THE UNIVERSITY OF MANITOBA

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

by

MICHAEL T. BRADLEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

WINNIPEG, MANITOBA

January 1980

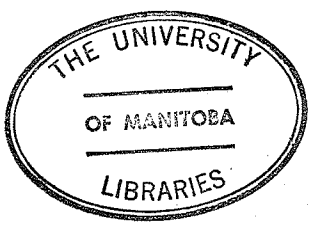
ACCURACY DEMONSTRATIONS, THREAT, AND THE DETECTION OF DECEPTION: CARDIOVASCULAR, ELECTRODERMAL, AND PUPILLARY MEASURES

BY

MICHAEL TIMOTHY BRADLEY

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY



ACKNOWLEDGEMENTS

I would like to thank my committee for their fine and scholarly aid in the production of this dissertation. Michel Pierre Janisse, my major advisor, had an immeasurable influence far beyond the confines of this work in teaching me an appreciation and understanding of research. John Schallow's careful and thorough criticism was much valued and led to an improved dissertation. Anise Mikhail's emphasis on theory broadened my perspective embedding the research product in a larger context. Roland Penner's donation of time, energy, and expressions of interest were most important and highly appreciated by me.

My wife Yvonne and children Michael and Arlyn while not contributing directly provided the environment, love and support, which at times seemed so necessary for the successful completion of this undertaking.

ABSTRACT

Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupiliary Measures

Interrogators have assumed that convincing suspects of the infallibility of the psychophysiological detection apparatus enhances accurate detection. The small amount of laboratory data available does not clearly support this assumption and data from field work has never been gathered. The present study was designed to assess the consequences of demonstrating varied levels of detection effectiveness to suspects prior to an interrogation. A mock crime involving money, threat of shock, and a field questioning technique were included to create a more realistic interrogation situation. Pupil size, pupil change, heart rate, heart change, and the galvanic skin response were used to assess detection. All subjects, half of whom had committed a mock crime, were instructed to appear innocent throughout a questioning session conducted by an interrogator blind to their actual guilt or innocence. In addition, half of the subjects in each group had also read that they would receive a painful but not permanently damaging electric shock if found guilty of the crime. Prior to the interrogation subjects received a demonstration of the physiological detection measures' effectiveness. Demonstrated effectiveness was manipulated through a series of rigged tests in which a subject chose a number from a set of cards and attempted to conceal the number. The feedback led subjects in different groups to believe they had been detected 0%, 33 1/3%, 66 2/3%, or 100% of the time. The full design was a 4 x 2 x 2 ANOVA with 4 levels of effectiveness, guilt or innocence, and threat or no threat of shock. Two interrogation techniques were used on

all subjects; the Backster test which involves accusatory questions about the crime paired with potentially incriminating questions about the suspects' past life, and the guilty knowledge test which involves questions on information only the guilty subject would recognize as relevant to the crime. As demonstrated effectiveness increased, detection ranks increased appropriately for guilty and innocent subjects using the Backster test. The results were less clear with the guilty knowledge test. Detection was at a minimum in the 0% effective condition but reached a maximum at the 33.3% level. Reasons for this result were discussed. Threat of shock affected only the rank scores derived from heart rate in the Backster test. The rank scores derived from all measures in both tests, except for pupil change in the Backster test, discriminated significantly between guilty and innocent subjects. Classification of guilty or innocent according to standard criteria from the literature was possible at significant levels for all measures but pupil size and pupil change in the Backster test. The galvanic skin response was the most accurate measure. The continuous data underlying the pupil and heart rank scores were analyzed and contributed to an understanding of the results.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

List of Tables

Chapter 1

Responses of physiological systems innervated by the autonomic nervous system (ANS) have been used as indicies of emotional, sensory and mental activity (Sternbach, 1966). Generally, these responses have been considered involuntary or at least difficult to control when a subject is exposed to meaningful stimuli and to the degree that this is true represent objective physiological manifestations of ongoing psychological phenomena.

"Lie detection" methods of police, government and employers have relied heavily on the measurement of autonomic responses as indicants of a suspect's attempts to conceal, mislead or lie during an interrogation (Inbau and Reid, 1966). Usually several measures are taken and recorded on a device known as a polygraph. Those who use the polygraph for applied or practical purposes claim that it is a scientific assessment of a suspect's guilt or innocence. However, many polygraph experts have difficulty when asked to verbalize just what the specific indicants of deception are. Overall, the judgement process appears to be very subjective (Davis, 1961). To the scientist this state of affairs is unsatisfactory. One preliminary task should be aimed at objectively delineating the responses concomitant with deception. Effort is needed to explain the so called physiological "lie response" in a way more congruent with the theories of psychophysiology. Further work is needed in identifying specific aspects of situations and individuals which make the detection of deception possible.

History of Lie Detection

Historically the belief that certain types of cognitive activity are

accompanied by perceptible physiological or behavioral alterations has been prevalent. This is especially true with regard to the detection of deception or "lie detection" (Boring, 1942; Larson, 1932; and Trovillo, 1939). Persons suspected of attempting deception were often subjected to special ordeals. It was believed that only a person not guilty of deception could pass these ordeals. Zoroaster proved the truth of his words by touching a red hot iron to his tongue nine times without scorching it, ancient Chinese were required to speak with their mouths full of rice to prove their innocence. In both cases if emotion interfered with salivation the suspects might have failed in their tasks. Witch doctors sometimes leapt at suspects smelling them feverishly. Distinctive odors indicated guilt and it is possible that the fear of being caught produced such odors. Another test had the suspect immerse his arm in boiling water and if it blistered the next day, he was considered guilty.

How effective these ordeals were in discriminating between guilty and innocent individuals is debatable. Some ordeals, such as the one reserved for the Roman Catholic clergy in the middle ages, were very unlikely to cause anyone to be declared guilty. The accused clergyman was instructed to eat a piece of barley bread and cheese while other clergymen prayed for an angel to stop the accused's throat if he was guilty. There is no recorded instance of a priest having been choked in this manner (Trovillo, 1939). Another ordeal involved the use of a very accurate balance beam. The accused was placed in one balance pan of the scale while the other side was carefully counterbalanced. A groove was filled with water for the purpose of detecting the slightest deflection either way. The suspect then stepped out of his end of the scale, listened while a judge exhorted the balance to discover the truth, and finally got back in. If he were lighter than before

he was considered innocent. Such a test depends more on how long the judge takes to make his speech than guilt or innocence since the body undergoes a constant loss of water of about 12 grams per hour. A long speech would free the accused (Trovillo, 1939).

More modern and scientific investigation into behavioral and physiological differences accompanying deception began before the turn of the century. As early as 1877 Mosso had observed the effects of fear on blood pressure and carried out investigations on physiological reactions to emotional stimuli. He was encouraged by his tutor Lombroso (Trovillo, 1939) who later assisted police in identifying criminal suspects through the use of a "hydrosphysmograph" to record blood pressure changes during interrogation (Inbau and Reid, 1953). Blood pressure was measured with this instrument by having the suspect place his hand in a vessel of water topped by a rubber seal. Pulsations of blood caused water level changes which affected an attached air filled tube and these changes were recorded on a revolving drum. In 1906 Jung studied differential reaction times to stimuli on which subjects hoped to deceive the experimenter (Orne, Thackray and Paskewitz, 1972). Marston (1917) used a sphygmomanometer to record blood pressure during questioning and reported 96% accuracy in detecting deception with the device. Luria (1932) showed that psychomotor responses can be impaired while the subject is lying. He required the subject to hold one hand steady while depressing a plunger with the other. Munsterberg (Trovillo, 1939) pointed to the possibility of using the galvanic skin response (GSR) for lie detection purposes. Larson (1921) put together an instrument capable of taking blood pressure, pulse and respiration all at once. Finally Keeler (1930) developed the polygraph to simultaneously measure respiration, relative blood pressure and the GSR. Although these measures have remained the major physiological measures in

"lie detection" since that time (Davis, 1961) it is interesting to note Keeler's observation on his apparatus:

> To begin with, there is no such thing as a 'lie detector'. There are no instruments recording bodily changes, such as blood pressure, respiration, or galvanic reflex, that deserve the name 'lie detector' any more than a stethoscope, a clinical thermometer, or a blood count apparatus with a microscope can be called an 'appendicitis detector'.
>
> However, deception, guilt, or innocence can be diagnosed from certain symptoms just as appendicitis, paranoia, or any other physical or mental disorder can be diagnosed. In every case, the examiner must make his diagnosis from tangible symptoms, using whatever mechanical aids he has at his disposal. (1934, p.153)

In spite of Keeler's caution Barland and Raskin (1973) report that the use of the polygraph as a "lie detector" has grown tremendously since the early 1950's for both criminal and commercial applications in North America and throughout the world. This growth has alarmed Lykken (1974) who sees lie detection as a psychological testing technique ignored by psychologists and practiced by individuals who ignore psychology.

Field Work

Field work or the practical application of lie detection in criminal investigations has dominated much of the work and literature in the detection

of deception. Orne, Thackray, and Paskewitz (1972) mention that investigations in the field have not been characterized by a systematic scientific approach. Practitioners involved in lie detection specifically try to structure a situation to achieve the goal of a successful diagnosis of deception. How they achieve this goal is, in part left up to the individual discretion of the investigator. Further, not many reports have attempted to validate findings, basically because in some cases the actual guilt or innocence of a suspect may never be known (Barland and Raskin, 1973). Field work is characteristically more of an art than a science. Inbau, Moenssen and Vitullo (1972) stressed that since the polygraph technique in criminal investigation involves a diagnostic procedure rather than a mechanical operation, an examiner must be intelligent and well educated, with suitable personality characteristics "to get along well with others and to be persuasive in his dealings with them". (1972, p. 153)

The actual recording of autonomic responses during questioning represents only the final part of a structured interview and interrogation session aimed at convincing the suspect of the infallibility of the lie detector. To accomplish this, individual interrogators may alter their style, mannerisms, subtle cues and tone of voice as they see fit. In addition many resort to demonstrating 100% detection accuracy with a rigged card test (Orne, Thackray, and Paskewitz, 1972). The scoring of the recorded physiological responses has not been specified in quantitative terms. Inbau, Moenssen and Vitullo (1972) write that in approximately twenty-five percent of the examinations conducted by a competant polygrapher, truthfulness or deception is so clearly disclosed that any layman could be shown the results and convinced of their significance. However, in sixty-five percent of the cases, the indicators are

sufficiently subtle as to require expert interpretation. This expert interpretation is carried out in the context of an investigation where the interrogator may already possess other information including the investigative knowledge and conviction of his colleagues as to the suspect's guilt. This alone may be a powerful source of bias affecting the subjective interpretation of objective records (Orlansky, 1962).

The problems of field work reviewed above may be discussed in terms more familiar to psychologists as problems of reliability, validity, and experimenter (examiner) bias.

Reliability. Barland (1972) sent polygraph charts from 72 subjects to 6 experienced field examiners for an evaluation of guilt or innocence using a numerical scoring technique. The resultant interrater correlation over 1080 pairs of judgements was .86 with the combined scores from respiration, skin resistance, and cardiovascular (blood pressure) responses which on an individual basis yielded reliabilities of .64, .90, and .76, respectively. Edel and Jacoby (1975) used charts from actual field situations (job interviews). The task for ten experienced evaluators was to agree as to whether a specific physiological reaction occured or failed to occur for each question. When each evaluator judged a chart never previously seen by either the percentage agreement was 96 for cardiovascular responses, 91 for GSR and 96 for respiration. When one evaluator had conducted the interview and thus had a first hand knowledge of the subject, the interrater reliability remained essentially the same. Horvath (1977) had 10 field trained examiners make blind judgements of charts from 112 criminal suspects. He found an interevaluator reliability coefficient of .89 for suspects whose guilt or innocence had been subsequently verified and .85 for suspects whose

cases were not verified. Thus, although criteria for evaluating responses have not been systematically explained there is in these few studies at least an indication of fairly high reliability among examiners. Since reliability is fundamental to a test and since there are numerous schools of polygraphy, eight recognized by the American Polygraph Association (Barland, 1973), many more examinations, similar to Horvath's (1977) study, are needed and could even be used to check on each evaluator's work.

Validity. Given reasonable levels of reliability, does the polygraph diagnostic procedure succeed in discriminating the guilty from the innocent? Certainly many examiners have published statistics reporting error rates of only 2% or 1% (Reid and Inbau 1966). Arther (1965) claimed incorrect decisions in only .05% of all his examinations making only 9 mistakes in 17 years! However, there are several problems with these kinds of estimates. For example Lykken (1974) has stated "Estimates given by professional polygraphers on the basis of their own experience are essentially worthless" (p. 734). Barland (1973) in conversation with field examiners found that many automatically assumed their decision verified unless specific evidence to the contrary was presented to them. This attitude ignores the problems found in establishing ground truth, that is whether a suspect is actually guilty or not. The inadequacies of various criteria such as confessions, police investigations and jury decisions are known. In many cases decisions based on job interviews have been included where there is no subsequent act or behavior to justify the hiring or failure to hire an interviewee because of his physiological responses to questions. Verified errors have been divided by the total number of cases examined whether verified or not thus the error rate may be deflated because actual errors amongst the

unverified cases are not only not included in the error total but are divided into the verified errors. As previously stated the criminal investigator has and uses a great deal more information in making his judgement than just the polygraph record. Confessions elicited while being interrogated or due to the threat of polygraph interrogation have often been counted as successful polygraph diagnoses (Orne, Thackray and Paskewitz, 1972).

