactions to noncritical items. Table 33 depicts the various countermeasures employed by the subject and the category to which it was assigned.

Table 34 shows the mean scores for subjects who employed one of the two categories of countermeasures and those for subjects who did not use any self-initiated countermeasures (Category III) with respect to the sum of critical item composite ranks, GSR maximum height downward drift, and GSR amplitude falling patterns. The table also presents the level of significance for the ANOVAs that compared the three category means for each dependent variable. Since certain subjects used two or three different countermeasures that would have resulted in their being classified in both the Category I and Category II countermeasure groups, only those subjects who used either no or just one method to attempt "beating the polygraph" were included in that analysis. Although far from significant (p = .40), it is interesting that the polygraph detection efficiency was the poorest

Table 33.--The frequency and description of two categories of self-initiated countermeasures employed by the subjects during the polygraph tests.

Category[a]	Method	Frequency
I	Concentrating on other things	21
I	Concentrating on pleasant things	13
I	Concentrating on math	1
I	Concentrating on future or past events	8
I	Concentrating on the previous question	6
I	Letting their mind roam	2
I	Rationalizing that no-one was actually shot	4
I	Imagining answering a different question	5
I	Meditating during the testing	5
I	Trying to fall asleep during the testing	2
I	Attempting to make their mind go blank	8
I	Attempting to forget the details of the mock crime	6
I	Breathing in a consistent fashion	8
I	Spraying deodorant on their fingers	2
I	Drinking alcoholic beverages prior to testing	2
I	Smoking marijuana prior to testing	2
I	Attempting to remain perfectly still	1
I	Using instrument noises for biofeedback	1
I	Responding vocally the same to all questions	1
I	Nonspecified method for remaining calm	10
II	Thinking about sexual fantasies	6
II	Thinking about adventurous situations	1
II	Thinking about unpleasant situations	2
II	Imagining different facts pertaining to their crime	18
II	Attempting to be nervous during non-critical items	9
II	Nonspecified method of increasing emotional responsivity	11
II	Attempting breathing pattern	15
II	Moving in an unnoticeable fashion	10
II	Altering vocal responses to the questions	2
II	Attempting to invoke goosebumps/shivers	1

[a]Category I = methods intended to decrease physiological responses to critical items; Category II = methods intended to increase physiological responses to noncritical items.

for Category I subjects ($\overline{X}$ = 9.73), surpassed by Category III ($\overline{X}$ = 9.66) and Category II subjects ($\overline{X}$ = 9.04).

Table 34.--A comparison of group means for three different countermeasure categories with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Countermeasure Category[a]			One-Way ANOVA Results
	I (n=84)	II (n=57)	III (n=95)	
Sum of critical item composite ranks[b]	$\overline{X}$=9.73	$\overline{X}$=9.04	$\overline{X}$=9.66	F(2,233)= .91 p=.40
GSR max. height downward drift patterns[c]	$\overline{X}$=1.73	$\overline{X}$=1.19	$\overline{X}$=1.76	F(2,233)=2.39 p=.09
GSR amplitude falling patterns[d]	$\overline{X}$=5.18	$\overline{X}$=2.77	$\overline{X}$=4.63	F(2,233)=3.41 p=.03

[a]Category I = countermeasures intended to decrease physiological responses to critical items; Category II = countermeasures intended to increase physiological responses to noncritical items; Category III = no countermeasures were employed.

[b]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[c]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[d]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Subjects attempting to increase their physiological responses to the noncritical items appeared to have increased their overall electrodermal responsiveness during the five tests. The group means for GSR maximum height downward drift were $\overline{X}$ = 1.73, $\overline{X}$ = 1.19, and $\overline{X}$ = 1.76 for Categories I through III, respectively, and $\overline{X}$ = 5.18,

$\overline{X} = 2.77$, and $\overline{X} = 4.63$ for those respective categories on the incidence of GSR amplitude falling patterns. For both of these measures, Category II subjects showed the highest degree of electrodermal responsiveness. The differences among the means were significant with respect to GSR amplitude falling patterns ($p = .03$), however were not significant for the GSR maximum height downward drift patterns ($p = .09$). A Duncan post hoc comparison of the GSR amplitude falling pattern means indicated a statistically significant difference between Category II and Category III; however, the Scheffe procedure found no statistically significant differences between the three groups.

Performance Expectancy Scores

As stated earlier, immediately after the demonstration test the subjects were asked to indicate their degree of certainty that they would "beat" three out of the five polygraph tests. After the testing was completed, they were asked how certain they were that they had "beaten" three out of the five tests. Their responses were collapsed into the following three groups: (1) will succeed, (2) neutral, and (3) will fail. These group means were compared with respect to the sum of critical item composite rank values, GSR amplitude falling patterns, and GSR maximum height downward drift patterns. Table 35 presents the means and levels of significance determined by one-way ANOVAs for subjects' performance-expectancy classifications after the demonstration test, whereas Tables 36 and 37 indicate that information for subjects' performance expectancy after the actual test and their combined performance expectancies for the two responses, respectively.

Table 35.--A comparison of group means for three different categories of performance expectancy responses given immediately after the "demonstration" test with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Performance Expectancy After Demonstration Test			One-Way ANOVA Results
	Will Succeed (n=105)	Neutral (n=93)	Will Fail (n=72)	
Sum of critical item composite ranks[a]	$\bar{X}$=9.44	$\bar{X}$=9.44	$\bar{X}$=9.48	F(2,267)=.004 p=1.0
GSR max. height downward drift patterns[b]	$\bar{X}$=1.60	$\bar{X}$=1.49	$\bar{X}$=1.68	F(2,267)= .27 p=.76
GSR amplitude falling patterns[c]	$\bar{X}$=3.81	$\bar{X}$=3.99	$\bar{X}$=5.29	F(2,267)=1.76 p=.17

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

There were no significant differences among the performance expectancy classifications attained after the "demonstration" test, for the actual test, or on their combined values for any of the three aforementioned dependent variables.

Table 36.--A comparison of group means for three different categories of performance expectancy responses given after the actual polygraph tests with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Performance Expectancy After Actual Tests			One-Way ANOVA Results
	Will Succeed (n=112)	Neutral (n=75)	Will Fail (n=83)	
Sum of critical item composite ranks[a]	$\bar{X}$=9.58	$\bar{X}$=9.33	$\bar{X}$=9.42	$F(2,266)= .14$ $\underline{p}=.87$
GSR max. height downward drift patterns[b]	$\bar{X}$=1.57	$\bar{X}$=1.47	$\bar{X}$=1.71	$F(2,266)= .46$ $\underline{p}=.64$
GSR amplitude falling patterns[c]	$\bar{X}$=4.09	$\bar{X}$=3.40	$\bar{X}$=5.30	$F(2,266)=2.50$ $\underline{p}=.08$

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

Attitudinal Responses

The subjects were asked to complete an attitudinal survey that was included in the follow-up questionnaire. The survey comprised 15 statements that required the subjects to indicate whether they agreed, disagreed, or were neutral with respect to each item. These statements are presented in Table 38.

A comparison of the sum of critical item composite rank value group means for students marking agree, neutral, or disagree for each

Table 37.--A comparison of group means for three different categories of the combined performance expectancy scores from before and after the actual polygraph tests with respect to one measure of polygraph detection efficiency and two measures of electrodermal responsiveness.

Dependent Variable	Combined Performance Expectancies			One-Way ANOVA Results
	Will Succeed (n=68)	Neutral (n=163)	Will Fail (n=39)	
Sum of critical item composite ranks[a]	$\bar{X}$=9.82	$\bar{X}$=9.28	$\bar{X}$=9.54	$F(2,267)= .70$ $p=.50$
GSR max. height downward drift patterns[b]	$\bar{X}$=1.65	$\bar{X}$=1.51	$\bar{X}$=1.80	$F(2,267)= .56$ $p=.57$
GSR amplitude falling patterns[c]	$\bar{X}$=3.72	$\bar{X}$=4.07	$\bar{X}$=6.03	$F(2,267)=2.49$ $p=.09$

[a]The higher the mean, the less detection efficiency exhibited by the sum of critical item composite ranks. Possible range: 5-25.

[b]The higher the mean, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

[c]The higher the mean, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

of the statements is presented in Table 39. Only the following question demonstrated a significant difference among the dependent variable means for the three categories: "When I took the polygraph test, I definitely wanted to beat it" (p = .04). The mean sum of critical item composite rank values for subjects agreeing with that statement was $\bar{X}$ = 9.66 (indicating the lowest detection efficiency rate), compared to $\bar{X}$ = 7.91 for those who were neutral and $\bar{X}$ = 8.05 for those disagreeing with it. Duncan post hoc procedures indicated a significant difference between the individuals who wanted to beat the test and those who were neutral; however, the Scheffe procedures indicated no significant differences between the three groups.

Table 38.--The order and wording of statements contained in the attitudinal survey.

Statement Number	Statement
1	People who can beat the polygraph are probably less moral than those who cannot beat it.
2	People who can beat the polygraph are probably more intelligent than those who cannot do it.
3	People who can beat the polygraph probably have better control over their emotions than those who cannot.
4	People who can beat the polygraph probably lie more frequently than those who cannot beat it.
5	I would rather have someone as a friend who could beat the polygraph than someone who could not beat it.
6	When I took the polygraph test, I definitely wanted to beat it.
7	I wanted to beat the polygraph to get the extra credit.
8	I wanted to beat the polygraph to prove to myself that I could control my emotions (keep my cool).
9	I wanted to beat the polygraph to show I could "beat the system."
10	My friends would be impressed if I beat the polygraph test.
11	I found the experiment informative.
12	I was tense throughout the experiment.
13	Committing the mock murder made me feel very uncomfortable.
14	I liked shooting the gun.
15	I enjoyed the experiment.

Table 39.--A comparison of the sum of critical item composite rank means for three categories of responses to statements contained in the follow-up questionnaire.

Statement Number	Response Categories[a] Agree	Neutral	Disagree	One-Way ANOVA Results
1	$\bar{X}$=10.38 n=29	$\bar{X}$=9.56 n=43	$\bar{X}$=9.34 n=192	F(2,261)=1.34 p=.26
2	$\bar{X}$=9.24 n=89	$\bar{X}$=9.63 n=45	$\bar{X}$=9.61 n=130	F(2,261)= .41 p=.67
3	$\bar{X}$=9.48 n=228	$\bar{X}$=8.47 n=17	$\bar{X}$=10.55 n=19	F(2,261)=1.93 p=.15
4	$\bar{X}$=9.42 n=86	$\bar{X}$=9.22 n=59	$\bar{X}$=9.67 n=119	F(2,261)= .40 p=.67
5	$\bar{X}$=8.81 n=16	$\bar{X}$=9.60 n=177	$\bar{X}$=9.40 n=70	F(2,260)= .50 p=.61
6	$\bar{X}$=9.66 n=238	$\bar{X}$=7.91 n=16	$\bar{X}$=8.05 n=10	F(2,261)=3.36 p=.04
7	$\bar{X}$=9.61 n=235	$\bar{X}$=8.79 n=14	$\bar{X}$=8.20 n=15	F(2,261)=1.75 p=.18
8	$\bar{X}$=9.62 n=200	$\bar{X}$=9.56 n=35	$\bar{X}$=8.55 n=29	F(2,261)=1.41 p=.25
9	$\bar{X}$=9.35 n=117	$\bar{X}$=9.79 n=60	$\bar{X}$=9.42 n=86	F(2,260)= .38 p=.68
10	$\bar{X}$=9.51 n=69	$\bar{X}$=9.76 n=99	$\bar{X}$=9.17 n=95	F(2,260)= .83 p=.44
11	$\bar{X}$=9.49 n=246	$\bar{X}$=9.88 n=12	$\bar{X}$=8.92 n=6	F(2,261)= .18 p=.83
12	$\bar{X}$=9.59 n=170	$\bar{X}$=8.78 n=29	$\bar{X}$=9.55 n=65	F(2,261)= .82 p=.44
13	$\bar{X}$=9.68 n=73	$\bar{X}$=9.23 n=37	$\bar{X}$=9.46 n=154	F(2,261)= .26 p=.77
14	$\bar{X}$=9.83 n=110	$\bar{X}$=8.91 n=85	$\bar{X}$=9.67 n=69	F(2,261)=2.16 p=.12
15	$\bar{X}$=9.48 n=253	$\bar{X}$=9.31 n=8	$\bar{X}$=11.17 n=3	F(2,261)= .43 p=.65

[a]The higher the mean value in each response category, the less detection efficiency was exhibited by the sum of critical item composite rank scores. Possible range: 5-25.

A comparison of the GSR maximum height downward drift means for students marking agree, neutral, or disagree is presented in Table 40 for each of the 15 statements. The only question demonstrating a significant difference among the three means was "People who can beat the polygraph probably have better control over their emotions than those who cannot" ($p = .05$). Subjects agreeing with that statement had a mean GSR maximum height downward drift value of $\overline{X} = 1.57$, compared to means of $\overline{X} = 1.35$ and $\overline{X} = 2.47$, respectively, for subjects marking neutral and disagree. Duncan post hoc procedures indicated the mean score for subjects disagreeing with that statement was significantly higher than for the other two groups; however, the Scheffe procedure found no significant differences between the three categories.

The last dependent variable on which the three classifications were compared was GSR amplitude falling patterns. A comparison of the means for each of the statements with respect to that variable is presented in Table 41. The means for the following statement were highly significantly different ($p = .0005$): "My friends would be impressed if I beat the polygraph test." Both Duncan and Scheffe post hoc comparisons indicated subjects agreeing with that statement had significantly fewer GSR maximum increase falling patterns ($\overline{X} = 2.31$) than subjects who were either neutral ($\overline{X} = 4.52$) or who disagreed with it ($X = 5.67$). The statement "Committing the mock murder made me feel very uncomfortable" also yielded significantly different GSR amplitude falling pattern means for the three groups ($p = .04$). Subjects agreeing with that statement had the highest mean number of

Table 40.--A comparison of GSR maximum height downward drift means for three categories of responses to statements contained in the follow-up questionnaire.