A few respectable validity studies have been reported. Lyon (1936) found 40 cases in a random selection of 100 which could be confirmed by outside criteria and of those 40 which included seven truthful and thirty-three deceptive judgements no incorrect decisions were found. Bersh (1969) had a panel of four experts evaluate all the facts of each case and then decide independently the guilt or innocence of the person without knowing the outcome of the case or polygraph examination. When all four panel members were unanimous in their judgement they agreed with the polygraph examinations 92.4% of the time. Of course such a panel can still make errors of classification but this remains a practical way of handling the independent validation criteria requirement. Barland (1973), who has an extensive background in field work, in his review of the literature suggested that the polygraph technique may be accurate in somewhat more than 90% of the cases where a definite decision is made by the examiner, if the test is made under ideal conditions. Horvath (1977) had ten field trained polygraph examiners evaluate the polygraph records of 112 criminal subjects. When ground truth had been established by confessions, correct calls averaged 64.1%. Agreement with the original examiners judgement was 62.1% for unverified cases. These substantially lower validity rates were obtained because all other investigative details were excluded and judgements were made on the polygraph records alone thus suggesting that other variables concerned

with the examiner's perception and knowledge of a suspect play a large role in the final decision.

Examiner bias. Orne (1973) has argued that the primary variables in lie detection are psychological in nature. The techniques involved in lie detection give ample leeway for such influence to be involved. During the pretest interview the examiner may form an impression that will influence his interrogation and interpretation of the polygraph records while at the same time influencing the suspect as to how credible he might appear. Barland (1973) pointed out that two schools of thought exist: (1) all information, behavioral cues and impressions gathered throughout the interrogation should be brought to bear during the interpretation of the records, (2) all extraneous information while sought out during the examination should be excluded during the polygraph chart interpretations. Certainly for the second proposition some of this information must be hard to ignore and not without influence in the final judgment process. Horvath (1977) pointed out another kind of bias in interpreting polygraph charts. False-positives, calling an innocent suspect guilty, occured more frequently than false negatives. Lykken (1978) reported an instance in which the interrogator classified 98% of the guilty as guilty at the expense of classifying 55% of the innocent as guilty. Part of the complexity of determining reliability, validity and examiner bias in field work is that there are two major facets that must be assessed. First, as Lyon (1936) did, the judgements from the whole interrogation procedure including the interview and polygraph recording, may be evaluated. Examiner bias in interviewing and interpreting can be fully operative but the result may be a highly effective and accurate assessment. Second, only the physiological records could be evaluated, so that the

adequacies and failings of the measurement device alone may be understood. This second approach may not eliminate the original examiner's ability to "make" a response happen by careful management of a pretest interview but it does mean a judgment of the record in absence of extraneous information. Also it does not eliminate judgment bias, e.g., the favored tendency to classify most responses as guilty, (Horvath, 1977) but objective scoring criteria could be developed to correct this.

Questioning Techniques in the Field

Practitioners in criminal lie detection have developed and come to rely on certain techniques and procedures to present questions during an interrogation. These can be classified as either guilty person or guilty knowledge techniques.

The guilty person technique. The guilty person technique (Lykken, 1960) or "undisguised question method" (Burack, 1955) is a modification of a direct confrontation question such as "Are you guilty?" The suspect is asked several questions, some relevant, some irrelevant to the crime. If responses to relevant questions differ from those to irrelevant questions the suspect is considered guilty. Unfortunately, questioning a suspect about whom he may have killed can yield responses interpretable as deceptive even though the suspect is innocent (Burack, 1955). Orne, Thackray and Paskewitz (1972) have called for more investigation of "false positives" where because of innocent fear, for example, a suspect may respond as if guilty.

Inbau and Reid (1966) have suggested the addition of control questions which are irrelevant to the crime being investigated but are questions to which a suspect will probably respond with a lie. Specifically, in the

Reid Control Question Technique, several relevant questions are asked along with several control questions. The particular control questions are worked out for each individual suspect by the interrogator in a pre-lie detection test interview. This interview is crucial for the wording of control questions in such a manner as to ensure that all suspects will lie or be troubled by them. This strategy is to divert the attention of the innocent suspect from the relevant questions. The guilty suspect should still remain more concerned about the relevant questions. Questions such as "Have you ever taken anything of value?" might be used. It is emphasized that all questions should be answerable by a simple "yes" or "no" and that doubts or ambiguities or reservations will be reflected in the polygraph readings (Barland, 1973). Suspects often do confess to misdemeanors but still remain uncomfortable with the question or questions.

If the lie reaction to the control question is as great or greater than the response to the relevant question then the suspect is considered innocent. If not he is considered guilty. An additional question, called the guilt complex question, about a fictitious crime of the same seriousness may be asked and if the suspect's reactions to this question are equal to or greater than the response to the relevant question then the suspect is considered innocently nervous. Backster (1969) introduced a time specification on the control questions so that they clearly did not pertain to the crime at hand.

Podlesny, Barland, and Raskin (1976) generated a Backster question sequence for a mock crime experiment and it is displayed in Appendix A. The first three questions serve primarily as buffers to habituate initial responding. Questions 4, 6 and 8 are control or irrelevant questions while

questions 5, 7 and 10 are the relevant questions. Question 8 is the guilt complex question. The Reid Control Question technique would typically use control questions starting with "Have you ever....?" and thus have no time specification.

Podlesny, Raskin and Barland (1976) interrogated experimental subjects who were either guilty or innocent of a mock crime by either the Backster or Reid Control Question techniques. Excluding inconclusive tests, 94% of the decisions were correct with the Backster technique while 83% were correct with the Reid technique. The differences were not significant suggesting that time specification has no effect. The guilt complex question was significantly less effective than both control question techniques in discriminating guilty from innocent.

The above techniques were developed on the assumption that a "lie response" was being measured rather than an emotional reaction to the content and implications of such questions. The literature does not support the conception of a "lie response" (Kugelmass, Lieblich and Bergman, 1967). Day (Note 5) has stated that a lie is not a critical factor in causing a detectable physiological response. Orne, Thackray and Paskewitz (1972) suggest that "no specific physiological responses are pathognomic of lying" (1972,p. 775).

The disguised question technique. Lykken (1960) has advocated a detection technique which removes the burden of a special "lie response" as being involved in physiological detection. The technique, actually anticipated by Lee (1953) with his "secondary relevant questions", termed the "disguised questions test" (Burack, 1955) or the "guilty knowledge technique" (Lykken, 1960) relies on the differential impact of questions

on knowledge only the guilty person could have. Since the innocent suspect could have no information about certain aspects of the crime it is assumed that questions about those aspects would be considered irrelevant and non-threatening to him and thus not evoke strong reactions. Several questions may be set up in a multiple choice format (Lykken, 1959) so that the probability of large responses to all correct (guilt implicating) questions by chance alone becomes very low. The suspect does not need to reply but merely listens to the alternatives for a differential recognition response to occur (Lykken, 1959). Thus, no lie need be involved in guilt detection. The basic assumption (Lykken, 1974) that a guilty suspect will show a stronger autonomic response to what he recognizes as a significant alternative than he would have without such knowledge seems much more scientifically sound. Unfortunately, there are limitations on this method. The details may simply be unavailable to the interrogator or on the other hand be so widely known that the innocent subject is aware of them.

Measures Used in the Field

According to Inbau and Reid (1966) the ANS responses measured for the detection of deception in field work have been generally limited to respiratory responses, cardiovascular responses, and the galvanic skin response. The following review concentrates on these measures.

Respiratory responses. Inbau and Reid (1966), acknowledged experts in the lie detection field, consider respiration to be their most reliable measure. Unfortunately, although respiration measures have long been used in the detection of deception (Trovillo, 1939), field workers have not identified a specific response as an indicant of deception. Instead they have tended to judge any marked change from the baseline of breathing rate (cycles per second) and/or amplitude as indicative of deception. Thus, a

suspect could speed up or slow down his breathing rate and/or increase or decrease amplitude and any of these responses would arouse the interrogator's suspicion. Davis, (1961) has noted that respiration in the early part of an interrogation session is often irregular and as such is not a good indicator of deception responses. However, in a longer test session discrimination becomes much better. The early irregularity is thought to be a response to the general interrogation situation (Davis, 1961).

Horvath and Reid (1972) may have overcome some difficulties associated with respiratory measures. They found, in a field investigation, that differences between respiratory responses to critical and neutral questions were enhanced when the suspects were requested to remain mute during the interrogation session. The enhancement was attributed to the elimination of sources of variability associated with talking. Horvath and Reid (1972) have outlined several respiratory irregularities associated with an audible answer. Having the suspect remain mute eliminates distortions where a suspect may either inhale or suppress inhalation just to give an audible answer. Answers given at the height of inhalation can produce substantial distortion. Subjects who prepare for an audible answer by physical movement cause distortion, as do those who loudly bellow their answers, feel compelled to talk in addition to a "yes" or "no", or have throat irritation when they respond. Unfortunately, Horvath and Reid (1972) have only presented selected samples of the polygraph record. These samples illustrate instances of relatively dramatic differences between neutral and critical questions but Horvath and Reid (1972) failed to supply data of the over all rate of detecting deception responses.

Cardiovascular responses. Blood pressure is the measure relied on by most practitioners in the field (Davis, 1961). A measure of relative blood pressure is obtained by inflating an arm or wrist cuff to a point equal to the pressure half way between systolic and diastolic blood pressure levels. Unfortunately, the apparatus can be painful and dangerous since the cuff pressure far exceeds vein pressure and does not allow the blood to return from the arm or the hand. It is quite possible that the pain produced using the pressure method may result in reactions of other autonomically controlled responses including blood pressure itself. In spite of this, Davis (1961) has reported that blood pressure is one of the better indicators of deception. A drawback with the blood pressure measure is that discrimination is poor (almost nil) in the early part of the session. However, it does improve greatly later on (Davis, 1961).

The galvanic skin response. It is not clear how useful the GSR is for lie detection in the field. Davis (1961) has concluded that the GSR is the best indicator of deception in short time intervals but poor in longer questioning periods. Inbau and Reid (1966) criticized the GSR because they were unable to obtain a high degree of accuracy with it. The GSR is easily triggered but slow in recovery and in a routine examination the next question may be introduced before recovery to baseline is complete. In a series of questions the GSR may adapt out simply because each new question is asked before the GSR has returned to the level it was at when the prior question was asked.

The opposite problem of failure for the GSR to occur has also been brought up. Woodworth and Schlosberg (1965) have documented evidence suggesting that the GSR is an inadequate measure during strong emotion. Part of this evidence

is dependent on Darrow's (1936) findings that adrenalin, contrary to the expected effect, seems to inhibit the GSR. Such evidence would be consistent with the claim of Inbau and Reid (1966) that the measure is poor for field work since many suspects could be highly emotional.

Another possibility for the failure of the GSR in field work has been raised by Ferguson (1966). He reported that a commonly used field instrument employing a self centering pen feature reduced the effectiveness of the instrument 75-80 percent. However, this may not be the whole story since Gustafson in an unpublished paper (cited in Orne, Thackray and Paskewitz, 1972) found no difference in detection with a commonly used field instrument, the Stoelting Deceptograph, in comparison to a Beckman Dynograph.

Laboratory Research

In general, although the polygraph "lie detector" is construed as a scientific instrument (which it is), ample room in its application is left for uncontrolled and nonystematic sources of variance. Previously, it was mentioned how various factors relegate the field interrogation procedures to an art. In spite of this psychophysiological detection of deception (PDD) situations are appropriate for laboratory investigations for at least two reasons: (1) to increase our understanding of variables normally uncontrollable in field work (Podlesny, Raskin and Barland, 1976), (2) to use as a paradigm for the study of "cognitive, motivational, attentional, and learning factors involved in the physiological responsivity to verbal stimuli." (p. 780, Orne, Thackray and Paskewitz, 1972).

Podlesny, Raskin and Barland (1976) have noted the major advantages and disadvantages associated with the laboratory study of PDD. Ground truth, that

is the actual guilt or innocence of the suspect, may be specified. Variables may be controlled and manipulated. Large quantities of high quality physiological data may be obtained using sophisticated laboratory equipment instead of portable field equipment. However, all laboratory PDD paradigms are only simulations of field situations. This unavoidable element of artificiality may reduce the generalizability of laboratory results to field contexts.

Detection in laboratory research. Just as in field work, laboratory approaches to the detection of deception can be divided into deception tests and information tests. Deception tests are based on the assumption that differential responsivity occurs when a suspect is deceptive and such responsivity may be the result of changes in arousal, attention, emotion or information processing (Podlesny, and Raskin, 1977). The relevant-irrelevant and control question techniques have already been introduced in the previous section on field work. The underlying assumption for information tests is that the possession of knowledge about which is the critical item, usually embedded among matched items, is sufficient to evoke a differential physiological response (Lykken, 1974).

Laboratory Paradigms in Detection

The card test is structured so that the subject is known to be guilty of attempting to deceive the experimenter about a particular item of information. Cards or numbers are presented to the subject who is instructed to choose one, keep his choice in mind and answer "no" to all questions of the nature "Is it card (or number)_____?" The result of this format is that the subject tells the truth on all cards except for the one selected, which in following the instructions he automatically lies about.

The experimenter's task is to identify the specific card or number on which the subject is attempting deception. This procedure has been common because of its ease and convenience of administration (Alpert, Kurtzberg, and Friedhoff, 1963; Block, Rouke, Salpeter, Tobach, Kubis and Walch, 1952; Burtt, 1921; Geldreich, 1941; Landis and Wiley, 1926; Langfeld, 1921; Obermann, 1939; Van Buskirk and Marcuse, 1954) but it also bears the least resemblance to field work (Podlesny and Raskin, 1977). Gustafson and Orne (1964) formulated a guilty person version of the card test by including blanks in the deck. In this situation the interrogator's task is not to identify the specific bit of information that each subject has but to find which individuals have information to conceal.

Some attempts at making detection studies more realistic or involving have included using items of personal significance (Cutrow, Lucas, Parks and Thomas, 1972; Lykken, 1960) and playing the role of an espionage agent (Clark, Note 4). Some researchers have combined both the personal items and espionage features into one design (Orne and Thackray, 1967; Thackray and Orne, 1968a, 1968b). Thackray and Orne (1968b) found the personally relevant material more detectable than the code words which gained their importance only in the context of the experiment.