Statement Number	Response Categories[a] Agree	Neutral	Disagree	One-Way ANOVA Results
1	$\bar{X}$=1.66 n=29	$\bar{X}$=1.63 n=43	$\bar{X}$=1.61 n=192	F(2,261)= .01 p=.99
2	$\bar{X}$=1.37 n=89	$\bar{X}$=1.68 n=45	$\bar{X}$=1.76 n=130	F(2,261)=1.60 p=.20
3	$\bar{X}$=1.57 n=228	$\bar{X}$=1.35 n=17	$\bar{X}$=2.47 n=19	F(2,261)=3.04 p=.05
4	$\bar{X}$=1.55 n=86	$\bar{X}$=1.63 n=59	$\bar{X}$=1.66 n=119	F(2,261)= .13 p=.88
5	$\bar{X}$=1.25 n=16	$\bar{X}$=1.62 n=177	$\bar{X}$=1.71 n=70	F(2,260)= .53 p=.59
6	$\bar{X}$=1.63 n=238	$\bar{X}$=1.81 n=16	$\bar{X}$= .90 n=10	F(2,261)=1.11 p=.33
7	$\bar{X}$=1.64 n=235	$\bar{X}$=1.50 n=14	$\bar{X}$=1.33 n=15	F(2,261)= .29 p=.75
8	$\bar{X}$=1.64 n=200	$\bar{X}$=1.66 n=35	$\bar{X}$=1.45 n=29	F(2,261)= .18 p=.83
9	$\bar{X}$=1.50 n=117	$\bar{X}$=1.68 n=60	$\bar{X}$=1.74 n=86	F(2,261)= .59 p=.55
10	$\bar{X}$=1.30 n=69	$\bar{X}$=1.68 n=99	$\bar{X}$=1.80 n=95	F(2,260)=1.97 p=.14
11	$\bar{X}$=1.59 n=246	$\bar{X}$=1.83 n=12	$\bar{X}$=2.00 n=6	F(2,261)= .29 p=.75
12	$\bar{X}$=1.60 n=170	$\bar{X}$=1.07 n=29	$\bar{X}$=1.90 n=65	F(2,261)=2.75 p=.07
13	$\bar{X}$=1.84 n=73	$\bar{X}$=1.43 n=37	$\bar{X}$=1.56 n=154	F(2,261)=1.01 p=.37
14	$\bar{X}$=1.49 n=110	$\bar{X}$=1.47 n=85	$\bar{X}$=2.00 n=69	F(2,261)=2.64 p=.07
15	$\bar{X}$=1.61 n=253	$\bar{X}$=1.81 n=8	$\bar{X}$=2.65 n=3	F(2,261)= .49 p=.61

[a]The higher the mean value in each response category, the more frequently GSR maximum height downward drift patterns were produced. Possible range: 0-5.

Table 41.--A comparison of GSR amplitude falling pattern means for three categories of responses to statements contained in the follow-up questionnaire.

Statement Number	Response Categories[a] Agree	Neutral	Disagree	One-Way ANOVA Responses
1	$\bar{X}$=4.69 n=29	$\bar{X}$=3.74 n=43	$\bar{X}$=4.42 n=192	F(2,261)= .33 p=.72
2	$\bar{X}$=3.80 n=89	$\bar{X}$=4.78 n=45	$\bar{X}$=4.56 n=130	F(2,261)= .68 p=.51
3	$\bar{X}$=4.39 n=228	$\bar{X}$=2.76 n=17	$\bar{X}$=5.21 n=19	F(2,261)= .94 p=.39
4	$\bar{X}$=4.35 n=86	$\bar{X}$=3.83 n=59	$\bar{X}$=4.59 n=119	F(2,261)= .37 p=.69
5	$\bar{X}$=3.44 n=16	$\bar{X}$=4.41 n=177	$\bar{X}$=4.42 n=70	F(2,260)= .24 p=.79
6	$\bar{X}$=4.21 n=238	$\bar{X}$=6.00 n=16	$\bar{X}$=4.80 n=10	F(2,261)= .83 p=.44
7	$\bar{X}$=4.39 n=235	$\bar{X}$=2.57 n=14	$\bar{X}$=5.27 n=15	F(2,261)= .95 p=.39
8	$\bar{X}$=4.21 n=200	$\bar{X}$=4.51 n=35	$\bar{X}$=5.03 n=29	F(2,261)= .30 p=.74
9	$\bar{X}$=4.15 n=117	$\bar{X}$=4.43 n=60	$\bar{X}$=4.58 n=86	F(2,261)= .16 p=.86
10	$\bar{X}$=2.31 n=69	$\bar{X}$=4.52 n=99	$\bar{X}$=5.67 n=95	F(2,260)=7.88 p=.0005
11	$\bar{X}$=4.16 n=246	$\bar{X}$=6.25 n=12	$\bar{X}$=7.67 n=6	F(2,261)=1.96 p=.14
12	$\bar{X}$=4.33 n=170	$\bar{X}$=3.45 n=29	$\bar{X}$=4.77 n=65	F(2,261)= .58 p=.56
13	$\bar{X}$=5.57 n=73	$\bar{X}$=2.89 n=37	$\bar{X}$=4.10 n=154	F(2,261)=3.32 p=.04
14	$\bar{X}$=3.61 n=110	$\bar{X}$=3.84 n=85	$\bar{X}$=6.13 n=69	F(2,261)=5.14 p=.006
15	$\bar{X}$=4.19 n=253	$\bar{X}$=8.50 n=8	$\bar{X}$=6.00 n=3	F(2,261)=2.55 p=.08

[a]The higher the mean value in each response category, the more frequently GSR amplitude falling patterns were produced. Possible range: 0-25.

GSR amplitude falling patterns ($\overline{X} = 5.57$), compared to $\overline{X} = 2.89$ for subjects who were neutral and $\overline{X} = 4.10$ for those disagreeing with the statement. A Duncan post hoc comparison indicated that the "neutral" group's mean was significantly different than the mean for subjects agreeing with the statement; however, no group mean differences were found using the Scheffe technique. The last statement that produced significantly different means for the three groups was "I liked shooting the gun" ($\underline{p} = .006$). Both Duncan and Scheffe procedures indicated that subjects agreeing with that statement or who were neutral with respect to it ($\overline{X} = 3.61$ and $\overline{X} = 3.84$, respectively) had significantly fewer GSR amplitude falling patterns than those who disagreed with it ($\overline{X} = 6.13$).

ANOVA Results Examining the Effects of Three Placebo Conditions and Three Feedback Conditions on Both the Accuracy of the Polygraph and on Two Measures of Electrodermal Responsiveness

Earlier in the Results section, a strong relationship between sex and GSR maximum height scores was demonstrated. Sex was also shown to be strongly related to the incidence of both GSR maximum height downward drift and GSR amplitude falling patterns. To compensate for the relatively high incidence of electrodermal nonresponsiveness, as measured by these two variables, the sum of critical item composite rank was developed to serve as an alternative measure of detection efficiency. In this section, the effects of the three placebo conditions, the three feedback conditions, and sex are examined with respect to both the sum of critical item composite rank values and the two measures of electrodermal responsiveness.

The results of the ANOVA examining the relationship of sex, the three levels of placebo, and the three feedback conditions on GSR amplitude falling patterns are presented in Table 42. The main effect for sex was the only significant source of variation found. As illustrated earlier, females had significantly more GSR amplitude falling patterns than did males (p = .001). The only other source of variation approaching significance was the drug x sex x feedback interaction (p = .065). Thus, of the three major independent variables examined in this study, only sex was found to have a major influence on the incidence of GSR maximum increase falling patterns.

Table 42.--Analysis of variance: GSR amplitude falling patterns by sex, three feedback conditions, and three placebo conditions.

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	874.8	33.4	.001
Placebo	2	33.6	1.3	.278
Feedback	2	46.4	1.8	.172
Sex x placebo interaction	2	17.6	.7	.511
Sex x feedback interaction	2	23.0	.9	.416
Placebo x feedback interaction	4	21.4	.8	.516
Sex x placebo x feedback interaction	4	58.6	2.2	.065
Residual	252	26.2	n/a	n/a

The results of the ANOVA examining the relationship of sex, the three placebo conditions, and the three levels of feedback on GSR maximum height downward drift are presented in Table 43. Once again, only the main effect for sex was found to be significant (p = .001).

As with GSR amplitude falling patterns, females had significantly more GSR maximum height downward drift patterns than did males.

Table 43.--Analysis of variance: GSR maximum height downward drift patterns by sex, three feedback conditions, and three placebo conditions.

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	87.8	39.1	.001
Placebo	2	5.4	2.4	.093
Feedback	2	.6	.3	.757
Sex x placebo interaction	2	4.2	1.9	.157
Sex x feedback interaction	2	3.0	1.3	.262
Placebo x feedback interaction	4	2.0	.9	.476
Sex x placebo x feedback interaction	4	3.9	1.7	.142
Residual	252	2.2	n/a	n/a

Table 44 depicts the results of the ANOVA that examined the relationship between sex, the three levels of feedback, and the three levels of placebo on the sum of critical item composite rank values. Not one of the main or interaction effects was found to be significant. Thus the three levels of feedback and placebo and the two sex conditions did not have a significant effect on the detection efficiency of the polygraph, as measured by the sum of critical item composite rank values. These findings also suggest that the relationships between sex and GSR amplitude and GSR maximum height reported earlier may have resulted from the lack of electrodermal responsiveness, which was more prevalent among female than male subjects.

Table 44.--Analysis of variance: sum of critical item composite rank values by sex, three feedback conditions, and three placebo conditions.

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	5.9	.6	.451
Placebo	2	12.8	1.2	.293
Feedback	2	13.6	1.3	.271
Sex x placebo interaction	2	.4	.0	.764
Sex x feedback interaction	2	1.4	.1	.878
Placebo x feedback interaction	4	8.3	.8	.525
Sex x placebo x feedback interaction	4	5.2	.5	.732
Residual	252	10.4	n/a	n/a

Summary

This chapter contained a discussion of the data-analysis methods used in this study, as well as the results of those tests. Included was an examination of the relationships between the specific variables, based on the data collected from the subjects' polygraph charts, performance expectancy self-reports, and follow-up questionnaires. Chapter IV contains a discussion of the findings presented in this chapter.

CHAPTER IV

DISCUSSION

Introduction

The primary focus of this chapter is the discussion of the findings presented in the Results chapter. This is accomplished by first addressing the effects the sex, feedback, and placebo conditions had on both the detection efficiency of the polygraph and on electrodermal responsiveness. This is followed by a discussion of the interaction effects produced by various combinations of those independent variables. Finally, the other findings presented in the Results chapter are addressed such as the detection efficiency of the four dependent variables, the reliability of the measurements obtained from the instrument used to objectively score respiration, and the association between certain responses on the questionnaires and both polygraph detection efficiency and electrodermal responsiveness. In each of the aforementioned areas the discussion centers on the factors that may have affected the results and how compatible the findings were with respect to those reported by other related studies.

After the results have been discussed, the limitations of the study are presented. That subsection focuses on the principal factors that restrict the degree to which the study's findings can be generalized to field situations. Close attention should be paid to this section, since several reasons are cited that necessitate one's exercising caution when drawing inferences from this study to the field.

The Effect of the Subject's Sex

Although there were several significant differences reported between male and female subjects in the Results chapter, almost all of those differences appear to be related to electrodermal responsiveness. Due to the system employed to score GSR maximum height in an objective fashion, the disproportionately large number of GSR maximum height downward drift patterns produced by females resulted in an increase in their GSR maximum height mean values. These inflated GSR maximum height values for females appear to have subsequently led to the significant ANOVA main effects for sex with respect to that dependent variable. In addition, the inflated values resulting from the GSR maximum height downward drift patterns appear to have also been the principal factor leading to the significant MANOVA main effect that was reported for sex.

Similarly, the significant main effect for sex with respect to GSR amplitude on the fifth polygraph test was apparently influenced by the large number of females producing GSR amplitude falling patterns on that test. Since the relative frequency of GSR amplitude falling patterns was much less than for GSR maximum height downward drift patterns, it is possible that this significant effect did not occur until the fifth polygraph test because its respective differences were too small until they were further inflated by the effects of habituation.

Although certain studies comparing the electrodermal activity of men and women have reported differences (Davis, 1932; Graham, Cohen, & Shmavonean, 1966; Kimmel & Hill, 1961; Kimmel & Kimmel, 1965; Plutchik, 1964; Rein, 1926; Shmavonean, Yarmat, & Cohen, 1965), these differences have not been noted in detection-of-deception experiments (Cutrow, Parks, Lucas, & Thomas, 1972). In retrospect, the absence of electrodermal-activity sex differences in prior detection-of-deception experiments might have been due to: (1) few of these studies included women in their sample, (2) sex was generally not isolated as a separate independent variable for analysis purposes, and/or (3) the experimental situation and testing procedures used in prior studies may not have caused the same degree of differentiation between males and females found in this study.

Since a lower degree of electrodermal responsiveness for females was not anticipated in this study and the project was not designed to examine the causes of that phenomenon, one can only speculate using the data at hand why it occurred. It appears equally plausible that the principal cause was either physiologically or psychologically based. Several alternative explanations for that phenomenon will be hypothesized in the paragraphs to follow; however, it is important for the reader to keep in mind that they are only hypotheses and that the actual cause(s) will not be determined until future studies specifically address this question.

It is possible that certain physiological differences between males and females contributed to the females showing significantly less electrodermal responsiveness during the polygraph tests. For example, differences in the thickness of the skin between males and females may have affected their levels of skin resistance. In addition, the females' menstrual cycles or other sex-related factors affecting their endocrine balance could have affected their level of electrodermal activity (Barton, 1940; Burr & Musselman, 1936).

There are also a multitude of potential psychological factors that could have contributed to the lack of responsiveness found among many of the females. Most of the psychological factors that will be addressed are based necessarily on the assumption that in 1977 when the subjects were tested there were differences between the male and female subjects with respect to their values, attitudes, interests, aspirations, etc. The degree of differentiation between male and female roles seems to have declined dramatically over the last few decades. If the trend continues, male and female roles should grow closer together, resulting in less differentiation between the sexes on the aforementioned factors.

Although it seems unlikely, certain females may have found committing a mock murder so repugnant that they repressed their recall of the entire situation. It is also possible that shooting the gun at an image of a person and taking the medication were sufficiently stressful for certain females that they perceived the polygraph testing as a much more relaxing and pleasant aspect of the study. The fact that the polygraph examiner was male might have also had a differential effect on the female subjects. However, it appears more likely that more females than males found the experimental mock murder so out of context with their conception of acceptable behavior that they were to some extent psychologically insulated from producing electrodermal responses to the stimuli. Females may have also found it more difficult than males to imagine themselves ever being in a position similar to the experimental situation, which would probably result in the experiment being less ego involving for them.

In addition, the females were significantly less certain that they would be able to beat three out of the five polygraph tests after the demonstration test than the males. Therefore, it is possible that their lower performance expectancies regarding their ability to deceive the examiner might also have had an impact on their electrodermal responsivity.

The Effect of the Feedback Conditions

The failure of the feedback conditions to produce a significant effect on detectability was somewhat surprising since several other detection-of-deception studies reported that feedback had a significant effect upon subsequent detection (Ellson, Davis, Saltzman, & Burke, 1956; Golden, 1971; Gustafson & Orne, 1965; Suzuki, Watanabe, & Shimizu, 1969). Barland (1972) has been the only other investigator who reported no significant feedback effect. In that study he used the same three feedback conditions utilized in this study. The feedback he provided pertained to a card test, after which he proceeded to employ a series of control-question tests pertaining to the subject's possible involvement in a mock theft. Barland reported that although subjects who were led to believe they successfully deceived the examiner on the card test were more certain they would beat the actual test than subjects who were informed they were detected on the card test, feedback had no significant impact on the detectability of their role in the experiment.

Barland (1972) hypothesized the following reasons might have negated the potential feedback effect in his study. First, he noted that regardless of the feedback conditions, the subjects' attitudes remained fairly positive toward the effectiveness of the polygraph. This was reinforced by information provided by the examiner that the polygraph is highly effective in detecting lies. Second, subjects in the feedback condition who were led to believe they successfully deceived the polygraph examiner were informed that part of the equipment might not have been properly attached. Finally, Barland suggested that feedback on the card test may not have been a direct enough inference for subjects to believe that the outcome of that test would have bearing on the results of the actual tests associated with the mock theft incident.

In this study subjects were asked specifically: (1) how certain they were that they would be able to "beat" three of the five polygraph tests pertaining to the mock murder after the demonstration test, and (2) how certain they were that they did "beat" three out of the five tests after they had taken them. In both instances feedback had a highly significant effect on the subjects' performance expectancies regarding the outcome of the five tests associated with their involvement in the mock murder. Subjects led to believe they "beat" the demonstration test were the most optimistic concerning their ability to "beat" the actual tests, followed by subjects who were not given any feedback and those who were led to believe the examiner would have correctly detected their card, respectively. Thus, the subjects' responses indicated that they did make inferences from the outcome of the card test to the outcome of the five actual tests. In addition, subjects who were led to believe that they successfully "beat" the polygraph on the card test were not informed that this might have been due to having the equipment improperly attached. Therefore, the lack of the three feedback conditions to produce significant differences on the dependent variables indicated that the feedback treatments might not have been strong enough to: (1) totally convince subjects in the feedback fail group that their efforts to deceive the examiner were hopeless and (2) convince subjects in the feedback pass group they had absolutely nothing to fear on the actual tests.

The inability of the feedback conditions to produce a significant effect on the accuracy of the polygraph might also reflect the effectiveness of the experimental situation in making it desirable for the subjects to attempt to deceive the examiner. All of the subjects were led to believe that "prior research has shown that subjects with superior intelligence and good emotional control could beat the polygraph." Gustafson and Orne (1963) reported that subjects in a laboratory detection-of-deception experiment who were told this type of statement had their critical items detected significantly more frequently than chance, whereas the detection rate for subjects who were not told the statement was within chance levels.

Subjects were also motivated to deceive the examiner by rewarding those subjects who beat three out of the five polygraph tests with sufficient extra credit points to make a half letter grade difference on their final grade in one of their criminal justice courses (Michigan State University's grading system reflects .5 numeric intervals). In addition, the nature of the subjects' mock murder assignment and the manner in which it was carried out probably made the experiment highly ego involving for most of the subjects. Therefore, the inability of the three feedback conditions to produce significantly different effects on the detection efficiency of the polygraph was probably due to both (1) their ineffectiveness in totally convincing the subjects that they would have the same fortune on the actual tests that they had on the card test, and (2) the success of the experimental situation in keeping the subjects motivated to deceive the examiner. However, the possibility that feedback might have an inconsequential effect on the polygraph's detection efficiency for most subjects must also be considered.

The Effect of the Placebo Conditions

The only significant main effects for the placebo conditions occurred on the fifth polygraph test. On that test there were significant differences between the placebo conditions on both GSR total length and GSR amplitude. Since these two variables were highly correlated with each other on the fifth test ($\underline{r}$ = .73), it was not surprising that significant differences were noted on both dependent measures instead of on just one. It was surprising, however, that the placebo conditions did not have more of an impact on the other polygraph tests or on the subjects' performance expectancies.

The reasons why significant placebo differences occurred on the fifth test but only on that test are obscure. It is possible that the significant differences represent only capitalizing on chance since numerous ANOVAs were conducted. It is also possible that differential habituation effects or the additional time the subjects had to reflect on the experimental treatments might have contributed to the significant differences. The higher GSR total length and amplitude mean ranks on the fifth test for subjects receiving the adrenergic placebos (indicating the polygraph had less detection efficiency for that placebo condition than for the other two placebo conditions) were consistent with the expected outcome. It was believed that subjects receiving the adrenergic placebo treatments would be more inclined to give up in their attempt at deceiving the examiner and that this would result in their being more likely to escape detection.

The reasons why significant placebo differences were not found on the other four polygraph tests are equally difficult to determine. As noted with the feedback treatments, the inability of the placebo treatments to produce consistent significant effects on the polygraph's detection efficiency probably reflects the relative ineffectiveness of placebo conditions psychologically to overcome the success of the experimental situation in keeping the subjects motivated to deceive the examiner. Once again, however, the possibility that the placebo effect might have an inconsequential impact on the detection efficiency of the polygraph for most subjects should also be considered.

The lack of a significant placebo effect on either of the two performance expectance measures was unexpected. The apparent reason feedback had a much greater effect on the subjects' performance expectancies lies partially in the temporal sequence of the treatments. Since the feedback was given after the placebos were ingested and much closer in time to when the five polygraph tests were to be administered, it is reasonable to assume feedback would have had a greater impact on the performance expectancies than the placebo conditions. In addition, the subjects had no prior personal knowledge concerning the effectiveness of the "medication" in this type of endeavor, which would also serve to make them more inclined to place credence in the feedback they received than in the unknown powers of the medication. Since no direct test was made to ascertain whether the subjects in the placebo groups believed they were receiving actual medication, it is also possible that despite the efforts of the investigator to convince them that the placebos contained active medication, certain subjects might have suspected they were given placebos.

The Effect of Various Combinations of the Sex, Feedback, and Placebo Conditions

The only significant multivariate interaction was the sex x placebo interaction occurring on the fourth polygraph test. In that interaction the detection efficiency rates for females were the poorest with respect to respiration and GSR amplitude in the placebo control groups, whereas those groups had the highest detection efficiency rates for males on those dependent measures. This suggests that the males who were unable to rationalize poor performance on some type of medication tried harder, whereas the females who initially had lower performance expectations were less motivated under those circumstances. However, given the large number of multivariate significance tests run, the lack of homogeneity of dispersion on the fourth test, and the fact that no other multivariate sex x placebo interactions were discovered, it is difficult to determine whether this significant interaction is meaningful.

The only other significant interaction effects found in this study were two univariate feedback x placebo GSR maximum height interactions occurring on the first and fourth polygraph tests. Even though the interactions for those two tests occurred for the same combination of independent and dependent variables, their nature was considerably different. On both tests and detection efficiency was the poorest for the feedback pass subgroups when they were paired with placebo control subgroups. Conversely, the feedback fail-placebo control group had a higher detection efficiency rate on both tests than the feedback fail-tranquilizer placebo group. However, the nature of the maximum height responses for the feedback pass groups and the feedback fail adrenergic placebo group varied considerably on the two tests. This inconsistency coupled with the lack of similar significant interaction effects on the other three polygraph tests also makes these interactions suspect. Although it is also possible the change reflects that as time passed the subjects in the adrenergic placebo groups were less inclined to try to overcome the "effects" attributed to the medication.

The Reliability of the Procedures Used to Measure GSR Total Height and Respiration Total Height

As stated in the Results chapter, the reliability of the procedures employed to measure the total length of the respiration patterns was $\underline{r}$ = .95, whereas the reliability coefficient for the GSR total length was $\underline{r}$ = .94. These coefficients were based on a comparison of the original measurements contained on 30 randomly selected charts with the values derived from a subsequent measurement of those charts. The reliability coefficients of the ranks for the five items on each of the five polygraph tests were also calculated. The reliability of the GSR total length ranks was $\underline{r}$ = .73, whereas the reliability coefficient of the ranks based on the subjects' respiration patterns was $\underline{r}$ = .78. The fairly large drop in both of those dependent variable reliability coefficients from the values based on the actual measurements to the values based on their ranks was due primarily to the loss of information in going from interval to ordinal level data and the small range of values possible on the ranks (one to five).

Although the investigator was forced to develop a rather crude instrument to measure both the respiration and GSR total length patterns, there are currently several highly sophisticated digitizers on the market that would undoubtedly increase the reliability of these measurements.

The Accuracy of the Different Physiological Responses in Detecting Deception

The slight superiority of respiration over the electrodermal measures in detecting deception in this study was not anticipated. Other laboratory detection-of-deception studies have consistently reported GSR to be more accurate than either respiration or cardiovascular measures in detecting lies (Barland & Raskin, 1975; Ellson et al., 1952; Kubis, 1962; Kugelmass & Lieblich, 1966; Podlesny & Raskin, 1978; Thackray & Orne, 1968; Violante & Ross, 1964). One of the reasons contributing to respiration having the highest percentage of correct identifications of the critical items was the large number of charts containing GSR amplitude falling and GSR maximum height downward drift patterns. When the inconclusive GSR patterns were eliminated from these calculations, GSR maximum height was the most accurate index of deception on all of the polygraph tests.

Another factor that might have contributed to respiration's relatively high degree of detection efficiency was the manner in which it was objectively scored. To the best of the investigator's knowledge, this study was the first one that used the length of the respiration pattern as a dependent variable. Given the scoring system employed in this study, this method also served to counteract the effects of GSR responses that were augmented by individuals attempting to beat the polygraph by taking deep breaths, since it was the suppression of the respiration pattern that was indicative of deception.

The intercorrelations between respiration, GSR amplitude, and GSR maximum height were low. In fact, respiration was not correlated ($\underline{r} < |.1|$) with any of the electrodermal measures. What is even more important was that each of those measures was able to differentiate between the critical and noncritical items at a level significantly better than chance. This supports the rationale for using multiple dependent measures to increase the accuracy of the examiner's decisions in detection-of-deception testing.

A Comparison Between the Guilty-Knowledge Accuracy Levels Attained in This Study and the Levels Reported in Other Guilty-Knowledge Detection-of-Deception Experiments

Several laboratory detection-of-deception studies incorporated a guilty-knowledge paradigm into their experimental design (Ben Shakhar, Lieblich, & Kugelmass, 1970; Davidson, 1968; Lieblich, Ben Shakhar, & Kugelmass, 1976; Lykken, 1959, 1960; Podlesny & Raskin, 1978). However, before comparing the accuracy levels obtained in this study with the levels obtained by its predecessors, some important methodological differences should be noted.

In three of the six aforementioned studies (Ben Shakhar et al., 1970; Lieblich et al., 1976; Lykken, 1960), the goal of the detection-of-deception testing was to identify subjects by independently matching known personal information with the subjects' responses to questions concerning that information. The polygraph question format used in those three studies consisted of asking a series of questions containing one relevant and several irrelevant alternatives about each category of personal information examined (i.e., Is your father's name: (a) John? (b) Tom? etc.). The formulation of alternatives and the subsequent matching were based on subsets of five subjects. Therefore, the matching procedure required the experimenter to determine which of five sets of personal information most closely corresponded with the physiological responses on a given polygraph chart.

Using the method described above, Lykken (1960) reported achieving a 100 percent accuracy level in his study (N = 20), while Ben Shakhar et al. (1970) and Lieblich et al. (1976) reported accuracy levels of 77 percent (N = 27) and 62 percent (N = 30), respectively. The perceptibly higher level of accuracy reported by Lykken (1960) might be due to: (1) differences in the experimental design (Lykken trained and encouraged his subjects to employ countermeasures that would create false positive responses); (2) subtle differences in scoring procedures (Lykken's scoring procedure was designed to counteract the effects of subjects creating false positives in a systematic fasion); and/or (3) differences in subject characteristics (Lykken's subjects were affiliated in various capacities with an American medical school, Ben Shakhar et al.'s subjects were Israeli college students, and Lieblich et al.'s subjects consisted of Israeli prisoners).

The results obtained by the three other guilty-knowledge studies (Davidson, 1968; Lykken, 1959; Podlesny & Raskin, 1978) were much more consistent with each other. Davidson (1968) reported an accuracy level of 98 percent (N = 48), Lykken (1959) reported 94 percent (N = 49; two possible crimes considered separately), and Podlesny and Raskin (1978) reported 90 percent (N = 20). In all three of these studies, all of the reported errors were false negatives. These three studies were discussed in detail in the Review of the Selected Literature section because of their similarity to the present study. All three involved a mock crime situation and employed essentially the same question and scoring procedure utilized in this study. This, then, raises the question of why only 58 percent of the subjects in this study had scores of five or greater when their GSR amplitude (skin resistance) responses were scored using the Lykken (1959) procedure, when the accuracy levels of the other three aforementioned studies were 90 percent or higher based solely on skin conductance responses.

Since electrical resistance is a function (the reciprocal) of electrical conductance and only relative differences in electrodermal activity were considered in deriving the aforementioned accuracy levels, differences between skin conductance and skin resistance should not have been a major factor affecting the results. However, it is possible that the electrodermal channel contained in the field polygraph used in this study was not as precise as the highly sophisticated laboratory equipment used in the Davidson (1968), Lykken (1959), and Podlesny and Raskin (1978) studies, which might have affected the results. Although Barland and Raskin (1978) cited numerous studies in their presentation of a convincing argument that dismisses this hypothesis, it may warrant further examination.

Perhaps the most obvious factor reducing the GSR amplitude accuracy levels attained in this study was the fact that only "guilty" subjects were tested. As mentioned earlier, all of the errors reported in the Davidson (1968), Lykken (1959), and Podlesny and Raskin (1978) studies were false negatives (guilty subjects incorrectly classified as innocent). In this study, theoretically approximately 20 percent of all innocent subjects would have been misclassified as guilty if scores of five or greater (based on the probability distribution presented in Table 20) were selected as the criteria for "guilty" classifications. However, the percentage of estimated false positives would be considerably smaller if the questioning procedure was altered to make it more consistent with the Davidson (1968) and Lykken (1959) studies by adding one more question series and selecting scores of six or more as the guilty classification criterion.

Just as interesting as the comparatively low GSR amplitude detection efficiency level attained in this study were the relatively high respiration and composite score accuracy levels that were achieved. Podlesny and Raskin (1978) reported that in their guilty-knowledge experiment, respiration scores did not significantly discriminate between guilty and innocent subjects. Their mean guilty-knowledge respiration score for guilty subjects, based on the same number of questions used in this study, was 3.8 ($\underline{N}$ = 10). In this study the mean respiration guilty-knowledge score was 5.6 ($\underline{N}$ = 270), which was significantly higher than the theoretical population mean for innocent subjects ($\underline{z}$ = 23.7). In addition, 69 percent of the subjects had respiration guilty-knowledge scores of five or greater, and 53 percent had scores of six or more. The difference between the level of accuracy attained by respiration in this study and the one reported by Podlesny and Raskin (1978) might be due to differences in measurement procedures. Respiration total length was computed in this study, whereas Podlesny and Raskin examined respiration amplitude and cycle time. Naturally, subtle differences in the two experimental situations may also have contributed to this discrepancy.

The highest accuracy level found in this study that was based on the Lykken (1959) scoring procedure was attained using the composite rank values. These composite rank values were derived by summing the GSR maximum height, GSR amplitude, and respiration ranks associated with each response, then ranking these composite values within each question series. Using this method, 82 percent of the subjects had guilty-knowledge scores of five or more. Since theoretically approximately 20 percent of all innocent subjects would have been expected to attain a score of five or more (based on the probability distribution presented in Table 20), one could predict that if innocent subjects were included in this study, the overall accuracy of the guilt/innocence categories would have been approximately 80 percent using Lykken's (1959) procedure on the composite rank values and a guilty criterion score of five or more.