Mock crime situations have been devised in an attempt to create realism and approximate field conditions. Mock crimes involve subjects in a set of behaviors construed as criminal for the purpose of the study and have ranged from the simple (Cutrow, Parks, Lucas and Thomas (1972) to the very elaborate (Davidson, 1968). Lykken (1959) had some subjects "murder" an individual whom they had accused of cheating in a poker hand while others "stole" property from an office desk. Barland and Raskin (1975), Podlesny Raskin

and Barland (1976) and Podlesny, and Raskin(1978) had subjects take money or a heavy gold ring from a desk drawer. Raskin and Hare (1978) had prisoners steal $20 from a prison office. In one experimental setup Berrien (1942) had subjects kill a rat with a knife. All of these studies have reported high rates of detection.

Measures used in Laboratory Research

Measures used in laboratory detection of deception have not been limited to those popular in the field. A much wider array of measures both autonomic and nonautonomic have been employed. Cutrow, Parks, Lucas and Thomas (1972) have used eye blink latency, eye movements and voice latency. Several cardiovascular measures have been used including blood pressure (Marston, 1921), plethysmographic monitoring of pulse (Brown, 1967) and heart rate (Berkout, Walter and Adey, 1970). Obermann (1939) has used the electroencephalogram and pupillary dilation was investigated by Berrien and Huntington (1943).

Cardiovascular Responses

Blood pressure The early laboratory studies of Chappell (1929) and Marston (1921) reported high rates of detection, 87% and 94% respectively, using a blood pressure measure. However, more recently investigators such as Thackray and Orne (1968a) have not been able to detect deception using blood pressure measures. Orne, Thackray and Paskewitz (1972) have suggested that since earlier studies used manual devices to measure blood pressure and the measurement of the response was taken immediately after each question the response was time locked to the question more closely than with modern devices which inflate the cuff automatically at fixed intervals. Early findings with manual devices were not universally positive in any event.

Landis and Gullette (1925) reported only a 55% detection rate against a chance level of 50%.

Another measure of blood pressure has involved the use of the cardio cuff only partially inflated to a level between diastolic and systolic blood pressure. Though somewhat less painful and dangerous than the complete occlusion of arteries, pressure must be released periodically or the duration of questioning be reduced to prevent problems with the limb (Davis, 1961). Kugelmass, Lieblich, Ben-Ishai, Opatowski, and Kaplan (1968) reported significant detection rates with a card test in field conditions using this device. Barland and Raskin (1975) found significant discrimination between guilty and innocent subjects (psychology students) only on the first of three charts using this measure.

A low pressure cuff has been recently tested by Podlesny, Raskin and Barland (1976). Electronic amplification has made it possible to operate at cuff pressures of 50 mm/Hg instead of about 90 mm/Hg and has eliminated the problems due to occlusion of blood flow. They found that innocent subjects responded more to the control questions than the relevant questions and thus were judged correctly but guilty subjects did not show differential responses. The low pressure cuff was not effective with the Guilty Knowledge technique. Podlesny, Raskin and Barland (1976) suggested that raising the cuff pressure to 70 mm/Hg yields better discrimination according to their pilot data.

Other measures are available from the blood pressure cuff such as pulse rate and pulse amplitude. Podlesny and Raskin (1976) reminded us that when cuff pressure is below mean arterial pressure, increases in blood pressure decrease pulse amplitude and when cuff pressure is higher then increases in blood pressure increase pulse amplitude.

Heart rate Kugelmass and Lieblich (1966) and Kugelmass, Lieblich, Ben-Ishai, Opatowski and Kaplan (1968) have found that heart rate discriminates

no better than chance. However, analyses were limited to only 5 seconds following a response and Clark (Note 4) found that guilty subjects began a deceleratory phase after 7 to 8 seconds that allowed a 63.3% correct hit rate which was significant. Ellson, Davis, Saltzman and Burke (1952) cited in Orne, Thackray, and Paskewitz (1972) also noted a delayed deceleration with deception. Cutrow, Parks, Lucas and Thomas (1972) reported significant detection with increases in heart rate but also noted that the results for certain individuals showed consistent heart rate decreases across deceptive conditions. Podlesny, Raskin, and Barland (1976), Raskin and Hare (1978) and Podlesny and Raskin (1978) in mock crime contexts found a deceleration for guilty subjects which reached a peak for discrimination from innocent subjects around 11 seconds after relevant questions. The initial response is an acceleration for both guilty and innocent subjects, but by 5 seconds a differentially greater deceleration for guilty subjects began.

Other cardiovascular measures Podlesny, Raskin and Barland (1976) Raskin and Hare(1978) and Podlesny and Raskin (1978) found significant differences between guilty and innocent subjects, responding to relevant and control questions, for finger blood volume, finger blood volume response time and finger pulse amplitude. Cutrow, Parks, Lucas and Thomas (1972) also found finger pulse amplitude to discriminate between guilty and innocent subjects although Thackray and Orne (1968a) found detection rates no better than chance with the same measure. In the same experiment using finger blood volume Thackray and Orne (1968a) found significant differences.

Respiration

In 1914 Benussi (Davis, 1961) experimented with respiration as an

indicator of deception. Benussi developed the I/E ratio, time of inspiration divided by time of expiration, and found that the ratio increased during lying. Except for a few early studies, Burtt (1921) and Landis and Wiley (1926), this ratio has been neglected in both laboratory and practical work because it is difficult to delineate between periods of inspiration, expiration and rest (Davis, 1961). However, more readily quantifiable features of respiration such as breathing rate (cycles per second) and amplitude have been compared in the laboratory.

Respiration cycle time. Cutrow, Parks, Lucas, and Thomas (1972) found longer respiration cycle times (RCT) during deception. Barland and Raskin (1975) also found that guilty subjects show increased RCT after being asked relevant questions but not after control questions. However, Thackray and Orne (1968a) failed to find an effect and Raskin (1975) found that a guilt by question type interaction was produced mainly by innocent subjects slowing respiration cycle time to control questions and speeding it up to relevant questions. Guilty subjects showed no change. Podlesny, Raskin and Barland (1976) and Podlesny and Raskin (1978) reported no differences for RCT between guilty and innocent subjects. However they suspected instrumentation problems with a new strain gauge device that picked up a high level of noise (mainly, heart rate pulsations) and may have masked any effects. In a study using pneumatic bellows transducers, Raskin and Hare (1978) found that this measure did discriminate between the guilty and innocent.

Respiration amplitude Cutrow, Parks, Lucus and Thomas (1972) found that respiration amplitude (RA) decreased during deception. Thackray and Orne (1968a) reported similar findings but only for stimuli of personal relevance. Barland and Raskin (1975) did not find a significant difference

when comparing RA following guilt and control questions but did find differences when the changes from pre and post stimulus levels were compared. Podlesny, Raskin and Barland (1976) reported no differences using RA but again, as mentioned above, suspected their instrumentation.

Electrodermal Measures

Skin resistance In contrast to skin resistance response (SRR) or its reciprocal the skin conductance response (SCR) being regarded as a very poor indicant of deception in the field (Inbau and Reid, 1966) it has been one of the most reliable and sensitive indicants in the laboratory (Cutrow, Parks, Lucas, and Thomas, 1972; Davis, 1961). As Podlesny and Raskin (in press) point out "Virtually no reported experiment has failed to find significant discrimination between truth and deception using measures of SCR and SRR." (p. 26). Laboratory investigators (Clark, Note 4; Cutrow, Parks, Lucus and Thomas, 1972; Podlesny, Raskin and Barland, 1976; and Thackray and Orne, 1968a) who have used multiple ANS indices agree that the SRR is superior to other variables in the detection of deception. Orne, Thackray and Paskewitz (1972) report that field investigators think the SRR may be too responsive to any stimulus in a real life situation where the emotional or motivational level of the subject may be very high. However, Kugelmass and Lieblich (1966) reported successful detection using a group of police cadets who had been told that their future careers depended on their ability to avoid detection. Kugelmass, Lieblich, Ben-Ishai, Opatowski and Kaplan (1968) found that prisoners given a card test prior to a real interrogation were readily detectable through the SRR. Thus, it is apparent that the SRR can be valuable under high stress conditions.

Podlesny, Barland and Raskin (1976) and Podlesney and Raskin (1978)

examined a variety of features of their electrodermal measures. SCR rise time did not discriminate among the guilty and innocent but both SCR recovery half-time and SCR recovery half time width did.

Skin potential response Whereas the SRR is obtained by the application of an external current through electrodes on the skin surface the skin potential response (SPR) is obtained by measuring the voltage difference between the skin's surface and interior (Edelberg, 1972). A relatively inactive area (for example, the earlobe) is chosen for a reference site rather than actually inserting an electrode in the skin. Thackray and Orne (1968a) found significant discrimination between guilty and innocent subjects in their mock agent paradigm with the SPR. Podlesny, Raskin and Barland (1976) using both the initial negative and following positive waves of the SPR response found that guilty subjects gave a larger negative SPR to relevant questions while innocent subjects responded to control questions. No such effect was found with just the positive component but it did (as had the negative component) discriminate successfully in the guilty knowledge technique.

Ocular Responses

The eye has been another organ studied in deception. Berrien (1942) counted eye movements during deception and found that the gaze of guilty suspects became more fixed or steady during a lie. Berrien and Huntington (1943) have detected deception by measuring increases in pupil size. They found a pupillary response with a slow negatively accelerated dilation, lasting 1-5 seconds, followed by a rapid constriction yielded correct discrimination in 70% of the cases.

Generally, except for the aforementioned study, the pupillary response has been neglected in the detection of deception until recently. Bradley

and Janisse, in a series of experiments, used the pupil response in the card test paradigm, to investigate verbal response parameters and habituation (Note 1), individual differences (Note 2) and varied levels of subjective impressions of the probability of detection (Note 3). They found (a) no differences between verbalizing or not verbalizing during PDD, (b) a reduction in successful detection over repeated trials, (c) differences in detectability between high and low Machiavellians and (d) higher detection rates with subjects uncertain as to whether they would be detected or not. Clark (Note 4) used a secret-agent paradigm to investigate and compare the pupillary response with heart rate and GSR. All three measures yielded significant detection rates but the correlations among them were low. The only significant relationship was that of the pupillary response with heart rate ($r=-.18$). The most accurate measure for detection was the GSR, followed by the pupillary response and then heart rate.

Miscellaneous Measures

A wide variety of measures have been tested in the laboratory with varying degrees of success. Cutrow, Parks, Lucus and Thomas (1972) found evidence that eyeblink latency and voice latency could successfully be used to discriminate among guilty and innocent items. Obermann (1939) presented some suggestive data that the electroencephalogram could be used in detection. Thackray and Orne (1968a) had limited success measuring oxygen saturation levels of the blood. Podlesny and Raskin (1977) have reviewed the literature in electronic voice analysis and concluded that a better determination of the specific measures to be used is necessary. They could find no scientific evidence that voice analysis was as effective as standard polygraph procedures, "and little evidence that the results

exceed chance levels." (p. 38, in press). Horvath (1978) in a comparison of the Dektor Psychological Stress Evaluator, a popular voice device, with the GSR found the voice detectors performance at chance levels while the GSR performed beyond chance.

In summary there are many questions still to be asked about the measures used in the detection of deception. The GSR, the most accurate measure in the laboratory does not seem to perform as well in the field (Barland and Raskin, 1973). Factors such as the additional stress of the field situations or the interaction with the effects of other measures (e.g. the pain caused by a blood pressure cuff) do not seem to degrade the performance of the GSR (Kugelmass, Lieblich, Ben-Ishai, Opatowski and Kapplan, 1968). Possibly field examiners have failed to adequately exploit the GSR. Cardiovascular measures, except for blood pressure which has mixed reports of success, have not been very accurate in detecting deception in the laboratory. Respiratory measures have been difficult to objectively quantify while still maintaining the accuracy of detection that experienced field workers such as Inbau and Reid (1966) claim to obtain by subjective judgement. And, as mentioned above, work on the pupillary response has begun in at least two laboratories (Janisse, in press; Clark, Note 4). Many other measures (e.g. eyeblink latency, oxygen saturation) have been introduced to laboratory studies and some hold promise for both practical and theoretical work.

Possibly the most exciting work involves the weighted contribution of measures to improve prediction (Cutrow, Parks, Lucus and Thomas 1972). This could be used to advantage whenever measures successfully discriminate but still retain some degree of independence. Clark (Note 4) pointed out that some measures may be more sensitive to certain factors and knowledge of this information could be used in refining the weightings given to any measure under specific conditions.

Laboratory Research and Field Work Compared

Caution is necessary when drawing inferences from the laboratory to the field, since the consequence of the test in a criminal interrogation or employment interview can be very important to the individual involved while an interrogation in the laboratory may be of almost no importance to the subject.

The degree to which laboratory and field situations may differ is considerable and these differences may have great effect on the results of investigations. The level of affect or arousal may be much higher in the field. Kugelmass and Lieblich (1966) presented data which showed that the mean pulse rate for 22 suspects undergoing an actual criminal interrogation was 91.6 beats per minute while the mean pulse rate for police cadets taking a card test was 83.7 beats per minute in both an experimentally induced stress condition and a more neutral condition. Heart rates for the police cadets were available from a previous medical examination and the average was 72.1 beats per minute. Thus while the affective levels may be higher in a detection experiment than in a medical examination they are not as high as in criminal interrogation.

Another apparent difference occurs with respect to levels of motivation. Investigators in the field have assumed that the suspect's concern or motivation to avoid detection results in enhanced lie responses (Horvath and Reid, 1972). Gustafson and Orne (1963) supported this contention in a laboratory study using college students who were told that only those with superior intelligence and emotional control could avoid detection. Students motivated to avoid detection by these instructions were easier to detect. However, neither Day (Note 5) nor Lieblich, Naftali, Shmueli and Kugelmass

(1974) found any effect of varying levels of motivation. Lieblich et al. (1974) used instructions concerned with intelligence and emotional control similar to those employed by Gustafson and Orne (1963). One difference between the two experiments was the use of the subject's own name as the critical stimulus by Lieblich et al. which resulted in such high detection rates over all conditions that a difference could not emerge. Davidson (1968) also failed to find motivational differences with monetary incentives ranging from 10¢ to $1.00 and $25.00 to $50.00 because again the detection rates were uniformly high in all groups.