The Effect of Habituation

The effect of habituation on electrodermal activity was reflected by the substantial increase in GSR amplitude falling patterns associated with each subsequent polygraph test. Electrodermal habituation effects referred to in this section do not refer to those associated with the repeated presentation of exactly the same stimulus questions, rather to a repetition of various critical and noncritical items pertaining to the same crime. The significantly lower frequency of falling patterns on the critical items compared to the rate on the noncritical items on each of the five polygraph tests is consistent with the findings of other investigators (Ben Shakhar, 1977; Geldreich, 1941; Jones and Wechsler, 1928).

The habituation phenomenon once again raises the question of what factors reduce electrodermal responsiveness. The decrease in the novelty or arousal level of the stimulus as suggested by Berlyne et al. (1963), changes in the endocrine balance following high levels of activation as suggested by Darrow (1936), or possibly hydration or something analogous physiologically to fatigue effects at the neural, dermal, or epidermal levels could each be a contributing factor. However, the resolution of this question is beyond the scope and intent of this particular study. It should also be noted that since the sequence of the five tests remained consistent for all subjects, it is possible that the nature of the tests and their order contributed to the empirical relationship being discussed.

The Relationship Between the Subjects' Responses to Certain Questions Contained on the Questionnaires and Both Polygraph Accuracy and Electrodermal Responsiveness

This section of the Discussion chapter will examine the relationship between the subjects' responses to certain questions contained on the questionnaires and both polygraph detection efficiency and electrodermal responsiveness. For the sake of brevity, only questions producing statistically significant findings will be discussed. However, the reader should keep in mind that statistical significance is dependent upon the sample size within each of the categories being compared. It is entirely possible that several of the other questions would have yielded significant differences if the composition of the subjects was altered to have a more equally balanced set of responses to the various questions. Conversely, since a large number of questions was asked, it is possible that some of the statistically significant findings that will be discussed merely resulted from capitalizing on chance. Also, any causal interpretation of the significant relationships presented may be erroneous, since the relationships could be spurious in nature. For example, a question on the amount of time the subjects spent on applying make-up would have undoubtedly produced significant differences with respect to electrodermal responsiveness, since the responses would have reflected a clear-cut differentiation between males and females. Therefore, all of the findings derived from the subjects' responses to the questions should be viewed more as generated hypotheses rather than as confirmation of the relationships that will be discussed.

The comparison of church attendance with electrodermal responsiveness indicated subjects that either did not attend church at all or attended more than 25 times within the year prior to their participation in the study had significantly more GSR maximum height downward drift and GSR amplitude falling patterns than those attending between 1 and 25 times. This finding could be construed as suggesting that the nonreligious and devoutly religious subjects found the experiment less ego involving than the moderately religious subjects. Naturally, there could also be a multitude of other equally plausible explanations for this finding.

The relationship between GSR amplitude falling patterns and the type of self-initiated countermeasures employed by certain subjects appears much clearer. The subjects who attempted to beat the polygraph by trying to decrease their emotional responses to the critical items had significantly more GSR amplitude falling patterns than the subjects who attempted to create false-positive responses to the noncritical items. This finding indicates that subjects who attempted psychologically to detach themselves from the polygraph tests had less electrodermal responsiveness than those having to pay more attention to the questions in order to know when to attempt to create their false-positive responses. In addition, some of the subjects who attempted to create false-positive responses probably were successful in increasing some of their responses, which would obviously increase the electrodermal responsiveness values.

Several of the different categories of the subjects' responses (agree, neutral, disagree) to statements contained on the attitudinal survey also demonstrated statistically significant differences with respect to polygraph accuracy and electrodermal responsiveness. However, most of these significant differences appeared only with respect to the GSR amplitude falling pattern values attained by the subjects.

The only statement depicting a significant relationship between the responses invoked by it and the sum of critical item composite rank values was "When I took the polygraph test, I definitely wanted to beat it." The detection efficiency of the polygraph for subjects agreeing with that statement was significantly poorer than for those who were neutral or disagreed with it. Although this finding is contrary to what one would expect if fear of detection was the only psychological mechanism involved in invoking the increased psychological responses to the critical items, related findings have been reported and discussed by other researchers (Ben Shakhar, 1977; Gustafson & Orne, 1965). Since the subjects who were less motivated to beat the polygraph probably did not have as much fluctuation in their emotional responses and they probably were less inclined to employ countermeasures, their responses may have been more stable, which might account for the higher accuracy.

The only statement demonstrating a significant relationship between the incidence of GSR maximum height downward drift patterns and the responses invoked by it was "People who can beat the polygraph probably have better control over their emotions than those who cannot." Subjects who either agreed with that statement or who were neutral had significantly fewer GSR maximum height downward drift patterns than those disagreeing with it. This suggests that those subjects who accepted responsibility to some degree for the outcome of the polygraph were more attentive and motivated to beat it. Since all of the subjects were told that "subjects with superior control over their emotional reactions are more successful in beating the polygraph" to enhance their ego involvement and add an additional element of jeopardy to the experiment, the response might also reflect their degree of skepticism regarding the experimental procedures.

Three of the statements indicated a significant relationship between incidence of GSR amplitude falling patterns and the responses invoked by them. Subjects agreeing with the statement, "My friends would be impressed if I beat the polygraph test," had significantly fewer falling patterns than those who were neutral or disagreed with it. This suggests that subjects who perceived beating the polygraph would enhance their social status among their peers were more motivated and attentive during the polygraph tests, which subsequently increased their degree of electrodermal responsiveness.

Subjects agreeing with the statement, "Committing the mock murder made me feel very uncomfortable," had significantly more GSR amplitude falling patterns than those who indicated they were neutral. This suggests that subjects who found committing the mock murder repugnant had their electrodermal responsiveness decreased by the type of psychological, physiological, or combined mechanisms that were discussed previously in the Habituation and Sex Effect sections of this chapter. This premise was also supported by the relationship between the incidence of GSR amplitude falling patterns and the subjects' responses to the statement, "I liked shooting the gun." Subjects who agreed with that statement or who were neutral had significantly fewer falling patterns than those disagreeing with it. However, in both cases it is suspected that the females were over-represented in the categories reflecting the larger number of GSR amplitude falling patterns.

Limitations of the Study

The principal purpose of this study was to examine the effects of sex, placebos, and feedback on the detection of deception. Although the effects of these variables on subjects in laboratory situations were of some interest, the impact of these variables as they relate to field situations was considered far more important. One might assume, then, that the most logical place to conduct the research would have been in the field. This would have greatly increased the generalizability of the study, since the data would have been collected under normal field circumstances using standard field procedures. Unfortunately, the solutions to research design problems in this area are seldom that simple. Although field studies permit greater generalizability, there are inherent problems associated with them in establishing ground truth and in maintaining adequate controls.

In this study the two most important elements were knowing ground truth and being certain that each subject was given the appropriate stimuli. This study was the end product of an attempt to design an ethically sound study in which the two aforementioned elements were maximized with minimal loss of generalizability. However, as can be seen from the limitations listed below, any inferences from this study to field situations will have to be drawn cautiously.

1. Probably the most serious limitation affecting the study's generalizability was the artificial situation the testing was based upon. The subjects were not responsible for actually committing a crime, nor were they placed in the same jeopardy that a real criminal

faces during a deception test. However, to make the situation as realistic and ego involving as possible, the subjects did have to act out a mock murder. To heighten the subject's motivation to deceive the examiner, each subject was told that "although it is difficult to beat the lie detector, research has shown that people with superior intelligence and good powers of concentration are able to control their emotions well enough to succeed." As an additional incentive, subjects who successfully deceived the examiner were awarded substantially more extra credit than those who were detected.

2. Another limitation of the study is that the type of questioning procedure that was used in this experiment is seldom used in the field. Although the guilty-knowledge paradigm bears a fairly close resemblance to the peak-of-tension technique, it differs considerably from the more commonly used field control-question format. The guilty-knowledge technique was selected for use in this study because it is a relatively simple procedure; other researchers have obtained excellent results with it in similar situations (Davidson, 1968; Kubis, 1962; Lykken, 1959; Podlesny & Raskin, 1978); and it can be used by an examiner to try to determine which crime out of several possibilities the subject committed. Since guilty people would be more inclined to take medication which they believe will help them beat the lie detector, all of the subjects were required to commit one of several crimes. Therefore, the last reason presented for using the guilty-knowledge technique was of particular importance.

3. In this study only galvanic skin response and respiration were measured. This differs from field conditions where most polygraph

examiners monitor their subject's cardiovascular activity in addition to those two parameters. The principal reasons cardiovascular responses were not monitored in this study were to avoid subject discomfort and the difficulties associated with objectively scoring cardiovascular responses.

4. The subjects used in this experiment consisted of volunteer college students enrolled in criminal justice classes at Michigan State University during Fall 1977 term. Although these subjects might be similar in nature to some applicants who receive a pre-employment screening test, they were probably more homogeneous as a group and were probably considerably different in psychological make-up than suspects who are given polygraph tests for alleged criminal involvement. This reduces the generalizability of the study, since such characteristics might have affected the subjects' responsiveness to the treatment conditions. In addition, differences in their genetic make-up and/or their moral values might also have affected their physiological responses. The population of subjects described above was chosen, despite these problems, due to their accessibility, the convenience of conducting the study at the university, and the ability to award them nonfinancial rewards.

5. All of the mock murder victims (priest, fireman, police officer, army officer, surgeon) selected for use in this study would probably be considered as value-positive figures to most individuals in our society. The nature of the occupations selected for inclusion in this study may have subsequently affected the generalizability of the results.

6. The results stemming from the biographical data sheet and the follow-up questionnaire may have been affected by two factors. First, not all of the subjects may have interpreted the questions the same. For example, when asked about the number of times they attended church in the last 12 months, certain subjects may have included visits to church for weddings, funerals, etc. Second, the formulation of discrete categories for certain variables such as grade point average, age, church attendance, etc., for analysis purposes was to some extent arbitrary. The incorporation of the cut-off points that were selected, as opposed to others that could have been chosen, may have affected both the results and their generalizability.

7. One of the major problems associated with the generalizability of the placebo treatment was that normally people who take "medication" to help them "beat" the lie detector probably believe that it may help or they would not take it. In this experiment all subjects assigned to certain groups were given the placebo; in fact, half of those given the placebo were told that it would make it more difficult for them to "beat" the instrument. Even though the latter group has few direct practical applications, it provided data regarding detection-of-deception rates for subjects with a lower than normal performance expectancy. Another limitation of the placebo condition was that although several steps were taken to make the placebo as credible as possible, it is still unknown if the credibility of the placebo was higher or lower than home remedies.

8. The last major limitation of the study that will be discussed pertains to the feedback conditions. It is doubtful whether

any subjects in the "real world" would be given prior arbitrary feedback that would lead them to believe they could successfully deceive the examiner. Although card tests are administered by some field examiners as a stimulation procedure, many of them have the order of the cards arranged so they know which card the subject chooses (Reid & Inbau, 1966). Normally the only prior feedback a field subject might have that he previously "beat" a lie detector test would be if for some reason he did. Even if this was the situation, the generalizability of the above results is questionable since the two field-testing situations would probably differ considerably (examiner, procedure, location, temporal proximity), as well as all the other limitations that have been discussed. Once again, the data collected on this group have helped to provide insights regarding the effects of differing subject performance expectancies on the detection of deception.

Despite the aforementioned limitations, this study has generated a considerable amount of data regarding the effects of sex, placebos, and feedback on the detection of deception. Every effort was taken to design an ethically sound study that was as realistic and ego involving as possible. However, all inferences to the field from this study will still have to be drawn with extreme caution.

APPENDICES

APPENDIX A

INFORMED CONSENT FORM

APPENDIX A

INFORMED CONSENT FORM

I have freely consented to take part in a scientific study being conducted by Howard W. Timm under the supervision of Frank S. Horvath, Ph.D., Assistant Professor, School of Criminal Justice. I have been informed that the study is designed to test the effects of certain variables which may affect the ability of lie detection instruments to analyze deception. I am aware that I will be given a lie detector test concerning my involvement in a mock contract murder situation.

I understand that I may be given one of two drugs which are designed to affect my emotional responses during the lie detection experiment. I voluntarily consent to participate in this study with the understanding that both of the drugs I may be given have been found safe in other persons, but with the further understanding that not all of the effects of the drugs on lie detection are known.

The study has been explained to me and all of my questions have been satisfactorily answered.

I understand that I am free to discontinue my participation in the study at any time without penalty.

I understand that the results of the study will be treated in strict confidence and that I will remain anonymous. Upon completion of the study, results will be made available to me at my request within the aforementioned restrictions.

I understand that my participation in this study does not guarantee any beneficial results to me.

Signed________________________________

Date__________________________________

APPENDIX B

MEDICAL RECORD RELEASE FORM

APPENDIX B

MEDICAL RECORD RELEASE FORM

I, ______________________________, hereby authorize any physician connected with the Olin Health Center at Michigan State University to examine my medical records with the following conditions:

A. These records are used solely to determine if there is anything in my medical records that would indicate that the drug which may be given to me for experimental purposes would adversely affect me.

B. Only licensed physicians connected with the Olin Health Center are permitted to examine my medical records.

C. No specific information contained in these records may be released or discussed with any outside individuals.

D. The only information that may be released is whether or not my medical records indicate that I should not be given the drug and this information is only to be released to Frank S. Horvath, Ph.D., or those working under his immediate supervision.

Signature______________________

Date___________________________

Student Number_________________

APPENDIX C

MOCK MURDER CONTRACT

APPENDIX C

MOCK MURDER CONTRACT

Subject's Name:____________________

Subject #:____________________

Date:____________________

Your assignment is to participate in a mock contract murder. You are to simulate killing (<u>name of intended victim</u>), who is a (<u>victim's occupation</u>) in the Lansing area. You must shoot at the victim exactly <u>(#)</u> times. If you do so, the (<u>location of the Mafia family</u>) branch of the Mafia will pay you (<u># of dollars</u>). However, before you shoot the person, you must say, "(<u>victim's name</u>), I am shooting you for betraying the (<u>family's location</u>) branch of the Mafia."

A picture of the victim is located in the right-hand corner of this contract. Be certain that you are able to recognize the victim so you do not shoot the wrong person.

After you have completed the simulated murder, you will be given a lie detector test. The examiner will attempt to find out the name of your victim, what his occupation was, how many times you shot at him, the location of the Mafia branch that hired you, and how much you were paid. If you can successfully deceive him ("beat" the lie detector) on three or more of the tests, you will be awarded extra credit totalling 5 percent of the total points possible in your class. However, if you beat the lie detector less than three times, you will only receive 1 percent extra credit.

Prior research has shown that subjects with superior intelligence and good powers of concentration are able to control their emotional responses well enough to succeed.

In order to receive <u>any</u> extra credit, however, you must:

A. Say the required statement before shooting your victim.

B. Shoot the correct person.

C. Not inform the lie detection examiner whether or not you were given one of the drugs.

D. Not inform the examiner of any details concerning the mock contract murder you participated in.

E. Complete all of the lie detection tests.

F. Not discuss the experiment with any volunteers who have not already participated in the study.

If you violate any of these rules, you will be disqualified from the experiment and not receive <u>any</u> extra credit.