Barland (1973) listed a host of possible differences between the field and laboratory. Some of the more salient involved attitudes towards the examiner, purpose of the exam, expectations contingent upon various outcomes, the rapport between the examiner and suspect, the short term nature of stress in the experiment versus the long term stress in the real life situation. Also the ethics and mores of university students along with their beliefs about the polygraph may differ from that of the criminal population. Gustafson and Orne (1965) showed that student subjects who needed to be detected to demonstrate that they were not psychopathic were easier to detect in a subsequent examination.

Laboratory studies (Gustafson and Orne, 1965) yield the conclusion that requiring a subject to make an overt verbal response facilitates the detection rate over a situation requiring a subject to remain silent in response to the interrogation questions. By contrast, Horvath and Reid (1972) in a field investigation found that requesting the suspects to remain silent in response to interrogation questions enhanced the difference between ANS responses to critical and neutral questions. The response of

interest was respiration and, as noted earlier, part of the improvement was due to elimination of variation associated with an audible response. However, the GSR and blood pressure responses showed some indication of enhancement in the mute condition. Unfortunately, these measures were not systematically evaluated by the authors, making it difficult to tell if these were dramatic, but isolated results.

Further differences between field and laboratory investigations have occurred in relation to measures used. Some of the problems in this regard with the GSR, blood pressure and respiration have already been mentioned. The fact that there are discrepancies between field work and the laboratory should not deter investigation in the detection of deception. On the contrary, this should encourage more systematic experimentation to resolve and clarify these issues.

Podlesny, Raskin and Barland (1976) and Podlesny and Raskin (1977) approached the detection of deception with the express purpose of resolving differences between the field and laboratory. They advocated the use of the mock crime paradigm in an attempt to simulate actual field conditions while still retaining the advantages of the laboratory. They believe that subjects should be motivated with a meaningful reward for the judgment of "nondeceptive" in a polygraph examination and all studies should have non-deceptive controls. Finally, objective quantification of data should be employed and both statistical tests of significance and accurary of specific measures should be included in the report.

Theories of Psychphysiological Detection of Deception

Several attempts have been made to explain why in a theoretical context physiological responses are differentially enhanced to critical stimuli in

the detection of deception paradigm (Davis, 1961). The conditioned response theory suggests that critical stimuli play the role of conditioned stimuli. As conditioned stimuli they may evoke emotional responses that had been associated with these stimuli in the past. This theory appears reasonable when considering an intensely emotional field interrogation but does little to explain good results obtained in the laboratory where emotional involvement may be trivial (Davis, 1961). Conflict theory proposes that physiological disturbances occur when incompatible response tendencies are aroused at the same time. Habit may dispose a subject to answer a question truthfully and this would compete with a lie response (Davis, 1961). A third theory, based on the threat of punishment, states that the physiological responses are due to anticipation of negative consequences if the suspect is discovered in the attempt to deceive.

Day and Rourke (1974) noted that in each of the above mentioned theories there is the supposition that the suspect is aware of being in a lie detection situation. If such awareness is not necessary for the production of differential physiolgical responses, explanations based on fear of punishment or motivation to deceive the experimenter are not necessary. Day and Rourke (1974) included conditions in which subjects did not realize they were in a lie detection situation. They hypothesized that "short term familiarity" is a sufficient condition for differential physiological responses. For one short term familiarity group significant rates of detection were found giving support to the hypothesis that maintaining attention may be a sufficient condition for detection. The detection rates for this group did not differ from those of two other groups who were aware they were in a lie detection experiment and were motivated to deceive the experimenter.

The above result is compatible with data suggesting that differential physiological

responses in detection experiments are not contingent upon actual lying by the subject. Horvath and Reid (1972) could detect guilt when suspects remained mute to all questions. Gustafson and Orne (1965) could also detect critical items when subjects remained mute in response to questioning. Kugelmass, Lieblich and Bergman (1967) found differentially enhanced physiological responses to critical items when the subjects answered truthfully about these items and lied about the irrelevant stimuli. Berkhout, Walter and Adey (1970) found equivalent autonomic responses whether the subjects admitted or denied indulgence in certain sexual practices.

Day and Rourke (1974) noted the observation of Orne, Thackray and Paskewitz (1972) that the designated critical stimulus becomes a figure in a figure-ground relationship among stimuli and it is not why the subject attends to a given stimulus but the degree of attention that is paid to the stimulus that is important. Day and Rourke (1974) found the poorest detection rates for those who were bored or claimed their mind was wandering.

Lykken (1959) based the guilty knowledge test on a similar rationale; that the differential response is due merely to the fact that the guilty subject recognizes the critical item. No lie is necessary and the subject may remain silent throughout the interrogation.

The concept of the orienting response was used by Ben-Shakhar (1977) to explain how guilty knowledge tests work. Basically physiological responses evoked by novel or signal stimuli habituate over repeated presentations such that, with other conditions being equal, the size of a given physiological response is dependent upon the number of prior stimulus presentations (Sokolov, 1963). Ben-Shakhar (1977) in agreement with Orne, Thackray and Paskewitz (1972) and Lykken (1959) suggested that guilty subjects form just two categories

for the stimuli in a guilty knowledge test. One category is comprised of the one or few critical items while the other is comprised of the many, in a relative sense, neutral items. Over successive trials the presentation of neutral stimuli outnumber those of critical stimuli and thus the response to any given neutral stimulus is smaller because habituation has occurred. In this framework knowledge or recognition of critical items is the fundamental requirement and any factors that serve to differentiate and maintain that differentiation between the critical and neutral items could enhance detection. Thus Waid, Orne, Cook and Orne (1978) found well learned code items were more easily detected than poorly learned items. Lieblich, Naftali, Shmueli, and Kugelmass (1974) failed to find a motivation effect when the interrogator tried to detect a subject's own name embedded amongst other names. In this case all subjects whether motivated or not could easily categorize their own name as different from other names. Geldreich (1942) found 100% detection rates after presenting 20 to 50 buffer (neutral) items prior to any guilty information. Ben-Shakhar's (1977) formulation, which he calls the dichotomization theory, is elegantly simple, incorporates previous suggestions of "figure-ground" type relations, is tied to a well established concept, the orienting reflex, in psychology, and removes the burden of espousing a special lie response in the detection of deception. Ben-Shakhar (1977) has presented evidence for his theory with a demonstration of equally complete habituation of responding to critical stimuli by presenting as many critical as neutral stimuli. Unfortunately as he increased the number of critical stimuli he did not test for memory. If he had done so and subjects did remember correctly the up to four critical numbers it would be a convincing demonstration of a promising theory.

Ben-Shakhar's (1977) dichotomization theory may not be so readily applicable to the control question tests because of the more variegated nature of the items and the equal frequency of presentation of the past life and crime relevant items. Interestingly these crucial pairs of questions are prefaced with three or four general orientation questions. If the suspect lumped these into the same category as the past life questions then crime irrelevant questions would be presented at a rate of six or seven to every three crime relevant questions. Such a tendency would aid in detecting guilty subjects but would also result in the false classification of innocent subjects as guilty. Horvath (1977) pointed to just such a tendency. The optimal design of the control test should, if Ben-Shakhar's (1977) theory is accepted, promote a trichotomization of questions among the innocent (orientation, crime irrelevant and crime relevant) and a dichotomization among the guilty (crime irrelevant and crime relevant).

Purpose of the present research

The present research was conceived primarily to investigate two facets of the detection of deception. First, manipulations designed to show the "lie detector" operating at different levels of accuracy have been demonstrated to affect detection rates in the card test paradigm (Bradley and Janisse, 1979). The desire was to extend that finding to a mock crime paradigm while also revealing any potential interactions with threat of punishment. Second, Bradley and Janisse (Note 1) have explored the behavior of the pupillary response in the card test paradigm and wished to extend this to a mock crime situation while including comparison measures of GSR and heart rate.

Proposed research

The present study followed the recommendations of Podlesny and Raskin (1978) in examining the hypotheses in a mock crime situation which included both positive and negative (threat of shock) consequences contingent upon the outcome of the interrogation.

Only a few studies incorporating a mock crime with consequences have been carried out. Generally success rates in discriminating the guilty from the innocent have been so high that the analyses have been insensitive to possible effects due to factors such as motivation. Davidson (1968) had subjects attempt a mock murder. They could keep any money found on the "victim" if they successfully avoided PDD in a later interrogation. The money ranged from less than $1.00 to something between $25.00 to $50.00. Eleven of the twelve guilty subjects were detected with no differences between the two levels of money value. Thirty-six innocent subjects were also included in this experiment which used a guilty knowledge interrogation test and all of these were correctly judged as innocent. Podlesny, Raskin, and Barland (1976) instructed half of their subjects to steal from a secretary's desk and motivated both the guilty and innocent subjects with the promise of $10.00 for a PDD diagnosis of innocent. Control question tests yielded detection rates of 80% correct, 10% error and 10% inconclusive while the guilty knowledge test was 90% correct with no inconclusives.

For some unknown reason punishment or threat of punishment has received little attention in the mock crime paradigm. Barland and Raskin (1975) threatened innocent subjects with the loss of course credit for appearing guilty during the test. Guilty subjects were not threatened for appearing

guilty but could keep $10.00 taken in the crime for an appearance of innocent during PDD. Lykken (1959) in a study involving shock as punishment achieved an accuracy rate of 94% with suspects of a mock crime murder and/or theft. The shock immediately followed certain questions regardless of the physiological response or actual guilt or innocence of the suspect. In neither of these studies was punishment or threat manipulated. The authors merely assumed it was important. Interestingly enough, Lykken (1959) may have inadvertently created the conditions under investigation in the present experiment. Shocks noncontingent on the actual response or guilt of the subject could foster the belief among innocent subjects that the "lie detector" was very inaccurate while leading the guilty subjects to believe that it was very accurate. In any event the study of punishment or threat of punishment effects has been neglected and is in need of systematic investigation.

It would seem that more realism could be achieved in mock crime studies if all subjects were under the impression that an appearance of deception or guilt would lead to punishment since in practical field work even innocent suspects may entertain the possibility of false imprisonment. In addition innocent suspects in criminal investigations rarely receive positive rewards for a correct PDD diagnosis but rather they avoid punishment. Guilty suspects may gain in two ways following an incorrect diagnosis of innocence, first they avoid punishment and second they reap the rewards of the criminal endeavour. The present study included a mock crime with half of the guilty and innocent suspects threatened with shock for the appearance of deception. Guilty subjects gained a monetary reward for successful deception.

Field investigators have often assumed that critical responses would be

enhanced after the effectiveness of the autonomic lie detection technique had been demonstrated to the suspect (Barland & Raskin, 1975). The demonstration usually consisted of a card test arranged, unknown to the subject, as to the card to be chosen, recorded the autonomic responses and infallibly detected the correct card (Reid & Inbau, 1966). The test hopefully convinced the subject that this technique was effective in discovering all of the subject's lies.

Laboratory studies do not support this assumption. Instead, feedback of successful detection has either been ineffective or has actually resulted in a reduction of detectible responses to critical questions. In a mock crime situation, Barland and Raskin (1973) found no difference between three groups of subjects, one given feedback that the polygraph was perfectly effective, one that it was ineffective and a final group given no feedback. Ellson, Davis, Saltzman and Burke (cited in Gustafson and Orne, 1965) found subjects given feedback that the polygraph detected their attempts at lying were harder to detect on a subsequent trial.

Davis (1961) put forth an explanation to account for the decrease in detection rates found by Ellson, Davis, Saltzman and Burke (cited in Gustafson and Orne, 1965) by suggesting that physiological responses during lying were associated with a state of uncertainty. Davis (1961) viewed lying as an avoidance reaction with some, but not complete, confidence of success. Reasoning that the greater the subject's confidence in the validity of autonomic lie detection techniques the less his uncertainty in regard to detection, Davis (1961) hypothesized that a guilty suspect with the subjective impression that there was no prospect of successfully avoiding detection would show a minimum response while lying. "Apparently, when S is convinced that the instrument is infallible he is resigned and ceases to be

excited about the critical questions". (Davis, 1961, p. 154).

Lykken (1974), on the other hand, suggested that increasing a guilty suspect's confidence in autonomic lie detection would increase his autonomic responsiveness to the critical question. This autonomic responsiveness would be at a maximum if the guilty suspect had perfect confidence in the technique.

The hypotheses generated by Davis (1961) and Lykken (1974) predict opposite results when guilty subjects who are attempting to deceive have perfect confidence in the autonomic lie detection technique. That is, Lykken (1974) would predict that subjects detected on 100% of previous trials would be detected more often on subsequent trials. Davis (1961) on the other hand, would predict that such subjects would be detected less often.

Bradley and Janisse (Note 3) discovered that studies investigating this question used only a one trial demonstration and thus subsequent feedback could imply only that the apparatus was either perfectly effective or perfectly ineffective. By giving feedback after each of three trials Bradley and Janisse (Note 3) created four groups which had been detected on none, one, two, or three trials. The results supported the uncertainty hypothesis advanced by Davis (1961): that is the groups detected on only one (33.3%) or two (66.7%) of the three trials were easier to detect than those detected on none (0%) or three (100%) of three trials. It would however, be premature to discount the hypothesis proposed by Lykken (1974) since the paradigm was merely the card test and no form of incentives were used. Thus it was a purpose of the present study to manipulate the demonstrated effectiveness of the detection process and test the resultant outcome in the mock crime situation involving incentives.