APPENDIX D

BIOGRAPHICAL DATA SHEET

APPENDIX D

BIOGRAPHICAL DATA SHEET

Subject #____________

NAME: ____
Last First Middle

LOCAL ADDRESS: ________________________ PHONE: ________________

AGE: ________________ MAJOR: ________________

SEX: ________________ YEAR IN COLLEGE: ______

RACE: ________________ CURRENT CRIMINAL JUSTICE COURSES: ______________

ETHNIC BACKGROUND: ______________

FAMILY SIZE: ________________ GRADE POINT AVERAGE: ____

NUMBER OF PARENTS OR GUARDIANS WORKING: ____ RELIGIOUS PREFERENCE: ___

FAMILY INCOME (PARENTS OR GUARDIAN):

0 - 4,999 ____
5,000 - 9,999 ____
10,000 - 14,999 ____
15,000 - 19,999 ____
20,000 - 24,999 ____
25,000 - 29,999 ____
30,000 and over ____

CHURCH ATTENDANCE IN THE LAST YEAR:

0 Times ____
1 - 5 Times ____
5 - 10 Times ____
10 - 25 Times ____
25 - 35 Times ____
35 - 52 Times ____
over 52 Times ____

NUMBER OF PREVIOUS POLYGRAPH EXAMINATIONS:

Field: __________

Experimental: ________

COMMENTS:

APPENDIX E

QUESTIONNAIRE

APPENDIX E

QUESTIONNAIRE

Name ______________________________
Last First Middle

Please answer the following questions as honestly as possible. Your responses have nothing to do with how much extra credit you will receive and will be kept in strict confidence.

1. Did you spend any time thinking of ways to beat the polygraph?

____ YES

____ NO

2. Which of the following sources of information (if any) did you consult to try to find a method to beat the polygraph (mark more than one response if applicable).

____ A) Thinking about it by yourself.

____ B) Discussing it with friends or relatives.

____ C) Looking through reference books or articles already in your possession.

____ D) Looking through reference books or articles that you had to borrow from someone or the library.

____ E) Consulting with someone you believed to be an expert in the field.

____ F) Other, please explain.

3. How much time did you spend altogether trying to come up with a method to beat the polygraph?

____ Minutes (60 minutes = 1 hour)

4. Did you actually try to do something during the test that you thought might have helped you beat the polygraph (i.e., smoking marijuana before taking the test, thinking of adventurous or sexy things during the test, unnoticeable movement, etc.)?

____ YES

____ NO

If yes, please describe your method below in detail.

5. Are you left handed or right handed?

    ____ LEFT

    ____ RIGHT

Use the scoring key listed below to respond to the following statements.

1 = Strongly Agree
2 = Agree
3 = Partially Agree
4 = Neutral
5 = Partially Disagree
6 = Disagree
7 = Strongly Disagree

6. _____ People that can beat the polygraph are probably less moral than those who cannot beat it.

7. _____ People that can beat the polygraph are probably more intelligent than those who cannot do it.

8. _____ People that can beat the polygraph probably have better control over their emotions than those who cannot.

9. _____ People that can beat the polygraph probably lie more frequently than those who cannot beat it.

10. _____ I would rather have someone as a friend that could beat the polygraph than someone who could not beat it.

11. _____ When I took the polygraph test, I definitely wanted to beat it.

12. _____ I wanted to beat the polygraph to get the extra credit.

13. _____ I wanted to beat the polygraph to prove to myself that I could control my emotions (keep my cool).

14. _____ I wanted to beat the polygraph to show I could "beat the system."

15. _____ My friends would be impressed if I beat the polygraph test.

16. _____ I found the experiment informative.

17. _____ I was tense throughout the experiment.

18. _____ Committing the mock murder made me feel very uncomfortable.

19. _____ I liked shooting the gun.

20. _____ I enjoyed the experiment.

APPENDIX F

MANOVA AND ANOVA RESULTS

APPENDIX F

MANOVA AND ANOVA RESULTS

The following results were obtained by employing multivariate analysis of variance techniques in which four dependent variables and three independent factors were analyzed for each of the five polygraph tests. The three independent factors were: (1) sex (male, female); (2) drug (adrenegenic placebo, no placebo, and tranquilizer placebo); and (3) feedback (beat demonstration test, no feedback, failed demonstration test). The four dependent variables were: (1) the rank of the critical item with respect to the total length of the GSR for each of the five responses (CIRKTOT); (2) the rank of the critical item with respect to the maximum height attained by each of the five GSRs (CIRKHIT); (3) the rank of the critical item with respect to the maximum vertical increase of each of the five GSRs (CIRKINT); and (4) the rank of the critical item with respect to the total length of each of the five respiration patterns (CIRKRST). After each of the aforementioned seven character dependent variable abbreviations, another character will be placed to indicate the respective polygraph test the variable represents. For example, CIRKRST2 represents the rank of the critical item with respect to the total length of the five respiration patterns for the second polygraph test.

Multivariate and Univariate Tests of Significance With the Scores on All Five Polygraph Tests Added Together For Each Dependent Variable

Univariate Tests for Homogeneity of Variance

Variable . . CIRKTOTT

Cochran's C (14,18) =	.103, p = .388 (approx.)
Bartlett-Box F (17,30081) =	.733, p = .770

Variable . . CIRKHITT

Cochran's C (14,18) =	.078, p = 1.000 (approx.)
Bartlett-Box F (17,30081) =	.378, p = .990

Variable . . CIRKINTT

Cochran's C (14,18) =	.099, p = .514 (approx.)
Bartlett-Box F (17,30081) =	1.049, p = .401

Variable . . CIRKRSTT

Cochran's C (14,18) =	.090, p = 1.000 (approx.)
Bartlett-Box F (17,30081) =	.893, p = .582

Multivariate Test for Homogeneity of Dispersion

Boxs M =	206.49
F (170,38012) =	1.08, p = .251 (approx.)

Multivariate and Univariate Tests of Significance for the Sex x Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.045	.710	16.00	1008.00	.786
Hotellings	.046	.571	20.00	990.00	.934
Wilks	.956	.711	16.00	761.35	.784
Roys	.038	Not significant at $\alpha < .05$			

Univariate F-Tests with (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	15.52	2964.70	3.88	11.76	.33	.858
CIRKHITT	75.98	2866.80	18.99	11.38	1.67	.158
CIRKINTT	23.15	2827.97	5.79	11.22	.52	.724
CIRKRSTT	44.39	3458.00	11.10	13.72	.81	.521

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0; N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.074	1.183	16.00	1008.00	.275
Hotellings	.076	.943	20.00	990.00	.532
Wilks	.928	1.182	16.00	761.35	.277
Roys	.040	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	67.32	2964.70	16.83	11.76	1.43	.224
CIRKHITT	81.01	2866.80	20.25	11.38	1.78	.133
CIRKINTT	61.20	2827.97	15.30	11.22	1.36	.247
CIRKRSTT	54.64	3458.00	13.66	13.72	1.00	.411

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.021	.661	8.00	600.00	.726
Hotellings	.021	.660	8.00	496.00	.728
Wilks	.980	.660	8.00	498.00	.727
Roys	.018	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	.05	2964.70	.03	11.76	.00	.998
CIRKHITT	10.07	2866.80	5.03	11.38	.44	.643
CIRKINTT	17.84	2827.97	8.92	11.22	.80	.453
CIRKRSTT	6.50	3458.00	3.25	13.72	.23	.790

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2; M = 1/2; N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.029	.932	8.00	500.00	.490
Hotellings	.030	.929	8.00	496.00	.492
Wilks	.971	.931	8.00	498.00	.491
Roys	.023	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	7.47	2964.70	3.73	11.76	.32	.728
CIRKHITT	14.96	2856.80	7.48	11.38	.66	.519
CIRKINTT	4.76	2827.97	2.38	11.22	.21	.809
CIRKRSTT	11.47	3458.00	5.74	13.72	.42	.659

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.013	.415	8.00	500.00	.912
Hotellings	.013	.413	8.00	496.00	.913
Wilks	.987	.414	8.00	498.00	.912
Roys	.011	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	19.92	2964.70	9.96	11.76	.85	.430
CIRKHITT	18.36	2866.80	9.18	11.38	.81	.447
CIRKINTT	4.96	2827.97	2.48	11.22	.22	.802
CIRKRSTT	3.27	3458.00	1.63	13.72	.12	.888

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.019	.610	8.00	500.00	.770
Hotellings	.020	.507	8.00	496.00	.772
Wilks	.981	.609	8.00	498.00	.771
Roys	.015	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	35.84	2964.70	17.92	11.76	1.52	.220
CIRKHITT	28.67	2866.80	14.34	11.38	1.26	.285
CIRKINTT	33.08	2827.97	16.54	11.22	1.47	.231
CIRKRSTT	2.82	3458.00	1.41	13.72	.10	.902

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.072	4.843	4.00	249.00	.00089
Hotellings	.078	4.843	4.00	249.00	.00089
Wilks	.928	4.843	4.00	249.00	.00089
Roys	.072	4.843	4.00	249.00	.00089

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOTT	.33	2964.70	.33	11.76	.02	.866
CIRKHITT	158.70	2866.80	158.70	11.38	13.95	.00023
CIRKINTT	16.63	2827.97	16.63	11.22	1.48	.225
CIRKRSTT	5.07	3458.00	5.07	13.72	.37	.544

Multivariate and Univariate Tests of Significance for the First Polygraph Test

Univariate Tests for Homogeneity of Variance

Variable . . CIRKTOT1 CI RANK GSR TOTAL LENGTH--TEST ONE

Cochran's C (14,18) = .092, p = .861 (approx.)
Bartlett-Box F (17,30081) = 1.059, p = .391

Variable . . CIRKHIT1 CI RANK MAX HEIGHT--TEST ONE

Cochran's C (14,18) = .088, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .706, p = .799

Variable . . CIRKINT1 CI RANK GSP INCREASE--TEST ONE

Cochran's C (14,18) = .100, p = .496 (approx.)
Bartlett-Box F (17,30081) = .746, p = .757

Variable . . CIRKRST1 CI RANK RESPIRATION--TEST ONE

Cochran's C (14,18) = .095, p = .711 (approx.)
Bartlett-Box F (17,30081) = .955, p = .508

Multivariate Test for Homogeneity of Dispersion

Boxs M = 181.238
F (170,38012) = .946, p = .670 (approx.)

Multivariate and Univariate Tests of Significance for the Sex x Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, 1 = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.052	.832	16.00	1008.00	.650
Hotellings	.053	.659	20.00	990.00	.868
Wilks	.949	.827	16.00	761.35	.655
Roys	.027	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	7.41	430.17	1.85	1.71	1.09	.364
CIRKHIT1	3.75	329.33	.94	1.31	.72	.581
CIRKINT1	4.93	431.10	1.23	1.71	.72	.578
CIRKRST1	7.57	456.53	1.89	1.81	1.04	.385

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.091	1.462	16.00	1008.00	.106
Hotellings	.095	1.180	20.00	990.00	.264
Wilks	.911	1.471	16.00	761.35	.104
Roys	.063	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	5.86	430.17	1.46	1.71	.86	.490
CIRKHIT1	15.26	329.33	3.81	1.31	2.92	.022
CIRKINT1	2.10	431.10	.53	1.71	.31	.873
CIRKRST1	3.56	456.53	.89	1.81	.49	.743

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.017	.544	8.00	500.00	.823
Hotellings	.017	.542	8.00	496.00	.825
Wilks	.983	.543	8.00	498.00	.824
Roys	.014	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	1.90	430.17	.95	1.71	.56	.574
CIRKHIT1	.05	329.33	.03	1.31	.02	.980
CIRKINT1	.82	431.10	.41	1.71	.24	.788
CIRKRST1	.94	456.53	.47	1.71	.26	.772

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.036	1.153	8.00	500.00	.326
Hotellings	.037	1.151	8.00	496.00	.327
Wilks	.964	1.152	8.00	498.00	.327
Roys	.028	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	9.65	430.17	4.82	1.71	2.83	.061
CIRKHIT1	.90	329.33	.45	1.31	.34	.710
CIRKINT1	4.55	431.10	2.28	1.71	1.33	.266
CIRKRST1	3.43	456.53	1.71	1.81	.95	.389

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.016	.513	8.00	500.00	.947
Hotellings	.016	.510	8.00	496.00	.849
Wilks	.984	.512	8.00	498.00	.848
Roys	.012	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	1.04	430.17	.52	1.71	.30	.739
CIRKHIT1	2.76	329.33	1.38	1.31	1.06	.349
CIRKINT1	.02	431.10	.01	1.71	.01	.993
CIRKRST1	1.49	456.53	.74	1.81	.41	.663

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.026	.819	8.00	500.00	.586
Hotellings	.026	.821	8.00	496.00	.584
Wilks	.974	.820	8.00	498.00	.585
Roys	.024	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	1.18	430.17	.59	1.71	.35	.708
CIRKHIT1	6.54	329.33	3.27	1.31	2.50	.084
CIRKINT1	.14	431.10	.07	1.71	.04	.961
CIRKRST1	.16	456.53	.08	1.81	.04	.958

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.047	3.052	4.00	249.00	.018
Hotellings	.049	3.052	4.00	249.00	.018
Wilks	.953	3.052	4.00	249.00	.018
Roys	.047	3.052	4.00	249.00	.018

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT1	3.45	430.17	3.45	1.71	2.02	.157
CIRKHIT1	7.17	329.33	7.17	1.31	5.49	.020
CIRKINT1	.68	431.10	.68	1.71	.39	.530
CIRKRST1	3.79	456.53	3.79	1.81	2.09	.149

Multivariate and Univariate Tests of Significance for the Second Polygraph Test

Univariate Tests of Homogeneity of Variance

Variable . . CIRKTOT2 CI RANK GSR TOTAL LENGTH--TEST TWO

Cochran's C (14,18) = .097, p = .603 (approx.)
Bartlett-Box F (17,30081) = .879, p = .600

Variable . . CIRKHIT2 CI RANK MAX HEIGHT--TEST TWO

Cochran's C (14,18) = .093, p = .814 (approx.)
Bartlett-Box F (17,30081) = .616, p = .882

Variable . . CIRKINT2 CI RANK GSR INCREASE--TEST TWO

Cochran's C (14,18) = .089, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .980, p = .479

Variable . . CIRKRST2 CI RANK RESPIRATION--TEST TWO

Cochran's C (14,18) = .108, p = .257 (approx.)
Bartlett-Box F (17,30081) = .599, p = .895

Multivariate Test for Homogeneity of Dispersion

Boxs M = 197.613
F (170,38012) = 1.032, p = .384 (approx.)