Podlesny, Raskin and Barland (1976) have investigated various questioning techniques used in PDD and found that the Backster Control Question technique and Lykken's Guilty Knowledge technique are the two most effective. These techniques were used in the present experiment. Prior to receiving feedback, but after questioning, subjects were required to rank order the questions assigning a rank of one to the question thought to be the most disturbing in the sense that it would have elicited the most detectable physiological response. Later these responses were correlated with the ranks of the actual physiological responses. Subjects were also asked which of the techniques seemed most effective in detecting them. Previous investigations have failed to collect such self report data but it seems a necessary step toward understanding the cognitive or perceived impact of questions on subjects and the ultimate relations to physiological responses.

Three physiological measures were employed in this study, the galvanic skin response (GSR), heart rate and the pupil response. The GSR has been highly effective in virtually every laboratory study of deception (Barland and Raskin, 1973). Heart rate measures have not been so effective but yielded detection rates better than chance (Orne, Thackray and Paskewitz, 1972) and may be sensitive (responsive) to different factors in detection (Clark, Note 4). Although the pupillary response is purportedly a sensitive measure (Hakerem, 1970; Janisse, 1977), it has not been extensively investigated in the deception paradigm. Berrien and Huntington (1943) successfully discriminated guilty from innocent using the pupillary response but the accuracy of their instrumentation has been questioned (Young, 1959). Bradley and Janisse, in a series of experiments, used the pupil response in the card test paradigm, to investigate verbal response parameters and habituation (Note 1),

individual differences (Note 2) and varied levels of subjective impressions of the probability of detection (Note 3). In general they found: a) no differences between verbalizing or not verbalizing during PDD, b) a reduction in successful detection over repeated trials, c) differences in detectability between high and low Machiavellians and d) higher detection rates with subjects uncertain as to whether they would be detected or not. Clark (Note 4) used a secret agent paradigm to investigate and compare the pupillary response with heart rate and GSR. All three measures yielded significant detection rates but the correlations among them were low with only that of the pupillary response and heart rate reaching significance. The GSR yielded the most accurate discrimination with the pupil response second and heart rate third. No differences between the two levels of motivation based on rewards of one dollar or five dollars for successful deception were found. In the present study the three measures were used and compared in the more involving mock crime situation.

In summary three physiological measures were employed in PDD using the involving mock crime paradigm with negative incentives (avoidance of shock) motivating both guilty and innocent subjects to avoid a diagnosis of deception with the additional possibility of reward (keeping the money) for successful deception by the guilty subjects. Subjects' beliefs in the technique's ability to yield a correct diagnosis were manipulated.

It was hypothesized that guilty subjects would be detected more often than innocent subjects. This would be evidenced most in the no threat of punishment condition when guilty subjects are uncertain of the lie detectors accuracy, as previously found by Bradley and Janisse (Note 4), but not enough evidence was available to judge whether this relationship would hold in the punishment condition. It may be supplanted by the relationship

predicted by Lykken (1974) that subjects certain of detection will manifest the largest response. The detection rates of innocent subjects (false positives) were predicted to be influenced by the punishment and confidence manipulation only when responding to the control question technique and not to the guilty knowledge technique because, no matter how effective the manipulation, innocent subjects have no knowledge of the appropriate guilty items in the later technique.

The GSR was expected to be the most effective measure for detection, followed by the pupil response and heart rate (Clark, Note 4). Reviewing the literature, Janisse (1977) reported that other than finding a general lack of relationship among measures, heart rate and the pupil response were negatively correlated. This directional fractionation between heart rate and pupil response was expected to be demonstrated in the present experiment.

Chapter II

Method

Subjects

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

Apparatus

The pupillary response was measured by a Whittaker Space Sciences Eye View Monitor and Television Pupillometer designed to provide an accurate assessment of pupil diameter and to record the data in digital form on a Kennedy incremental tape recorder, Model 1600/360. Heart rate was assessed by a Whittaker Pulse Watch which was also designed to output the data on the same Kennedy recorder. The galvanic skin response was recorded on a two pen chart drive with one pen serving as an event marker. The medial phalanges on the first and third fingers served as recording sites and were fitted with Lafayette electrodes (cat.# 76602) after the fingers had been cleaned with a cotton swab dipped in alcohol. The electrodes, attached without electrolyte, were curved to conform to finger shape and were held in place by velcro wraps. The galvanic skin response (GSR) recordings of resistance were made in the A.C. automatic centering mode.

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

Procedure

Subjects who had left their name and telephone number some weeks earlier during a session of filling out personality inventories were contacted by phone and told to report to a specific room which was located near the detection laboratory. All subjects upon entering the room found a note taped to the table directing them to close the door and set a bell timer for five minutes to allow them the privacy and time to read and carry out the instructions which had been placed on the table before their arrival. Innocent subjects (See Appendix B) read that they were to sit in the room for the time interval and then open the door to wait for the arrival of the interrogator. Guilty subjects (See Appendix C) read that the room they were in was part of a store and that hidden at the back of a shelf under the table was one dollar in a medium sized white envelope. Each were to steal the dollar and hide it in their right front pant's pocket. After committing the crime they concealed the instructions, waited for the time interval to end, opened the door and sat waiting for the interrogator.

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

Half of the guilty and half of the innocent subjects were

informed that they would receive a painful but not permanently damaging electric shock if judged guilty.

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

The interrogator, cued by the bell of the minute timer, arrived at the open door,asked for the subject's name,accused him of the crime and lead him to the interrogation room. Once there the subject was shown the interrogation equipment and was given a little information about the lie detection process. Subjects in the threat of shock condition were shown the shock apparatus and electrodes and were reminded that they could receive shock whether guilty or innocent if they were judged guilty. It was also explained that those judged guilty if actually guilty would forfeit the stolen money. Of course, with a judgement of innocent a subject would be free to leave. At this point subjects were reminded that they could withdraw from the experiment and still retain their experimental credits. In either instance,continuing or not,subjects promised not to divulge details of the

experiment. Those who continued then filled out Spielberger's State Anxiety Scale.

In preparation for the interrogation the questions in the Backster Control test were reviewed. The crucial part of the test contains crime relevant questions paired with control questions (see Appendix D) and prior to the review subjects were reminded that they must on the one hand answer "No" to all crime relevant questions because the innocent were truly innocent while the guilty were attempting to conceal their guilt. On the other hand all subjects were to feel free in discussing questions about their past life so that ambiguities could be clarified making it possible to answer these questions with a "No". This was done to make the verbal responding comparable between each of the crime relevant and control question pairs and was achieved by making a standard minor modification, "Except for what you have told me....", in the control question. Questions in the Guilty Knowledge test were not reviewed and subjects were only informed that they would be asked a series of questions about items containing certain information only the guilty subjects knew (see Appendix E).

After being prepared for the interrogation, which included strapping on the shock electrodes in the threat of shock condition, subjects took a series of card tests ostensibly designed to assess their detectability. They were given three trials and after each were informed of the number the equipment indicated they had attempted to conceal. Depending on the group subjects were detected on none, one, two or three of three trials to create the impression that the detection measures were ineffective, sometimes effective or perfectly effective. In order to carry out this demonstration the

experimenter rigged the card tests so that unbeknownst to the subjects they chose a predetermined number. In this manner the experimenter would feed back the correct number for a hit or some other number for a miss regardless of the actual performance of the equipment. Subjects kept a concealed score of the results of this test by writing down the number they chose and the number indicated. They did this because the interrogator had informed them that he was very interested in studying this information in the future. In actuality the record served as a manipulation check and prevented the subject from being inattentive or forgetful of the results.

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

The questions in both techniques had been prerecorded for presentation

so that the interrogators vocal inflections would not vary across subjects. The recorded questions were spaced at twenty-second intervals and it took 8 minutes to deliver the 24 questions in the Guilty Knowledge test. The 9 questions, again, were recorded at twenty-second intervals, in the Backster test and took 3 minutes to present. There was a pause while the tape was rewound before the second and third presentations. The Backster test was individualized by dubbing in the subject's name and the phrase "Except for what you have told me" as necessary for the control questions. The questions for the series of three card tests used for the effectiveness manipulation had been pre-recorded at 10-second intervals and it took one minute to deliver each test.

Data Analysis

The data were analyzed on different levels. Detection scores most relevant to field work were derived by procedures described by Barland and Raskin (1975) for the Backster test and by Lykken (1959) for the Guilty Knowledge test. With the Backster test each control and relevant question pair was assigned a score of 1, 0, or -1, depending on whether the response to the control question was larger, the same, or smaller. The test had three such question pairs and was repeated three times for a scoring range between +9 and -9. Subjects with scores of +2 or greater were classed as innocent while those with -2 or less were classed as guilty. Scores between these values were not judged but categorized as inconclusive. In the Guilty Knowledge test each of the four question sequences received a score of 2, 1, or 0 depending on the relative magnitude of the response to the critical item. If the response to the critical item was largest it received a score of 2, if second largest, 1 and finally 0 for any other response magnitude. In this test

with four five-item sequences, plus a buffer item per sequence, the scores could range from 0 to +8. A score of 4 or more was classed as deceptive and less than 4 as innocent. The frequency of detection was reported from these scores.

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

All of the above analyses used the data from five dependent measures: pupil size, heart rate, pupil change, heart change and GSR. Pupil size and heart rate scores were, respectively, the highest and lowest values obtained in the response period while pupil change and heart change scores were these values subtracted from the average of the three seconds of base-line immediately prior to the appropriate question. This full metric (continuous) data was analyzed for additional information in the basic 4 x 2 x 2 between factors design augmented by three within factors, question type (control or crime relevant), question pair and trial number for the Backster test and two within factors, question type and question set for the Guilty Knowedge test. The automatic self centering baseline feature precluded the extraction of data at this level for the GSR.

Leaf blank to correct

numbering

CHAPTER III

RESULTS

Accuracy of Detection

The classification of subjects into categories of deceptive and non-deceptive in relation to whether they were actually guilty or innocent is displayed in Table 1 for each of the tests (Backster and Guilty Knowledge) and the five physiological measures.

In considering the Backster test it should be noted that an additional category of inconclusive was used. Examination of the χ^2 analyses in the table revealed that scores were non-randomly distributed across the six categories for all measures. When inconclusives were excluded from the analyses it was found that correct decisions were made more frequently than incorrect decisions with the GSR $\chi^2_1 = 59.78$, $p<.05$, heart rate $\chi^2_1 = 19.44$, $p < .05$, and heart change $\chi^2_1 = 5.96$, $p < .05$. Further analysis on absolute pupil size showed that subjects whether guilty or innocent tended to be classed as non-deceptive $\chi^2_1 = 34.63$, $p < 01$.

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

Three discriminant analyses were conducted on the rank scores to determine the best linear combination of the 5 dependent measures for differentiating between guilty and innocent subjects. The first two analyses involved the measures in each separate test while the third

Table I

Frequency of Detection

	Backster Test				Guilty Knowledge Test			
	Classification				Classification			
	Deceptive	Non-Deceptive	Inconclusive		Deceptive	Non-Deceptive		
GSR								
Guilty	58	13	25		57	39		
Innocent	9	56	31		11	85		
	χ^2_2 =	64.62	$p < .05$		χ^2_1 =	46.11	$p < .05$	
Pupil Size								
Guilty	17	41	38		51	45		
Innocent	11	52	33		26	70		
	χ^2_2 =	2.94	N.S.		χ^2_1 =	12.49	$p < .05$	
Heart Rate								
Guilty	50	12	34		38	58		
Innocent	25	37	34		17	79		
	χ^2_2 =	21.08	$p < .05$		χ^2_1 =	11.24	$p < .05$	
Pupil Change								
Guilty	33	21	42		32	64		
Innocent	30	25	41		19	77		
	χ^2_2 =	0.50	N.S.		χ^2_1 =	3.84	p	.05
Heart Change								
Guilty	34	20	42		43	53		
Innocent	19	32	45		17	79		
	χ^2_2 =	7.06	$p < .05$		χ^2_1 =	15.15	$p < .05$	

analysis combined measures and tests. Examination of the standardized discriminant function coefficients in Table 2 makes it clear that the GSR contributed most to differentiation in all instances. Heart rate contributed somewhat in the Backster test and pupil size in the guilty knowledge test.

The contribution of the GSR scores was proportionately so large that discriminatory power gained by a linear combination of these scores with those from the additional measures was insignificant. For example, with the Backster test, using no inconclusive category for comparison purposes, the GSR scores alone correctly classified 138 subjects while the combination managed only 148 $\chi^2 = 0.86$, $p < .50$. The same result occurred in the Guilty Knowledge test where the combination classified only one more subject correctly (142 versus 143). When tests were combined there was an increase in correct classification $\chi^2_1 = 8.46$, $p < .01$. In the main, this was due to the two sets of GSR scores working in conjunction.

Hypotheses Tests on Detection Scores

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

Backster test. No significant main effects were found for the levels of manipulated effectiveness over any of the five dependent measures. The shock factor was only marginally significant F (1, 176) = 5.079, $p < .025$ on change in heart rate such that subjects not threatened with shock tended to receive ranks more in the innocent direction ($\bar{X} = .67$) than

Table 2

Discriminant Analyses

	Backster	Guilty Knowledge	Combined
GSR	0.868		0.639
Pupil Size	0.058		0.099
Heart Rate	0.373		0.288
Pupil Change	-0.081		-0.143
Heart Change	0.166		0.040
GSR		-0.795	-0.596
Pupil Size		-0.282	-0.133
Heart Rate		-0.254	-0.211
Pupil Change		-0.110	-0.080
Heart Change		-0.214	-0.106

those threatened with shock ($\bar{X} = -0.41$). All of the dependent measures except for pupil change discriminated between the guilty and innocent: galvanic skin response F (1, 176) = 83.95, $p < .001$ ($\bar{X}_I = 1.92$, $\bar{X}_G = -1.98$), pupil size F (1, 176) = 3.91, $p < .05$ ($\bar{X}_I = 2.30$, $\bar{X}_G = 1.34$), heart rate F (1, 176) = 25.025, $p < .001$ ($\bar{X}_I = 0.68$, $\bar{X}_G = 1.46$), heart change, F (1, 176) = 8.725, p .004 ($\bar{X}_I = .83$, $\bar{X}_G = -.57$).

The levels of effectiveness factor was found to interact with guilt with the galvanic skin response scores F (3, 176) = 4.041, $p < .008$. The interaction is displayed in Figure 1. No other interactions with the galvanic skin response or any other dependent measure were found to be significant.

Guilty Knowledge Test. A three factor analysis of variance was conducted for each of the five dependent measures of galvanic skin response, pupil size, heart rate, change in pupil size and change in heart rate. Manipulated effectiveness was significant with the galvanic skin response F (3, 176) = 4.175, p <.01 and pupil size F (3, 176) = 3.364, $p < .05$. A trend analysis on the galvanic skin response data revealed significant quadratic F (1, 176) = 5.677, $p < .02$. and cubic trends (F (1, 176) = 4.135, p <.01, but no linear component. Only the quadratic component was significant with pupil size F (1, 176) = 4.36, p < .05 (see Figure 2).

None of the dependent measures significantly discriminated between the threat of shock conditions. All of the dependent measures discriminated significantly between guilt and innocence: GSR F (1, 176) = 72.87, $p < 0.001$ ($\bar{X}_I = 1.84$, $\bar{X}_G = 3.90$), pupil size F (1, 176) = 17.43, $p < 0.001$ ($\bar{X}_I = 2.55$, $\bar{X}_G = 3.52$), heart rate, F (1, 176)

Figure 1

Interaction of Manipulated Effectiveness by Guilt with the Backster Test Utilizing GSR

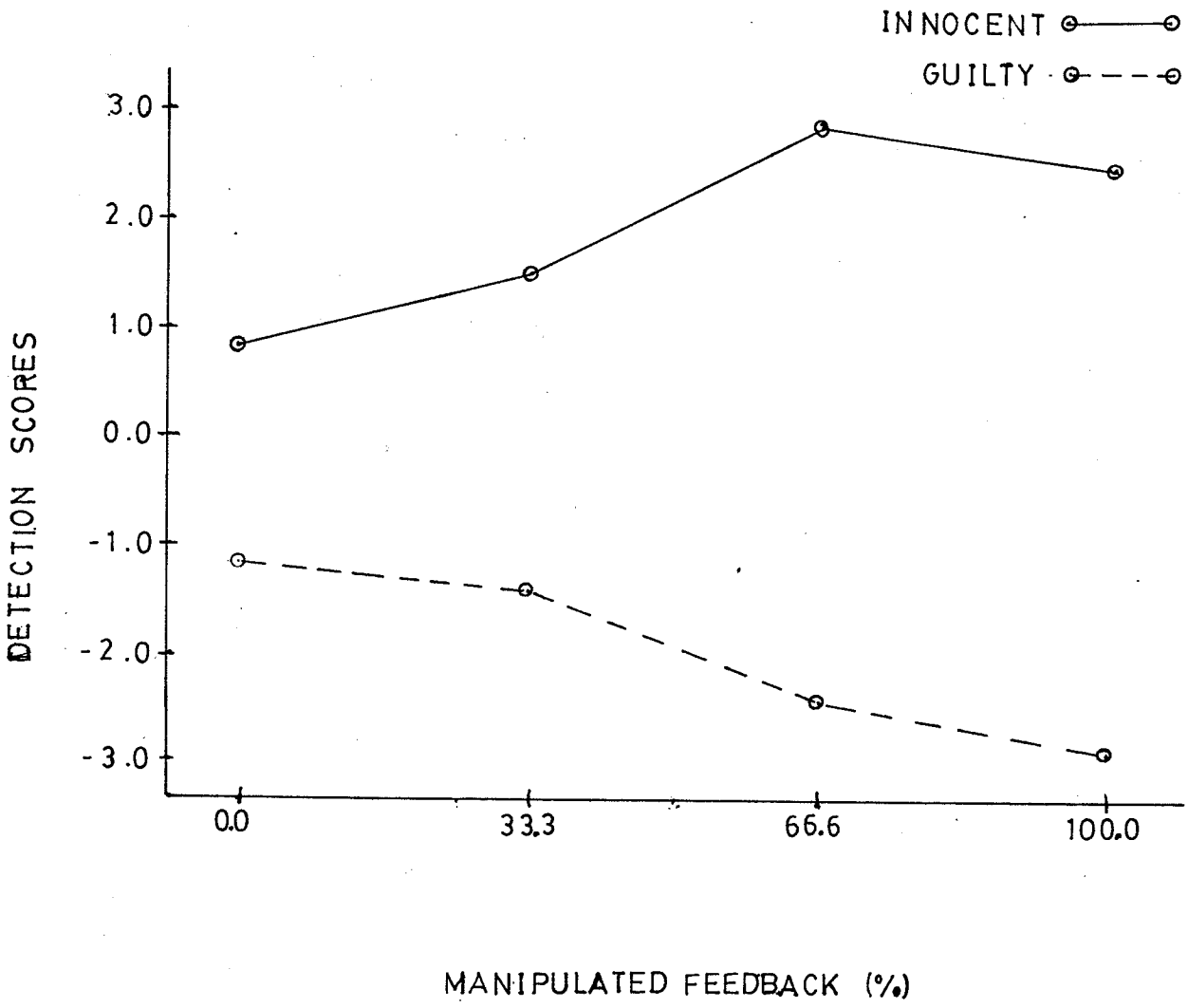
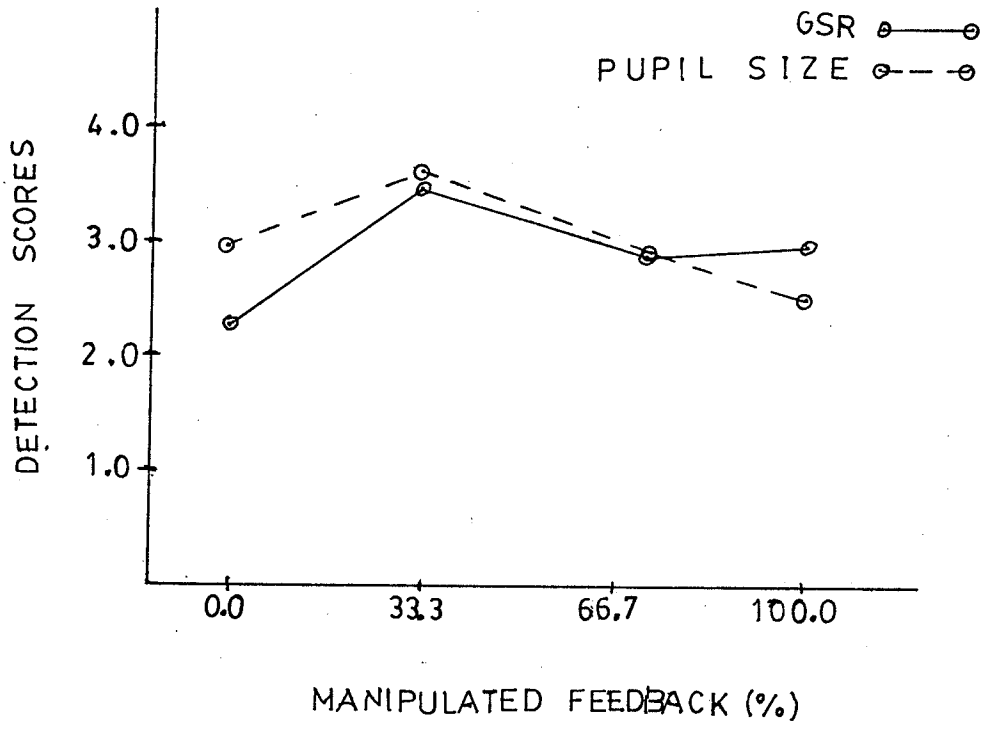

INNOCENT
GUILTY
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
DETECTION SCORES
0.0
33.3
66.6
100.0
MANIPULATED FEEDBACK (%)

Figure 2

Manipulated Effectiveness with the Guilty Knowledge Test Utilizing GSR and Pupillary Response


GSR
PUPIL SIZE
DETECTION SCORES
4.0
3.0
2.0
1.0
0.0
33.3
66.7
100.0
MANIPULATED FEEDBACK (%)

= 13.435, $p < 0.001$ ($\bar{X}_I = 2.12$, $\bar{X}_G = 3.06$), pupil change F (1, 176) = 5.083, $p < 0.025$ ($\bar{X}_I = 2.34$, $\bar{X}_G = 2.88$), heart change F (1, 176) = 13.979, $p < 0.000$ ($\bar{X}_I = 2.29$, $\bar{X}_G = 3.23$). None of the interactions, either two-way or three-way, for any of the five dependent measures were significant.

Actual Response Magnitudes

The measurement instruments for heart rate and pupil size allowed the recording of actual baseline and response magnitudes rather than just the relative (rank order) responses as with galvanic skin response. The analysis of the actual magnitude data, though not conducted in field research, contributes by allowing assessment of autonomic arousal in each of the conditions. Detection data reveals only the presence of a within subjects interaction between critical and non-critical questions whereas magnitude data reveals the overall level of arousal at which detection results may be occurring. The 8 major ANOVAS that follow are portrayed in Appendix F. Geiser Greenhouse (1958) conservative degrees of freedom were used in all within subjects analyses.

Baselines

The baselines recorded prior to the demonstrations of effectiveness but after subjects had carried out their instructions and been threatened with shock showed no differences on main effects or interactions on either pupil size or heart rate (see Table 3).

Backster Test

Four measures, pupil size, heart rate, pupil change and heart change were each examined in a six factor analysis of variance in which there were three between factors manipulated effectiveness, shock and guilt and three within factors trials, questions and question type

Table 3

Analysis of Variance of Baseline Data

Source of Variation	Measure	df	MS	F	p <
A (Effectiveness)	Pupil Size	3	288.73	1.33	.27
	Heart Rate	3	227.94	0.95	.42
B (Shock)	Pupil Size	1	219.13	1.01	.32
	Heart Rate	1	542.74	2.27	.13
C (Guilt)	Pupil Size	1	312.31	1.44	.23
	Heart Rate	1	8.16	0.34	.85
A B	Pupil Size	3	116.91	0.54	.66
	Heart Rate	3	97.73	0.41	.75
A C	Pupil Size	3	250.87	1.16	.33
	Heart Rate	3	13.67	0.66	.98
B C	Pupil Size	1	411.91	1.90	.17
	Heart Rate	1	57.99	0.24	.62
A B C	Pupil Size	3	5.07	0.02	.99
	Heart Rate	3	48.03	0.20	.90

(innocent - guilty).

Pupil size. No main effects were found for manipulated effectiveness, shock, or guilt. Pupil size was larger on the first trial than on the second and third trials F (1, 176) = 95.47, $p < .001$ ($\bar{X}_1$ = 101.504, $\bar{X}_2$ = 95.209, $\bar{X}_3$ = 95.672). Over each successive set of questions within a trial pupil size decreased F (1, 176) = 224.26, $p < .001$ ($\bar{X}_1$ = 101.203, $\bar{X}_2$ = 96.337, $\bar{X}_3$ = 94.844). An interaction between trials and question sets showed that pupil size decreased most rapidly over the first trial F (1, 176) = 33.85, $p < .001$ (see Figure 3).

A difference was found between guilty and innocent questions F (1, 176) = 57.97, $p < .001$ with pupil size in response to innocent questions larger ($\bar{X}_I$ = 98.173) than to guilty questions ($\bar{X}_G$ = 96.750). The questions factor interacted with the subjects actual guilt or innocence such that innocent subjects tended to have relatively smaller pupils in response to guilty questions than guilty subjects (see Figure 4)

An interaction between question type and trials revealed that the difference between the pupil size response to control and guilty questions was largest on the initial trial F (1, 176) = 12.63, $p < .001$ (see Figure 5). Question type also interacted with question set F (1, 176) = 32.19, $p < .001$ such that the difference between types became negligible by the final set (see Figure 6).