Multivariate and Univariate Tests of Significance for the Sex x Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.040	.645	16.00	1008.00	.849
Hotellings	.042	.516	20.00	990.00	.961
Wilks	.960	.644	16.00	761.35	.849
Roys	.033	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	.44	322.77	.11	1.28	.09	.987
CIRKHIT2	4.71	340.80	1.18	1.35	.87	.482
CIRKINT2	3.82	311.67	.96	1.24	.77	.544
CIRKRST2	2.93	430.80	.73	1.71	.43	.788

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.067	1.081	16.00	1008.00	.368
Hotellings	.069	.857	20.00	990.00	.644
Wilks	.934	1.077	16.00	761.35	.373
Roys	.036	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	3.33	322.77	.83	1.28	.65	.628
CIRKHIT2	6.10	340.80	1.53	1.35	1.13	.344
CIRKINT2	8.01	311.67	2.00	1.24	1.62	.170
CIRKRST2	3.73	430.80	.93	1.71	.54	.703

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.023	.723	8.00	500.00	.671
Hotellings	.023	.720	8.00	496.00	.674
Wilks	.977	.722	8.00	498.00	.672
Roys	.018	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	.17	322.77	.08	1.28	.07	.936
CIRKHIT2	2.60	340.80	1.30	1.35	.96	.384
CIRKINT2	2.24	311.67	1.12	1.24	.91	.406
CIRKRST2	.39	430.80	.19	1.71	.11	.894

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.013	.415	8.00	500.00	.912
Hotellings	.013	.413	8.00	496.00	.913
Wilks	.987	.414	8.00	498.00	.913
Roys	.010	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	.50	322.77	.25	1.28	.20	.822
CIRKHIT2	.29	340.80	.14	1.35	.11	.899
CIRKINT2	.62	311.67	.31	1.24	.25	.778
CIRKRST2	2.10	430.80	1.05	1.71	.61	.542

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.023	.736	8.00	500.00	.660
Hotellings	.024	.734	8.00	496.00	.661
Wilks	.977	.735	8.00	498.00	.661
Roys	.019	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	2.97	322.77	1.48	1.28	1.16	.316
CIRKHIT2	2.27	340.80	1.14	1.35	.84	.433
CIRKINT2	2.34	311.67	1.17	1.24	.94	.390
CIRKRST2	.83	430.80	.41	1.71	.24	.785

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.017	.551	8.00	500.00	.818
Hotellings	.018	.548	8.00	496.00	.920
Wilks	.982	.550	8.00	498.00	.819
Roys	.013	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	.82	322.77	.41	1.28	.32	.728
CIRKHIT2	3.56	340.80	1.78	1.35	1.32	.270
CIRKINT2	.92	311.67	.45	1.24	.37	.690
CIRKRST2	.94	430.80	.47	1.71	.28	.760

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.035	2.241	4.00	249.00	.065
Hotellings	.036	2.241	4.00	249.00	.065
Wilks	.965	2.241	4.00	249.00	.065
Roys	.035	2.241	4.00	249.00	.065

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	2.80	411.13	2.70	1.63	1.65	.199
CIRKHIT3	6.23	304.27	6.23	1.21	5.16	.024
CIRKINT3	.30	383.00	.30	1.52	.20	.657
CIRKRST3	4.54	398.53	4.54	1.58	2.87	.092

Multivariate and Univariate Tests of Significance for the Fourth Polygraph Test

Univariate Homogeneity of Variance Tests

Variable . . CIRKTOT4 CI RANK GSR TOTAL LENGTH--TEST FOUR

Cochran's C (14,18) = .085, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .263, p = .999

Variable . . CIRKHIT4 CI RANK MAX HEIGHT--TEST FOUR

Cochran's C (14,18) = .087, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .747, p = .755

Variable . . CIRKINT4 CI RANK GSR INCREASE--TEST Four

Cochran's C (14,18) = .083, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .665, p = .839

Variable . . CIRKRST4 CI RANK RESPIRATION--TEST FOUR

Cochran's C (14,18) = .096, p = .669 (approx.)
Bartlett-Box F (17,30081) = 1.155, p = .295

Multivariate Test for Homogeneity of Dispersion

Boxs M = 250.458
F (170,38012) = 1.307, p = .009 (approx.)

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.030	1.927	4.00	249.00	.107
Hotellings	.031	1.927	4.00	249.00	.107
Wilks	.970	1.927	4.00	249.00	.107
Roys	.030	1.927	4.00	249.00	.107

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT2	1.27	322.77	1.27	1.28	.99	.321
CIRKHIT2	2.70	340.80	2.70	1.35	2.00	.159
CIRKINT2	.53	311.67	.53	1.24	.43	.512
CIRKRST2	1.79	430.80	1.79	1.71	1.05	.307

Multivariate and Univariate Tests of Significance for the Third Polygraph Test

Univariate Homogeneity of Variance Tests

Variable . . CIRKTOT3 CI RANK GSR TOTAL LENGTH--TEST THREE

Cochran's C (14,18) = .098, p = .547 (approx.)
Bartlett-Box F (17,30081) = .555, p = .925

Variable . . CIRKHIT3 CI RANK MAX HEIGHT--TEST THREE

Cochran's C (14,18) = .098, p = .574 (approx.)
Bartlett-Box F (17,30081) = .584, p = .906

Variable . . CIRKINT3 CI RANK GSR INCREASE--TEST THREE

Cochran's C (14,18) = .104, p = .356 (approx.)
Bartlett-Box F (17,30081) = 1.063, p = .386

Variable . . CIRKRST3 CI RANK RESPIRATION--TEST THREE

Cochran's C (14,18) = .079, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = 1.197, p = .259

Multivariate Test for Homogeneity of Dispersion

Boxs M = 232.812
F (170,38012) = 1.215, p = .042 (approx.)

Multivariate and Univariate Tests of Significance for the Sex x Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.043	.692	16.00	1008.00	.804
Hotellings	.045	.551	20.00	990.00	.945
Wilks	.957	.690	16.00	761.35	.806
Roys	.031	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	1.04	411.13	.26	1.63	.16	.958
CIRKHIT3	6.99	304.27	1.75	1.21	1.45	.219
CIRKINT3	5.66	383.00	1.41	1.52	.93	.447
CIRKRST3	2.48	398.53	.62	1.58	.39	.814

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.080	1.274	16.00	1008.00	.206
Hotellings	.082	1.016	20.00	990.00	.440
Wilks	.922	1.274	16.00	761.35	.207
Roys	.046	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	4.96	411.13	1.24	1.63	.76	.552
CIRKHIT3	4.33	304.27	1.08	1.21	.90	.467
CIRKINT3	7.06	383.00	1.76	1.52	1.16	.328
CIRKRST3	4.93	398.53	1.23	1.58	.779	.540

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.023	.741	8.00	500.00	.655
Hotellings	.024	.740	8.00	496.00	.656
Wilks	.977	.740	8.00	498.00	.656
Roys	.020	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	1.52	411.13	.76	1.63	.46	.629
CIRKHIT3	1.65	304.27	.83	1.21	.68	.506
CIRKINT3	1.07	383.00	.54	1.52	.35	.703
CIRKRST3	5.01	398.53	2.50	1.58	1.58	.207

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.021	.660	8.00	500.00	.727
Hotellings	.021	.655	8.00	496.00	.732
Wilks	.980	.657	8.00	498.00	.729
Roys	.012	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	3.29	411.13	1.64	1.63	1.01	.366
CIRKHIT3	1.30	304.27	.65	1.21	.54	.585
CIRKINT3	2.49	383.00	1.24	1.52	.82	.442
CIRKRST3	2.34	398.53	1.17	1.58	.74	.478

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.031	.977	8.00	500.00	.453
Hotellings	.031	.970	8.00	496.00	.459
Wilks	.969	.974	8.00	498.00	.456
Roys	.019	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	5.15	411.13	2.57	1.63	1.58	.209
CIRKHIT3	2.99	304.27	1.49	1.21	1.24	.292
CIRKINT3	5.68	383.00	2.84	1.52	1.87	.156
CIRKRST3	1.90	398.53	.95	1.58	.60	.550

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.016	.504	8.00	500.00	.854
Hotellings	.016	.502	8.00	496.00	.855
Wilks	.984	.503	8.00	498.00	.854
Roys	.014	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT3	1.70	411.13	.85	1.63	.52	.595
CIRKHIT3	2.59	304.27	1.29	1.21	1.07	.344
CIRKINT3	3.52	383.00	1.76	1.52	1.16	.316
CIRKRST3	.21	398.53	.10	1.58	.07	.937

Multivariate and Univariate Tests of Significance for the Drug x Sex x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.048	.759	16.00	1008.00	.733
Hotellings	.049	.605	20.00	990.00	.911
Wilks	.953	.758	16.00	761.35	.735
Roys	.034	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	1.99	407.00	.497	1.62	.31	.873
CIRKHIT4	3.82	309.73	.96	1.23	.78	.541
CIRKINT4	.71	367.87	.18	1.46	.12	.975
CIRKRST4	5.59	385.87	1.40	1.53	.91	.457

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, 1 = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.085	1.362	16.00	1008.00	.153
Hotellings	.089	1.096	20.00	990.00	.347
Wilks	.917	1.368	16.00	761.35	.151
Roys	.057	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	13.03	407.00	3.26	1.62	2.02	.093
CIRKHIT4	16.59	309.73	4.15	1.23	3.37	.010
CIRKINT4	11.19	367.87	2.80	1.46	1.92	.108
CIRKRST4	3.33	385.87	.83	1.53	.54	.704

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.042	1.348	8.00	500.00	.217
Hotellings	.043	1.339	8.00	496.00	.222
Wilks	.958	1.344	8.00	498.00	.219
Roys	.026	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	5.31	407.00	2.66	1.62	1.64	.195
CIRKHIT4	.9	309.73	.04	1.23	.04	.964
CIRKINT4	2.27	367.87	1.13	1.46	.78	.460
CIRKRST4	3.92	385.87	1.96	1.53	1.28	.280

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.096	3.139	8.00	500.00	.002
Hotellings	.104	3.226	8.00	496.00	.001
Wilks	.905	3.183	8.00	498.00	.002
Roys	.087	Significant at $\alpha < .01$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	4.11	407.00	2.06	1.62	1.27	.282
CIRKHIT4	2.16	309.73	1.08	1.23	.88	.417
CIRKINT4	4.95	367.87	2.47	1.46	1.69	.186
CIRKRST4	4.01	385.87	2.00	1.53	1.31	.272

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.045	1.424	8.00	500.00	.184
Hotellings	.046	1.428	8.00	496.00	.182
Wilks	.956	1.426	8.00	498.00	.183
Roys	.038	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	1.27	407.00	.63	1.62	.39	.675
CIRKHIT4	1.45	309.73	.73	1.23	.49	.555
CIRKINT4	1.61	367.87	.81	1.46	.55	.576
CIRKRST4	2.41	385.87	1.20	1.53	.79	.457

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.034	1.084	8.00	500.00	.373
Hotellings	.035	1.076	8.00	496.00	.378
Wilks	.966	1.080	8.00	498.00	.376
Roys	.020	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	5.22	407.00	2.61	1.62	1.61	.201
CIRKHIT4	3.87	309.73	1.94	1.23	1.58	.209
CIRKINT4	2.08	367.87	1.04	1.46	.71	.492
CIRKRST4	2.67	385.87	1.34	1.53	.87	.419

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.034	2.181	4.00	249.00	.072
Hotellings	.035	2.181	4.00	249.00	.072
Wilks	.966	2.181	4.00	249.00	.072
Roys	.034	2.181	4.00	249.00	.072

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT4	.11	407.00	.11	1.62	.07	.792
CIRKHIT4	7.50	309.73	7.50	1.23	6.10	.014
CIRKINT4	1.13	367.87	1.13	1.46	.78	.379
CIRKRST4	1.48	385.87	1.48	1.53	.97	.326

Multivariate and Univariate Tests of Significance for the Fifth Polygraph Test

Univariate Tests for Homogeneity of Variance

Variable . . CIRKTOT5 CI RANK GSR TOTAL LENGTH--TEST FIVE

Cochran's C (14,18) = .079, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .260, p = .999

Variable . . CIRKHIT5 CI RANK MAX HEIGHT--TEST FIVE

Cochran's C (14,18) = .106, p = .285 (approx.)
Bartlett-Box F (17,30081) 1.257, p = .213

Variable . . CIRKINT5 CI RANK GSR INCREASE--TEST FIVE

Cochran's C (14,18) = .076, p = 1.000 (approx.)
Bartlett-Box F (17,30081) = .518, p = .945

Variable . . CIRKRST5 CI RANK RESPIRATION--TEST FIVE

Cochran's C (14,18) = .094, p = .761 (approx.)
Bartlett-Box F (17,30081) = .820, p = .671

Multivariate Test for Homogeneity of Dispersion Matrices

Boxs M = 237.093
F (170,38012) = 1.238, p = .029 (approx.)

Multivariate and Univariate Tests of Significance for the Sex x Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.024	.387	16.00	1008.00	.386
Hotellings	.025	.309	20.00	990.00	.999
Wilks	.976	.386	16.00	761.35	.986
Roys	.023	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	3.72	440.23	.93	1.75	.53	.712
CIRKHIT5	5.19	281.73	1.30	1.12	1.16	.329
CIRKINT5	3.39	418.90	.85	1.66	.51	.728
CIRKRST5	1.69	421.20	.42	1.67	.25	.908

Multivariate and Univariate Tests of Significance for the Drug x Feedback Interaction

Multivariate Tests of Significance (S = 4, M = 0, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.051	.817	16.00	1008.00	.666
Hotellings	.053	.654	20.00	990.00	.873
Wilks	.949	.817	16.00	761.35	.667
Roys	.037	Not significant at $\alpha < .05$			

Univariate F-Tests With (4,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	4.19	440.23	1.05	1.75	.60	.663
CIRKHIT5	.57	281.73	.14	1.12	.13	.972
CIRKINT5	6.99	418.90	1.75	1.66	1.05	.382
CIRKRST5	9.21	421.20	2.30	1.67	1.38	.242

Multivariate and Univariate Tests of Significance for the Sex x Feedback Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.019	.610	8.00	500.00	.770
Hotellings	.020	.607	8.00	496.00	.772
Wilks	.981	.609	8.00	498.00	.771
Roys	.015	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	1.96	440.23	.98	1.75	.56	.572
CIRKHIT5	1.30	281.73	.65	1.12	.58	.561
CIRKINT5	3.95	418.90	1.98	1.66	1.19	.307
CIRKRST5	2.02	421.20	1.01	1.67	.60	.547

Multivariate and Univariate Tests of Significance for the Sex x Drug Interaction

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.020	.630	8.00	500.00	.753
Hotellings	.020	.630	8.00	496.00	.753
Wilks	.980	.630	8.00	498.00	.753
Roys	.018	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	.986	440.23	.49	1.75	.28	.755
CIRKHIT5	1.65	281.73	.83	1.12	.74	.479
CIRKINT5	.62	418.90	.31	1.66	.19	.829
CIRKRST5	2.29	421.20	1.14	1.67	.68	.505

Multivariate and Univariate Tests of Significance for the Main Effect of Feedback

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.017	.521	8.00	500.00	.841
Hotellings	.017	.519	8.00	496.00	.842
Wilks	.983	.520	8.00	498.00	.841
Roys	.015	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	4.74	440.23	2.37	1.75	1.36	.259
CIRKHIT5	.59	281.73	.29	1.12	.26	.770
CIRKINT5	3.12	413.90	1.56	1.66	.94	.392
CIRKRST5	1.03	421.20	.51	1.67	.31	.735

Multivariate and Univariate Tests of Significance for the Main Effect of Drug

Multivariate Tests of Significance (S = 2, M = 1/2, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.037	1.19	8.00	500.00	.305
Hotellings	.038	1.19	8.00	496.00	.305
Wilks	.963	1.19	8.00	498.00	.305
Roys	.031	Not significant at $\alpha < .05$			

Univariate F-Tests With (2,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	11.47	440.23	5.73	1.75	3.28	.039
CIRKHIT5	1.25	281.73	.63	1.12	.56	.572
CIRKINT5	11.87	418.90	5.94	1.66	3.57	.030
CIRKRST5	.90	421.20	.45	1.67	.27	.765

Multivariate and Univariate Tests of Significance for the Main Effect of Sex

Multivariate Tests of Significance (S = 1, M = 1, N = 123-1/2)

Test Name	Value	Approx. F	Hypothesis D.F.	Error D.F.	Signif. of F
Pillais	.039	2.52	4.00	249.00	.042
Hotellings	.041	2.52	4.00	249.00	.042
Wilks	.961	2.52	4.00	249.00	.042
Roys	.039	2.52	4.00	249.00	.042

Univariate F-Tests With (1,252) D.F.

Variate	Hypothesis Sum of Sq.	Error Sum of Sq.	Hypothesis Mean Sq.	Error Mean Sq.	F	Signif. of F
CIRKTOT5	2.50	440.23	2.50	1.75	1.43	.232
CIRKHIT5	9.26	281.73	9.26	1.12	8.28	.004
CIRKINT5	6.53	418.90	6.53	1.66	3.93	.049
CIRKRST5	.53	421.20	.53	1.67	.32	.573

APPENDIX G

ANOVA RESULTS EXAMINING THE EFFECTS OF SEX, PLACEBO CONDITION, FEEDBACK CONDITION, AND ALL OF THEIR POSSIBLE INTERACTIONS ON FOUR MEASURES OF POLYGRAPH DETECTION EFFICIENCY THAT WERE BASED ON THE LYKKEN (1959) SCORING PROCEDURE

APPENDIX G

ANOVA RESULTS EXAMINING THE EFFECTS OF SEX, PLACEBO CONDITION, FEEDBACK CONDITION, AND ALL OF THEIR POSSIBLE INTERACTIONS ON FOUR MEASURES OF POLYGRAPH DETECTION EFFICIENCY THAT WERE BASED ON THE LYKKEN (1959) SCORING PROCEDURE

Analysis of variance: GSR amplitude scores based on the Lykken (1959) scoring procedure by sex, three feedback conditions, and three placebo conditions

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	9.3	1.6	.20
Placebo	2	9.1	1.6	.20
Feedback	2	.53	.1	.91
Sex x placebo interaction	2	2.4	.4	.66
Sex x feedback interaction	2	7.3	1.3	.28
Placebo x feedback interaction	4	10.7	1.9	.11
Sex x placebo x feedback interaction	4	2.6	.5	.76
Residual	252	5.7	N/A	N/A

Analysis of variance: GSR amplitude scores based on the Lykken (1959) scoring procedure by sex, three feedback conditions, and three placebo conditions

Source of Variation	D.F.	Mean Sqaure	F Value	Signif. of F
Sex	1	.3	.07	.80
Placebo	2	6.6	1.2	.30
Feedback	2	1.7	.31	.73
Sex x placebo interaction	2	3.2	.60	.73
Sex x feedback interaction	2	10.3	1.9	.15
Placebo x feedback interaction	4	6.6	1.2	.30
Sex x placebo x feedback interaction	4	3.6	.67	.62
Residual	219	5.4	N/A	N/A

Note: Subjects producing three or more GSR amplitude falling patterns on three or more of the five polygraph tests were excluded.

Analysis of variance: respiration scores based on the Lykken (1959) scoring procedure by sex, three feedback conditions, and three placebo conditions

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	4.8	.78	.38
Placebo	2	2.2	.36	.70
Feedback	2	1.2	.20	.82
Sex x placebo interaction	2	2.5	.41	.66
Sex x feedback interaction	2	2.7	.44	.65
Placebo x feedback interaction	4	4.3	.70	.60
Sex x placebo x feedback interaction	4	3.8	.61	.65
Residual	252	6.1	N/A	N/A

Analysis of variance: composite values with respect to GSR amplitude, GSR maximum height, and respiration based on the Lykken (1959) scoring procedure by sex, three feedback conditions, and three placebo conditions

Source of Variation	D.F.	Mean Square	F Value	Signif. of F
Sex	1	6.7	1.3	.26
Placebo	2	4.8	.90	.41
Feedback	2	3.6	.67	.51
Sex x placebo interaction	2	.78	.15	.86
Sex x feedback interaction	2	5.6	1.1	.35
Placebo x feedback interaction	4	.48	.09	.99
Sex x placebo x feedback interaction	4	3.1	.59	.67
Residual	252	5.3	N/A	N/A

BIBLIOGRAPHY

BIBLIOGRAPHY

Abrams, Stanley. "Polygraph Validity and Reliability: A Review." Journal of Forensic Studies 18 (1973): 313-26.

Abrams, S.; March, L.; Weir, R.; Zimmerman, C.; and Zonn, L. "APA Reviews the Polygraph, Employees, and Unions." Security World 9(6) (1972): 20-30.

________. "APA Reviews the Polygraph, Employees and Unions." Security World 9(8) (1972): 48-52.

Abrams, Stanley, and Weinstein, Edwin. "The Validity of the Polygraph With Retardates." Journal of Police Science and Administration 2(1) (1974): 11-14.

Alpert, M.; Kurtzberg, R. L.; and Friedhoff, A. J. "Transient Voice Changes Associated With Emotional Stimuli." Archives of General Psychiatry 8 (1963): 362-65.

Amick, D. J., and Walberg, H. J. Introductory Multivariate Analysis. Berkeley, Calif.: McCutchan Publishing Corp., 1975.

Arthur, R. O. "Non-Reactors." Journal of Polygraph Studies 11 (1977): 1-4.

Atkinson, John W. An Introduction to Motivation. Princeton: Van Nostrand, 1964.

________. "Towards Experimental Analysis of Human Motivation in Terms of Motives, Expectancies, and Incentives." In Motives in Fantasy, Action, and Society. Edited by J. W. Atkinson. New York: Van Nostrand, 1959.

Baesen, Henry; Chung, Chia-Mon; and Yang, Chen-Ya. "A Lie Detection Experiment." Journal of Criminal Law and Criminology 39 (1948): 532-37.

Barefoot, J. K., ed. "The Polygraph Technique." The Use of Polygraphs and Similar Devices by Federal Agencies. Hearings before a subcommittee of the Committee on Government Operations, House of Representatives, 93rd Congress, 2d session. Washington, D.C.: Government Printing Office, 1974.

Barland, Gordon. "Detection of Deception in Criminal Suspects: A Field Validation Study." Ph.D. dissertation, University of Utah, 1975.

________. "An Experimental Study of Field Techniques in Lie Detection." M.A. thesis, University of Utah, 1972.

Barland, Gordon H., and Raskin, David C. "Detection of Deception." In Electrodermal Activity in Psychological Research. Edited by William F. Prokasy and David C. Raskin. New York: Academic Press, 1973.

________. "An Evaluation of Field Techniques in Detection of Deception." Psychophysiology 12(3) (1975): 321-27.

Barton, D. S. "Electric Correlates of the Menstrual Cycle in Women." Yale Journal of Biological Medicine 12 (1940): 335-44.

Beecher, H. K. "Quantification of the Subjective Pair Experience." In Psychopathology of Perception. Edited by P. Hoch and J. Zubin. New York: Grune & Stratton, 1965.

Ben-Shakhar, G. "A Further Study of the Dichotomization Theory in Detection of Information." Psychophysiology 14 (1977): 408-13.

________; Lieblich, I.; and Kugelmass, S. "Guilty Knowledge Technique: Application of Signal Detection Measures." Journal of Applied Psychology 54 (1970): 409-13.

Benussi, Vittorio. "Die Atmungsymptome det Luge" (The Respiratory Symptoms of Lies). Achiv. f. die gesamte Psychologie 31 (1914): 244.

Berlyne, D. E.; Craw, M. A.; Salapatck, P. H.; and Lewis, J. L. "Novelty, Complexity, Incongruity, Extrinsic Motivation and the GSR." Journal of Experimental Psychology 66 (1963): 560-67.

Berrien, F. K. "Ocular Stability in Deception." Journal of Applied Psychology 26 (1942): 55-63.

________, and Huntington, G. H. "An Exploratory Study of Pupillary Responses During Deception." Journal of Experimental Psychology 32 (1943): 443-49.

Bersh, Philip. "A Validation Study of Polygraph Examiner Judgments." Journal of Applied Psychology 53(5) (1969): 399-403.

Bitterman, M. E., and Marcuse, F. L. "Cardiovascular Responses of Innocent Persons to Criminal Investigation." American Journal of Psychology 60 (1947): 407-12.

Block, Jack D. "A Study of Affective Responsiveness in a Lie-Detection Situation." Journal of Abnormal and Social Psychology 55 (1957): 11-15.

________; Rouke, F. L.; Salpeter, M. M.; Tobach, E.; Kubis, J. F.; and Welch, L. "An Attempt at Reversal of the Truth-Lie Relationships as Measured by the Psychogalvanic Response." Journal of Psychology 34 (1952): 55-66.

Blum, Richard. "The Polygraph Examination in Law Enforcement Personnel Selection." Police (Nov.-Dec. 1967): 60-74.

Buckalew, L. W. "An Analysis of Experimental Components in a Placebo Effect." The Psychological Record 22 (1972): 113-19.

Burns, P., and Cohen, E. SPSS-MANOVA. Vogelback Computing Center, Northwestern University. Document no. 413(Rev. A), June 1977.

Burr, H. S., and Musselman, L. K. "Bioelectric Phenomena Associated With Menstruation." Yale Yournal of Biological Medicine 9 (1936): 155-58.

Burtt, Harold. "A Pneumograph for Inspiration-Expiration Ratios." Psychology Bulletin 15(10) (1918): 325.

Business Week. "Corporate Lie Detectors Come Under Fire." January 13, 1973, pp. 89-90.

Chappell, M. N. Blood Pressure Changes in Deception." Archives of Psychology 17 (1929): 1-39.

Committee on Government Operations. Hearings: "Use of Polygraphs as 'Lie Detectors' by the Federal Government." Washington, D.C.: Government Printing Office, 1964.

________. House Report No. 94-795. "The Use of Polygraphs and Similar Devices by Federal Agencies. Washington, D.C.: Government Printing Office, 1976.

________. "Use of Polygaphs as 'Lie Detectors' by the Federal Government." House Report No. 88. Washington, D.C.: Government Printing Office, 1965.

Committee on the Judiciary, U.S. Senate. "Privacy, Polygraphs, and Employment." Washington, D.C.: Government Printing Office, 1974.

Cooley, W. W., and Lohnes, P. R. Multivariate Data Analysis. New York: John Wiley and Sons, 1971.

Dahm, Anthony E. "Study of the Field Use of the Psychological Stress Evaluator." Committee on Government Operations, Hearings: "The Use of Polygraphs and Similar Devices by Federal Agencies." Washington, D.C.: Government Printing Office, 1974.

Darrow, C. W. "The Galvanic Skin Reflex and Blood Pressure as Preparatory and Facilitative Functions." Psychological Bulletin 33 (1963): 73-94.

Davidson, P. O. "Validity of the Guilty-Knowledge Technique: The Effects of Motivation." Journal of Applied Psychology 52 (1968): 62-65.

Davis, R. C. "Electrical Resistance Before and After a Period of Noise Stimulation." Journal of Experimental Psychology 15 (1932): 108-17.

Dektor Counterintelligence and Security, Inc. The Psychological Stress Evaluator Training Manual. Distributed at PSE Orientation Courses, 500 Port Royal Road, Springfield, VA 21151.

Dykman, R. A.; Reese, W. G.; Gailbrecht, C. R.; and Thomasson, P. J. "Psycho-Physiological Reactions to Novel Stimuli; Measurement, Adaption, and Relationship of Psychological and Physiological Variables in the Normal Human." Annals of the New York Academy of Sciences 79 (1959): 43-107.

Easton, David. The Political System. New York: Alfred A. Knopf, 1953.

Edwards, W. "Measurement of Utility and Subjective Probability." In Psychological Scaling: Theory and Application. Edited by H. Gulliksen and S. Messick. New York: Wiley, 1960.

Ellson, D. G.; Davis, R. C.; Saltzman, I. J.; and Burke, C. J. "A Report of Research on Detection of Deception." Indiana University, Contract No. 6 Nonr-18011 Office of Naval Research, 1952.

Evans, Frederick J. "Placebo Response: Relationship to Suggestibility and Hypnotizability." Proceedings, 77th Annual Convention, APA, 1969.

Ferguson, Robert. "How Correct Are Critics of the--Lie Detector--Test?" Police (Jan.-Feb. 1968): 37-40.

________, and Miller, A. L. Polygraph in Court. Springfield, Ill.: Charles C. Thomas, 1973.

Field, L. H., and Williams, Mark. "A Note on the Scientific Assessment and Treatment of the Sexual Offender." Medicine, Science, and the Law 2(4) (1971): 180-81.

Geldreich, E. W. "Studies of the Galvanic Skin Response as a Deception Indicator." Transactions of the Kansas Academy of Science 44 (1941): 346-51. Cited by Orne, M. T.; Thackray, R. I.; and Paskewitz, D. A. "On the Detection of Deception." In Handbook of Psychophysiology. Edited by N. S. Greenfield and R. A. Sternbach. New York: Holt, Rinehart, and Winston, 1972.

Glass, G. V., and Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Golden, R. I. "Audio GSR Biofeedback in Polygraph Examinations." Presented at the American Polygraph Association Seminar, Atlanta, GA, August 1971. Cited by G. Barland. "An Experimental Study of Field Techniques in Lie Detection." M.A. Thesis, University of Utah, 1972.

Gottaschalk, L. A., and Gleser, G. C. The Measurement of Psychological States Through the Content Analysis of Verbal Behavior. Berkeley, Calif.: University of California Press, 1969.

Graham, L. A.; Cohen, S. I.; and Shmavonian, B. M. "Sex Differences in Autonomic Responses During Instrumental Conditioning." Psychosomatic Medicine 28 (1966): 264-71.

Greenfield, N. S., and Steinbach, R. A., eds. Handbook of Psychophysiology. New York: Holt, Rinehart and Winston, 1972.

Gustafson, Lawrence, and Orne, Martin. "The Effects of Heightened Motivation on the Detection of Deception." Journal of Applied Psychology 47 (1963): 408-11.

________. "The Effects of Task and Method of Stimulus Presentation on the Detection of Deception." Journal of Applied Psychology 48(6) (1964): 383-87.

________. "The Effects of Verbal Responses on the Laboratory Detection of Deception." Psychophysiology 2(1) (1965): 10-13.