Pupil change. As perceived effectiveness increased the pupil change in response to all questions was larger F (3, 176) = 4.0606, $p < .08$ ($\bar{X}_1$ = 9.842, $\bar{X}_2$ = 10.221, $\bar{X}_3$ = 9.926, $\bar{X}_4$ = 11.584). No other between main effects were significant. Pupil change was larger on the first trial than on the following trials F (1, 176) = 9.6324, $p < .01$

Figure 3

Significant Trial by Question Sets Interaction

on Backster Test Pupil Size

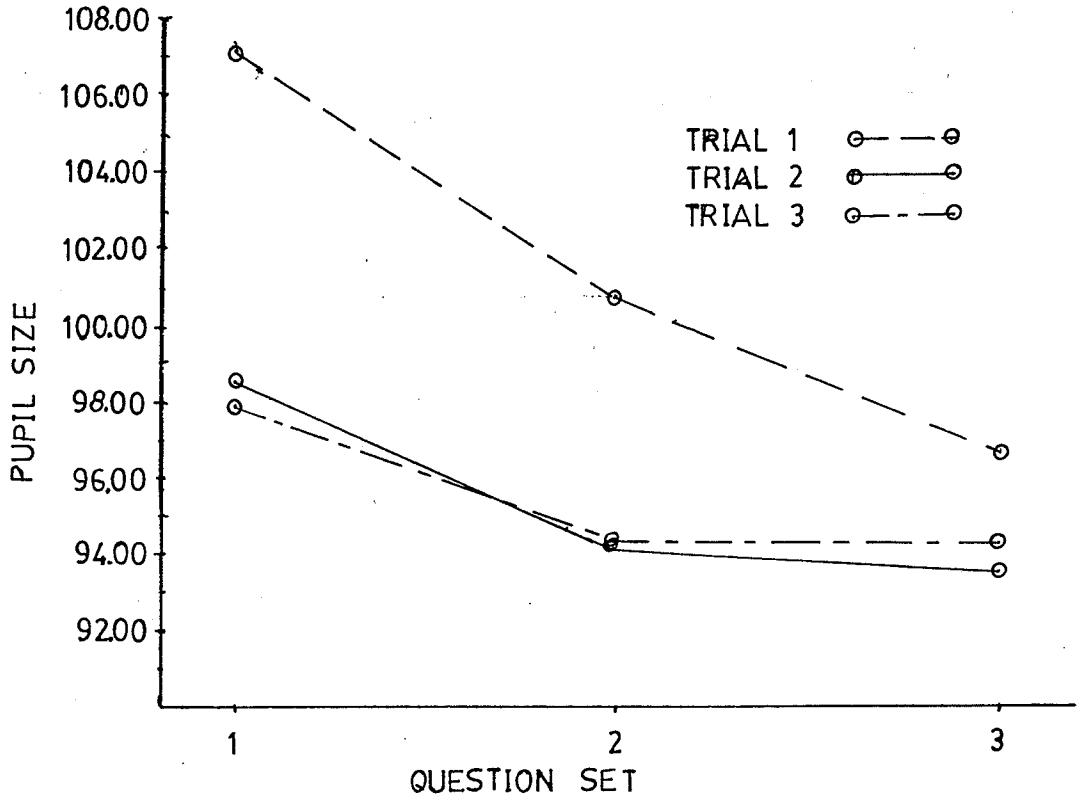

108.00
106.00
104.00
102.00
100.00
98.00
96.00
94.00
92.00
PUPIL SIZE
TRIAL 1
TRIAL 2
TRIAL 3
1
2
3
QUESTION SET

Figure 4

Significant Question Type by Guilt Interaction

on Backster Test Pupil Size

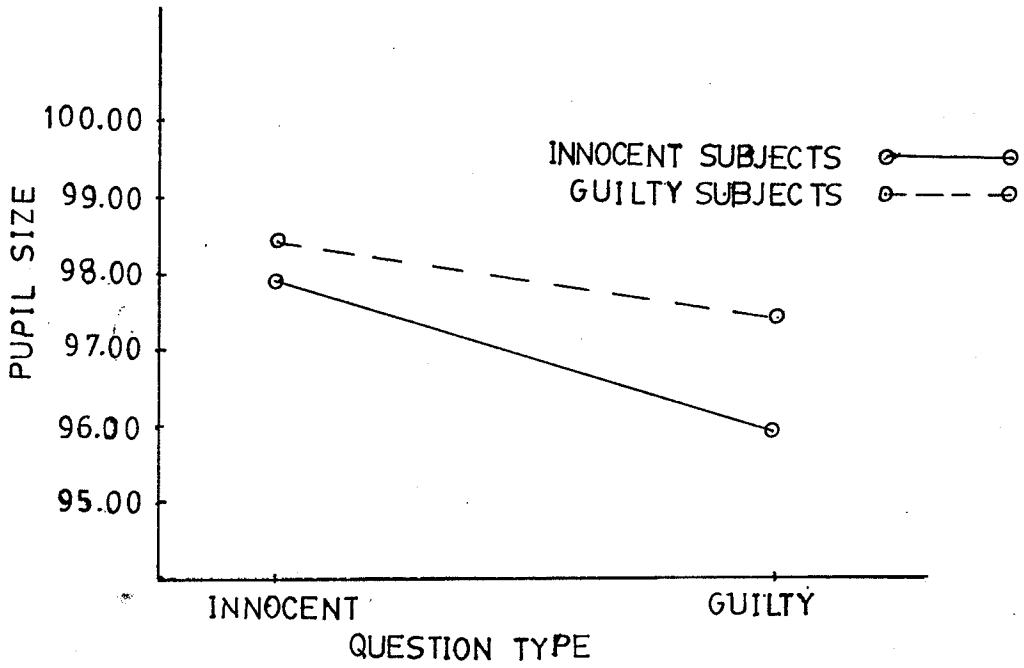

100.00
99.00
98.00
97.00
96.00
95.00
PUPIL SIZE
INNOCENT SUBJECTS
GUILTY SUBJECTS
INNOCENT
GUILTY
QUESTION TYPE

Figure 5

Significant Question Type by Trials Interaction

on Backster Test Pupil Size

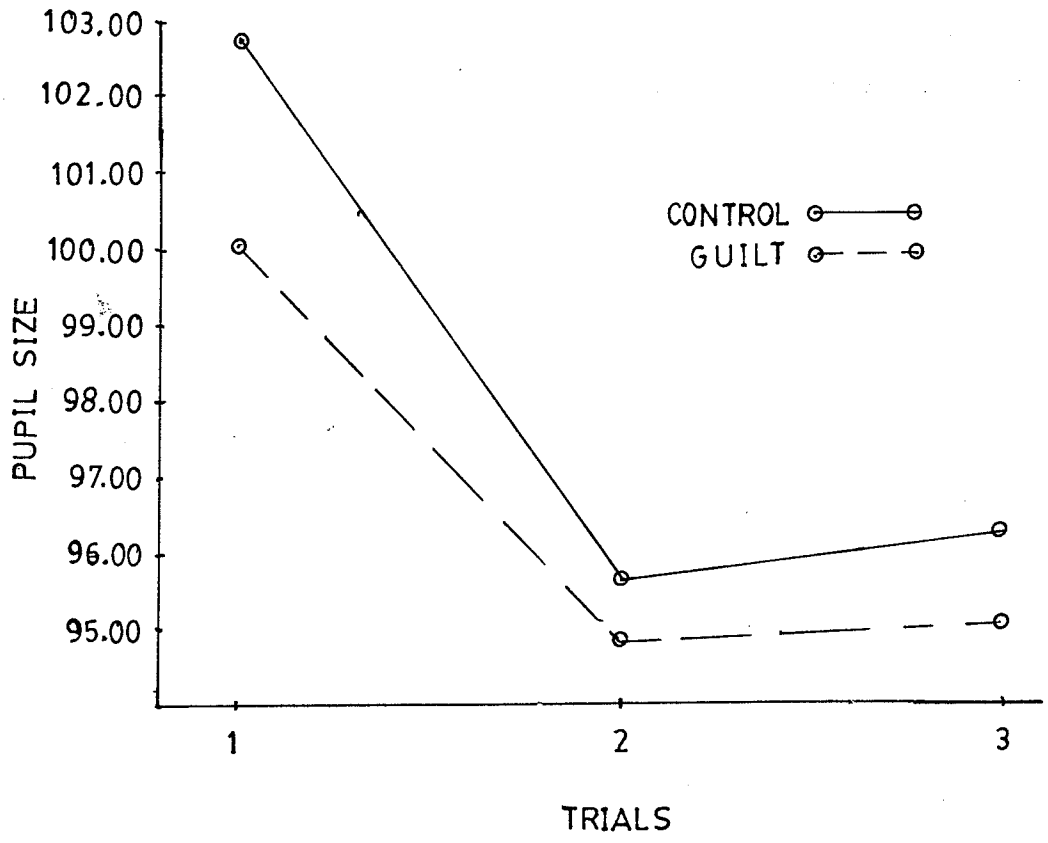
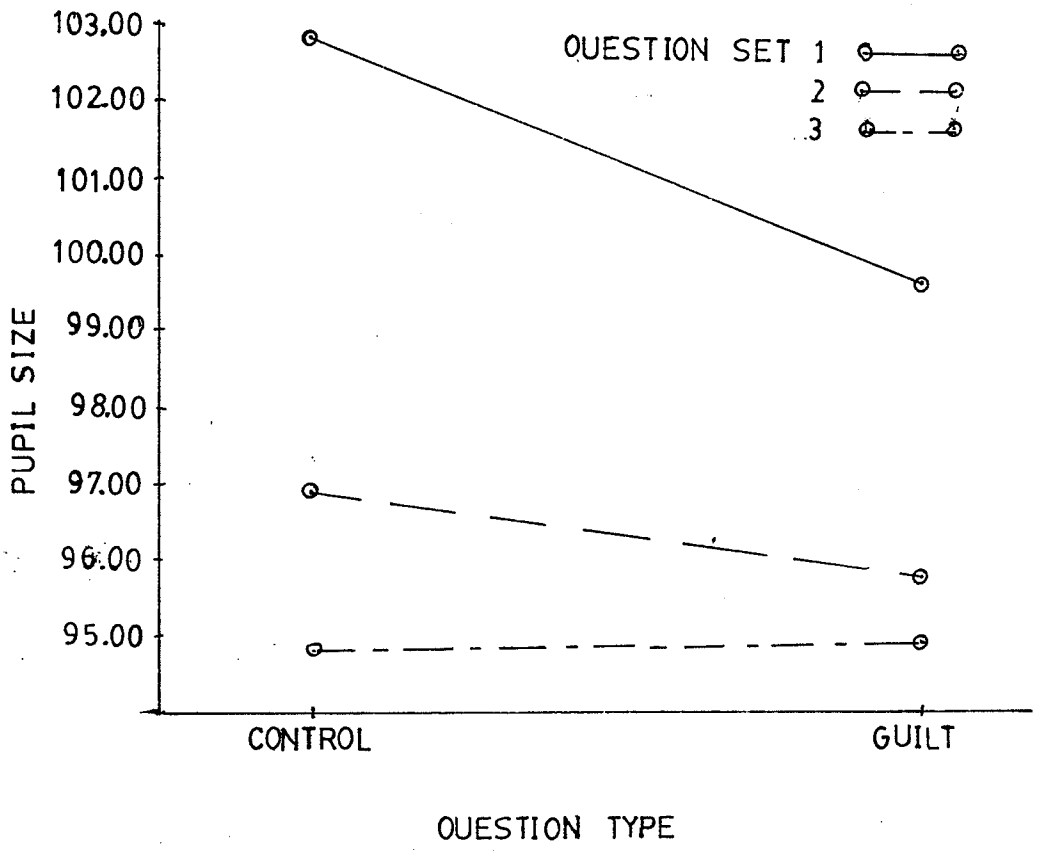

103.00
102.00
101.00
100.00
99.00
98.00
97.00
96.00
95.00
PUPIL SIZE
CONTROL
GUILT
1
2
3
TRIALS

Figure 6

Significant Question Type by Question Set

Interaction on Backster Test Pupil Size


103.00
102.00
101.00
100.00
99.00
98.00
97.00
96.00
95.00
PUPIL SIZE
QUESTION SET 1
2
3
CONTROL
GUILT
QUESTION TYPE

($\bar{X}_I$ = 10.981, $\bar{X}_2$ = 9.889, $\bar{X}_3$ = 10.282). Pupil change was smallest in response to the middle set of questions in each trial F (1, 176) = 6.2987, $p < .002$ ($\bar{X}_1$ = 10.578, $\bar{X}_2$ = 9.887, $\bar{X}_3$ = 10.688). An interaction between trials and questions revealed that the response to the initial set of questions in the first trial was larger than on all other responses F (1, 176) = 7.4839, $p < .001$. Pupil change was largest to guilt implicating questions F (1, 176) = 4.1836, $p < .042$ ($\bar{X}_I$ = 10.152, $\bar{X}_G$ = 10.616), but this did not interact with the actual guilt or innocence of the subject F (1, 176) = .72, $p < .397$ and thus was not a successful measure in discriminating between subjects. Finally, a marginally significant interaction between trials, question sets, question types, and threat of shock F (1, 176) = 3.99, $p < .05$ revealed that pupil change remains large and consistent across guilty questions in trials for subjects under threat of shock while change is more variable for subjects under no threat of shock.

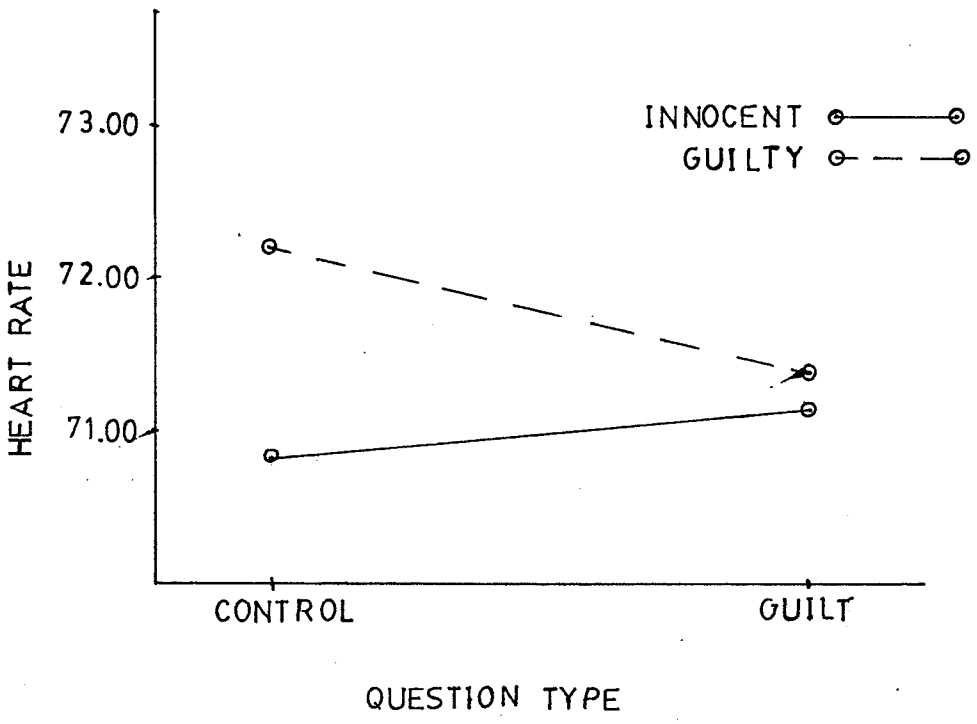
Heart rate. No between subjects main effects were significant. Heart rate was lower in response to guilty questions F (1, 176) = 4.429, $p < .0367$ ($\bar{X}_I$ = 71.524, $\bar{X}_G$ = 71.270) and this factor interacted with whether the subject was actually guilty F (1, 176) = 23.049, $p < .001$ such that the guilty subjects' heart rates were lower in response to guilty questions than in their responses to innocent questions, while the heart rates to guilty questions of innocent subjects were relatively elevated over their response to innocent questions (see Figure 7).