Harrelson, L. H. The Keeler Technique (Keeler Polygraph Institute Training Guide). Chicago: Keeler Polygraph Institute, 1964. Cited by G. H. Barland and D. C. Raskin. "Detection of Deception." In Electrodermal Activity in Psychological Research. Edited by W. F. Prokasy and D. C. Raskin. New York: Academic Press, 1973.

Henson, Donald, and Rubin, H. B. "Voluntary Control of Eroticism." Journal of Applied Behavior Analysis 4(1) (1971): 37-44.

Herman, D. J. "Privacy, the Prospective Employee, and Employment Testing--The Need to Restrict Polygraph and Personality Testing." Washington Law Review 47(1) (1971): 73-154.

Horvath, Frank. "The Accuracy and Reliability of Police Polygraphic ('Lie Detector') Examiners' Judgments of Truth and Deception: The Effect of Selected Variables." Ph.D. dissertation, Michigan State University, 1974.

________. "An Experimental Comparison of the Psychological Stress Evaluator and the Galvanic Skin Response in Detection of Deception." Journal of Applied Psychology 62 (1978): 338-44.

________. "The Police Candidate Polygraph Examination: Considerations for the Police Administrator." Police (June 1972): 33-39.

________, and Reid, John. "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception." Journal of Criminal Law, Criminology, and Police Science 62(2) (1971): 276-81.

Hunter, Fred, and Ash, Philip. "The Accuracy and Consistency of Polygraph Examiners' Diagnosis." Journal of Police Science and Administration 1(3) (1973): 370-75.

Inbau, Fred, and Reid, John. Lie Detection and Criminal Interrogation. Baltimore: Williams & Wilkins Co., 2nd ed., 1948, 3rd ed., 1953.

Jacobson, E. Progressive Relaxation. Chicago: University of of Chicago Press, 1938.

Jones, H. E., and Wechsler, D. "Galvanometric Techniques in Studies of Association." American Journal of Psychology 40 (1928): 607-12.

Keeler, L. "Scientific Methods of Crime Detection With the Polygraph." Kansas Bar Association Journal 2 (1941): 22-51.

Kimmel, H. D., and Hill, F. A. "A Comparison of Two Electrodermal Measures of Response to Stress." Journal of Comparative and Physiological Psychology 54 (1961): 395-97.

Kimmel, H. D., and Kimmel, E. "Sex Differences in Adaptation of the GSR Under Repeated Application of a Visual Stimulus." Journal of Experimental Psychology 70 (1965): 536-37.

Klump, C. S. "So You Want to Beat a Polygraph!" Security World (June 1965): 30-32.

Kradz, Michael P. "The Psychological Stress Evaluator." Committee on Government Operations, Hearings: "The Use of Polygraphs and Similar Devices by Federal Agencies." Washington, D.C.: Government Printing Office, 1974.

Kubis, Joseph. "Comparison of Voice Analysis and Polygraph as Lie Detection Procedures." Contract DAAD05-72-C-0217, U.S. Army Land Wayfare Laboratory, Aberdeen Proving Ground, Maryland, August 1973.

________. "Experimental and Statistical Factors in the Diagnosis of Consciously Suppressed Affective Experience." Journal of Clinical Psychology 6 (1950): 12-16.

________. "Studies in Lie Detection: Computer Feasibility Considerations." RADC-TR 62-205, 1962, Armed Services Technical Information Agency, Arlington, Virginia.

Kugelmass, S. Reactions to Stress. Air Force Office of Scientific Research Rep. 67-0530; Defense Documentation Center: AD-647467, January 1967.

________, and Lieblich, Israel. "Effects of Realistic Stress and Procedural Interference in Experimental Lie Detection." Journal of Applied Psychology 50(3) (1966): 211-16.

Lacey, John, and Lacey, Beatrice. "Verification and Extension of the Principle of Autonomic Response--Stereotypy." American Journal of Psychology 71 (1958): 50-73.

Landis, C., and Wiley, L. E. "Changes of Blood Pressure and Respiration During Deception." Journal of Comparative Psychology 6 (1926): 1-19.

Langfield, H. S. "Psychophysical Symptoms of Deception." Journal of Abnormal Psychology 15 (1921): 319-28.

Larson, John. Lying and Its Detection. Chicago: University of Chicago Press, 1932. Reprinted Montclair, N.J.: Patterson-Smith Co., 1969.

Lasagna, L.; Mosteller, F.; Von Felsinger, J.; and Beecher, H. K. "A Study of the Placebo Response." American Journal of Medicine 16 (1954): 770-79.

Lawler, Edward E. Motivation in Work Organizations. Monterey, Calif.: Brooks/Cole, 1973.

Laws, D. R., and Rubin, H. B. "Instructional Control of an Autonomic Sexual Response." Journal of Applied Behavior Analysis 2 (1969): 93-97.

Lee, Clarence. The Instrumental Detection of Deception. Springfield, Ill.: Charles C. Thomas, 1953.

Lewin, Kurt. Principles of Topological Psychology. New York: McGraw-Hill, 1936.

Lieblich, Israel; Ben-Shakhar, Gershon; and Kugelmass, Sol. "Validity of the Guilty Knowledge Technique in a Prisoners' Sample." Journal of Applied Psychology 61(1) (1976): 89-93.

Lippold, Olaf. "Physiological Tremor." Scientific American 224(3) (March 1971): 65-73.

Lombroso, Cesare. L'Homme Criminal. 2nd ed., 1895, 1:336-346. Also see Crime, Its Causes and Remedies. Translated by H. P. Horton, 1911, 137.

Lykken, David. "The GSR in the Detection of Guilt." Journal of Applied Psychology 43(5) (1959): 385-88.

________. "The Validity of the Guilty Knowledge Technique: The Effects of Faking." Journal of Applied Psychology 44(4) (1960).

MacNitt, Reginald. "In Defense of the Electrodermal Response and Cardiac Amplitude as Measures of Deception." Journal of Criminal Law and Criminology 33 (1942): 266-75.

Marston, William. The Lie Detector Test. New York: Richard K. Smith, 1938.

________. "Psychological Possibilities in the Deception Test." Journal of Criminal Law 11 (1921): 551-68.

________. "Studies in Testimony." Journal of Criminal Law and Criminology 15(1) (1924): 5.

________. "Systolic Blood Pressure Changes in Deception." Journal of Experimental Psychology 91(2) (1917): 117-63.

McClelland, David C. The Achieving Society. Princeton: Van Nostrand, 1961.

________. Personality. New York: William Sloane-Dryden Press, 1951.

________; Atkinson, John W.; Clark, Russell A.; and Lowell, Edgar L. The Achievement Motive. New York: Appleton, 1953.

Munsterberg, Hugo. On the Witness Stand. New York: Doubleday, Page & Co., 1908.

Nie, N. H.; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K.; and Bent, D. H. Statistical Package for the Social Sciences. 2nd ed. New York: McGraw-Hill, 1970.

Obermann, C. E. "The Effect of the Berger Rhythm of Mild Affective States." Journal of Abnormal and Social Psychology 34 (1939): 84-95.

Orlansky, Jesse. "An Assessment of Lie Detection Capability (Declassified Version)." Technical Report 62-16, Institute for Defense Analysis, Washington, D.C., 1964.

Orne, Martin T.; Thackray, Richard I.; and Paskewitz, David A. "On the Detection of Deception: A Model for the Study of Physiological Effects of Psychological Stimuli." In Handbook of Psychophysiology. Edited by Norman S. Greenfield and Richard A. Sternbach. New York: Holt, Rinehart, and Winston, 1972.

Peak, Helen. "Psychological Structure and Psychological Activity." Psychological Review 65 (1958): 325-47.

Plutchik, R. "Effect of Electrode Placement on Skin Impedance Related Measures." Psychological Record 14 (1964): 145-51.

Podlesny, John A., and Raskin, David C. "Effectiveness of Techniques and Physiological Measures in the Detection of Deception." Psychophysiology 5(4) (1978): 344-59.

Podlesny, J. A.; Raskin, D. C.; and Barland, G. H. Effectiveness of Techniques and Physiological Measures in the Detection of Deception. Report No. 76-5, Contract 75-NI-99-001, U.S. Department of Justice. Salt Lake City, Utah: University of Utah, Department of Psychology, August 1976.

Press, S. J. Applied Multivariate Analysis. New York: Holt, Rinehart, and Winston, 1972.

Prokasy, W. F., and Raskin, D. C., eds. Electrodermal Activity in Psychological Research. New York: Academic Press, 1973.

Raskin, David C. Psychopathy and Detraction of Deception in a Prison Population. Report No. 75-1, Contract 75 NI 99-0001. Salt Lake City, Utah: University of Utah, Department of Psychology, 1975.

________. "Statement at Hearings." Polygraph Control and Civil Liberties Protection Act. Hearings before the Subcommittee on the Constitution of the Committee on the Judiciary, U.S. Senate, 95th Congress. Washington, D.C.: Government Printing Office, 1978.

_______, and Hare, R. D. "Psychopathy and Detection of Deception in a Prison Population." Psychophysiology 15 (1978): 126-36.

Reid, John, and Inbau, Fred. Truth and Deception. 2nd ed. Baltimore: Williams and Wilkins Co., 1977.

Rein, H. "Die Elektrophysiologie der Hant." In Handbuch der Hant und Geschlechtskrankherten, pp. 43-91. Edited by J. Jadassohn. Berlin: J. Springer, 1929. Cited by S. Rothman. Physiology and Biochemistry of the Skin. Chicago: University of Chicago Press, 1954.

Rickels, K.; Hesbacher, P.; Weise, C.; Gray, B.; and Feldman, H. "Pills and Improvement: A Study of Placebo Response in Psychoneurotic Outpatients." Psychopharmacologia (Berl.) 16 (1970): 318-28.

Romig, Clarence. "The Status of Polygraph Legislation of the Fifty States," Part I. Police (September 1971): 54-61.

_______. "The Status of Polygraph Legislation of the Fifty States," Part II. Police (October 1971): 54-61.

_______. "The Status of Polygraph Legislation of the Fifty States," Part III. Police (November 1971): 55-61.

Rotter, J. B. "Generalized Expectancies for Internal Versus External Control of Reinforcement." Psychological Monographs 80 (1966): 1-28.

Ruckmick, C. A. "Truth About the Lie Detector." Address to A.A.A.S., December 29, 1936.

Runkel, J. E. "Lurin's Motor Method and Word Association in the Study of Deception." Journal of General Psychology 15 (1936): 23-37.

Shapiro, Arthur K. "Semantics of the Placebo." Psychiatric Quarterly 42 (1968): 653-95.

Shmavonian, B. M.; Yarmat, A. J.; and Cohen, S. I. "Relationship Between the Autonomic Nervous System and Central Nervous System in Age Differences in Behavior." In Behavior, Aging, and the Nervous System. Edited by A. T. Welford and J. E. Birren. Springfield, Ill.: Charles C. Thomas, 1965.

Slowik, Stanley M. "Correlations Between Marijuana Use and Theft Admissions During Polygraph Examinations." Paper presented at the American Polygraph Association Convention, Los Angeles, California, 1970.

________, and Buckley, J. "Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure, and GSR Recordings." Journal of Police Science and Administration 3 (1975): 305-309.

Smith, G. Alan. "Analysis of the Voice." Unpublished manuscript, Department of Psychology, Powick Hospital, Worchester, England, 1973. Distributed by Dektor Counterintelligence and Security, Inc., 5508 Port Royal Road, Springfield, Virginia 22151.

Sokolov, E. N. Perception and the Conditioned Reflex. New York: Pergamon Press, 1963.

Sternbach, R. A. "The Effects of Instructional Sets on Autonomic Responsivity." Psychophysiology 1 (1964): 67-72.

________; Gustafson, L. A.; and Colier, R. L. "Don't Trust the Lie Detector." Harvard Business Review 40(6) (1962): 127-34.

Stroebel, Charles F., and Glueck, Bernard C. "Biofeedback Treatment in Medicine and Psychiatry: An Ultimate Placebo?" Seminars in Psychiatry 5(4) (1973): 379-93.

Summers, Walter. "Science Can Get the Confession." Fordham Law Review 8 (1939): 334-54.

Suzuki, S.; Wantanabe, T.; and Shimizy, K. "Effects of Visual Feedback by the Skin Potential Responses." Japanese Journal of Psychology 40 (1969): 59-67. Cited by G. Barland, "An Experimental Study of Field Techniques in Lie Detection." M.A. thesis, University of Utah, 1972.

Tatsuoka, M. M. Multivariate Analysis. New York: John Wiley and Sons, 1971.

Thackray, Richard, and Orne, Martin. "A Comparison of Physiological Indices in Detection of Deception." Psychophysiology 4(3) (1958): 329-39.

Timm, Howard. "The Reliability and Validity of the Psychological Stress Evaluator in Analyzing Deception." NCJRS Microfiche Program, P.O. Box 24-36, Washington, D.C. 20024.

Timm, N. H. Multivariate Analysis With Applications in Education and Psychology. Monterey, Calif.: Brooks/Cole Publishing Co., 1975.

Trovillo, Paul. "Scientific Proof of Credibility." Tennessee Law Review 53(22) (1951): 743-66.

U.S. Army Military Police School. Polygraph Examiners Training Course, 1970. Cited by G. H. Barland and D. C. Raskin. "Detection of Deception." In Electrodermal Activity in Psychological Research. Edited by W. F. Prokasy and D. C. Rankin. New York: Academic Press, 1973.

Van Buskirk, D., and Marcuse, F. "The Nature of Errors in Experimental Lie Detection." Journal of Experimental Psychology 47 (1954): 187-90.

Violante, R., and Ross, S. "Research in Interrogation Procedures." Office of Naval Research, Rep. No. 707-65. Defense Documentation Center, AD-467624, October 1964.

Vroom, Victor H. Work and Motivation. Chicago: Markham, 1972.

Walters, C. Glenn, and Gunn, Bruce. "Appraising Retailers' Use of the Polygraph." Journal of Retailing 43(4) (1964): 10-21.

Weiner, Bernard. Theories of Motivation. Chicago: Markham, 1972.

________, and Sierad, Jack. "Misattribution for Failure and Enhancement of Achievement Strivings." Journal of Personality and Social Psychology 31(3) (1975): 415-21.

Weinstein, Edwin; Abrams, Stanley; and Gibbons, Donald. "The Validity of the Polygraph With Hypnotically Induced Repression and Guilt." American Journal of Psychiatry 126(8) (1970): 1159-62.

Wenger, M. W.; Clemens, T. L.; Coleman, D. R.; Cullen, T. D.; and Engel, B. T. "Autonomic Response Specificity." Psychosomatic Medicine 23 (1961): 185-93.

Wenger, M. W.; Engel, B. T.; and Clemens, T. K. "Studies of Autonomic Response Patterns; Rationale and Methods." Behavioral Science 2 (1957): 216-21.

Worth, James, and Lewis, Bernard. "An Early Validation Study With the Psychological Stress Evaluator." Unpublished manuscript, Washington and Lee University. Distributed by Dektor Counterintelligence and Security, Inc., 5508 Port Royal Road, Springfield, Virginia 22151.