Finally, a complex interaction between threat of shock, guilt, trials and question type showed that heart rate in the final trial of

Figure 7

Significant Question Type by Guilt Interaction

on Backster Test Heart Rate


73.00
72.00
71.00
HEART RATE
INNOCENT
GUILTY
CONTROL
GUILT
QUESTION TYPE

innocent subjects in the no shock condition tended to be greater than in any other condition F (1, 176) = 6.977, $p < .01$. Little interpretative significance can be derived from this result.

Heart change. No main effects between or within were significant. Perceived effectiveness interacted with threat of shock such that in each threat of shock condition the largest heart change occurred at different levels of uncertainty in the perceived effectiveness condition F (3, 176) = 2.83, $p < .03$. Subjects under threat of shock showed the largest change when the apparatus appeared relatively ineffective while subjects not under threat of shock reacted most when the apparatus was relatively effective (see Figure 8).

Guilt of subjects interacted with question type such that the largest heart change response of guilty subjects was to guilty questions while the largest response of innocent subjects was to innocent questions (see Figure 9).

Finally, a trials, question set, shock interaction revealed that the initial responses in the second trial under no threat of shock were large F (1, 176) = 4.265, $p < .01$. This is a relatively trivial result having little explanatory signific nce.

Guilty Knowledge Test

The dependent measures of pupil size, pupil change, heart rate and heart change were examined in a five factor analysis of variance with three between and two within factors. The four levels of manipulated effectiveness, threat or no threat of shock, and guilt or innocence made up the three between factors while question sets and question type made up the two within factors.

Pupil size. No significant between subjects main effects were

Figure 8

Significant Effectiveness by Shock Threat

Interaction on Backster Test Heart Change Data

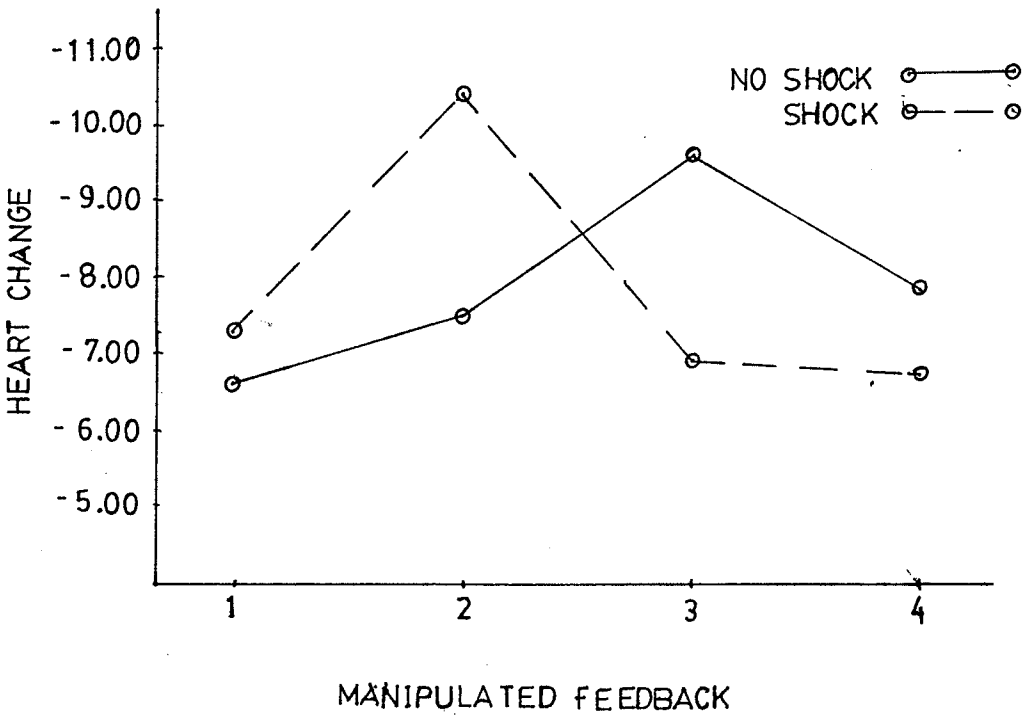
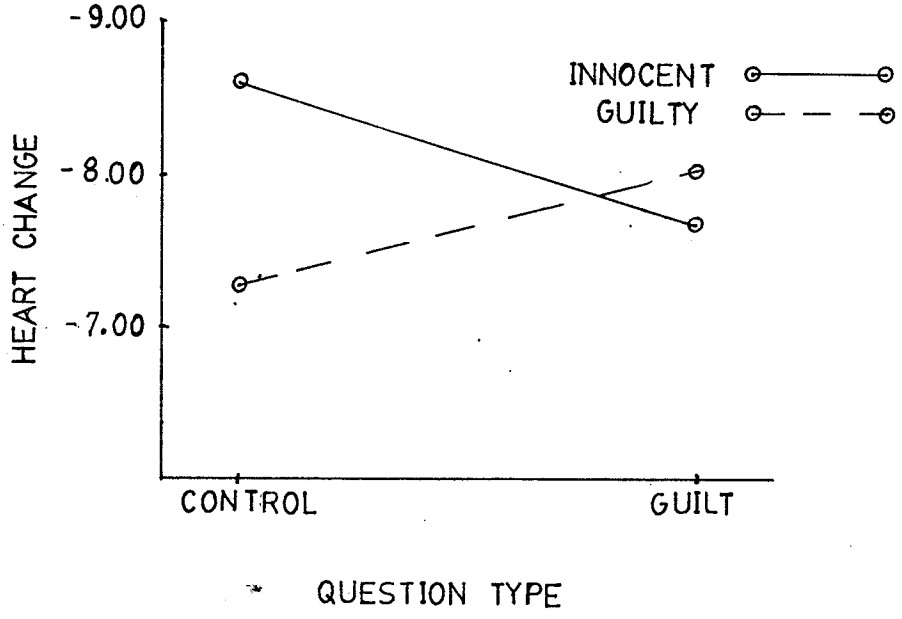

-11.00
-10.00
-9.00
-8.00
-7.00
-6.00
-5.00
HEART CHANGE
NO SHOCK
SHOCK
1
2
3
4
MANIPULATED FEEDBACK

Figure 9

Significant Guilt by Question Type Interaction

on Backster Heart Change Data


-9.00
-8.00
-7.00
HEART CHANGE
INNOCENT
GUILTY
CONTROL
GUILT
QUESTION TYPE

found. Pupil size declined over trials F (1, 176) = 121.9867, $p < .01$ ($\bar{X}_1$ = 104.905, $\bar{X}_2$ = 96.293, $\bar{X}_3$ = 94.782, $\bar{X}_4$ = 95.915). The response to guilty questions was larger than to innocent questions F (1, 176) = 20.0626, $p < .001$ ($\bar{X}_I$ = 97.466, $\bar{X}_G$ = 98.481) and this factor interacted with the actual guilt or innocence of the subjects such that guilty subjects responded most to guilt questions F (1, 176) = 11.44, $p < .001$ (see Figure 9).

Pupil change. The change in pupil size tended to become larger as perceived effectiveness increased F (3, 176) = 5.29, $p < .002$ ($\bar{X}_1$ = 9.827, $\bar{X}_2$ = 10.707, $\bar{X}_3$ = 9.707, $\bar{X}_4$ = 11.820). Guilty subjects had larger responses than innocent subjects F (1, 176) = 5.8222, $p < .02$ ($\bar{X}_I$ = 10.003, $\bar{X}_G$ = 11.028). This factor interacted with question type such that guilty subjects responded most to guilt questions F (1, 176) = 4.82, $p < .029$ (see Figure 11).

Finally, a significant question set factor showed that pupil change increased over question set F (1, 176) = 9.4442, $p < .001$ ($\bar{X}_1$ = 9.626, $\bar{X}_2$ = 10.222, $\bar{X}_3$ = 11.063, $\bar{X}_4$ = 11.151).

Heart rate. No between group main effects were found. Heart rate increased to a peak by the third question set F (1, 176) = 5.103, $p < .05$ ($\bar{X}_1$ = 71.085, $\bar{X}_2$ = 71.931, $\bar{X}_3$ = 72.186, $\bar{X}_4$ = 71.910). Heart rate levels in response to innocent questions were at higher levels for guilty questions F (1, 176) = 5.909, $p < .016$ ($\bar{X}_I$ = 71.967, $\bar{X}_G$ = 71.580). This factor interacted with the actual guilt of the subjects such that innocent subjects showed no response difference between question types while guilty subjects showed a relatively lower heart rate to guilt questions than to innocent questions F (1, 176) =

Figure 10

Significant Guilt by Question Type Interaction

on Guilty Knowledge Test Pupil Size

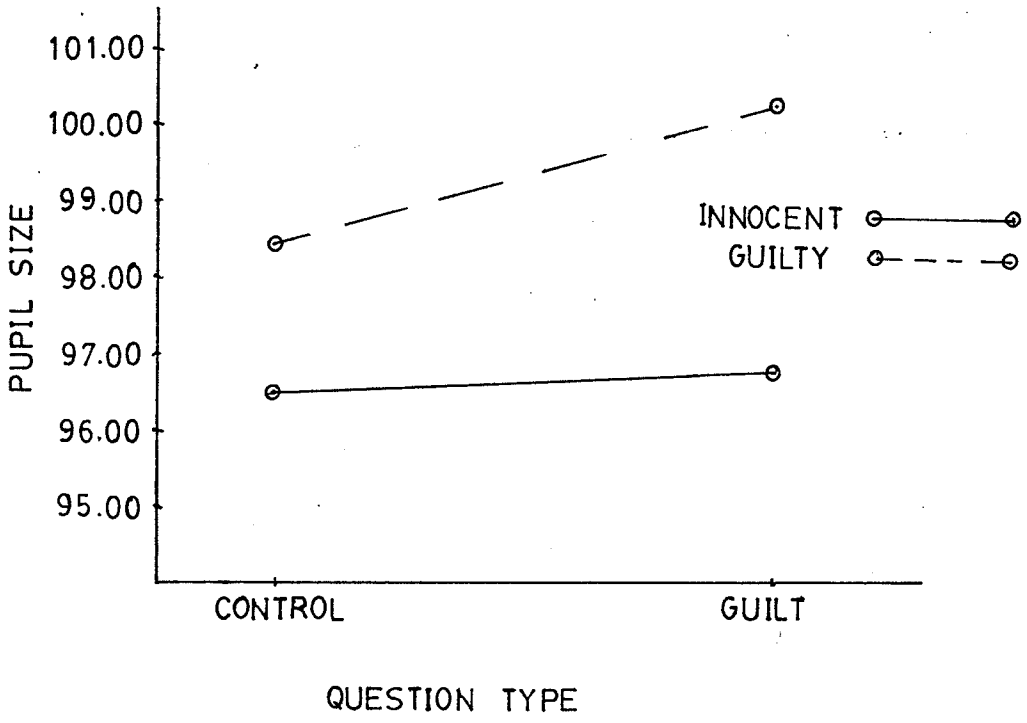
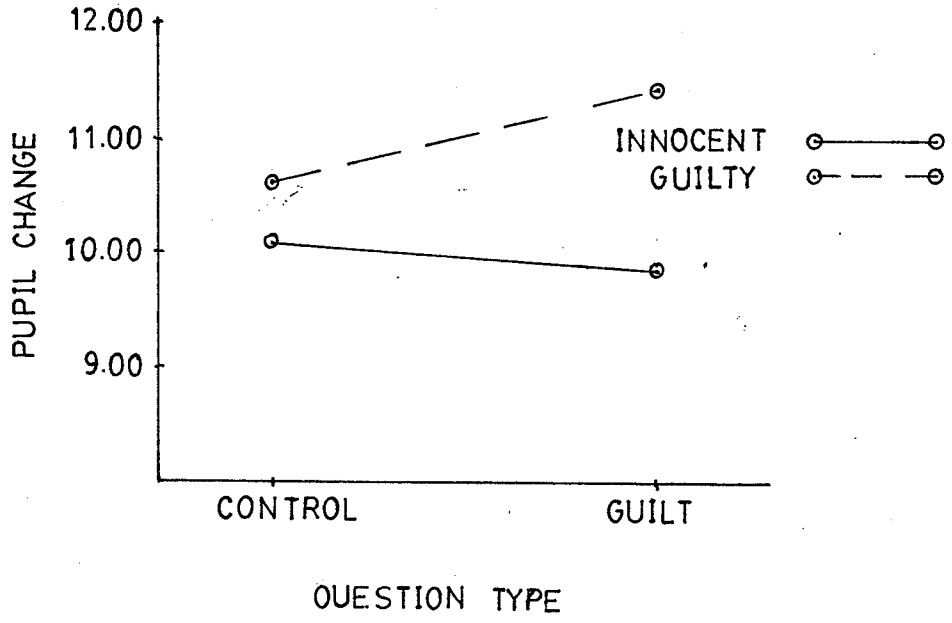

101.00
100.00
99.00
98.00
97.00
96.00
95.00
PUPIL SIZE
INNOCENT
GUILTY
CONTROL
GUILT
QUESTION TYPE

Figure 11

Significant Guilt by Question Type Interaction

on Guilty Knowledge Pupil Change Data


12.00
11.00
10.00
9.00
PUPIL CHANGE
INNOCENT
GUILTY
CONTROL
GUILT
QUESTION TYPE

5.918, $p < .016$ (see Figure 12).

Finally, question type interacted with threat of shock such that those threatened with shock had the lowest heart rates in response to guilty questions F (1, 176) = 6.1495, $p < .05$ (see Figure 13).

Heart change. No significant between groups main effects were found. Manipulated effectiveness interacted with threat of shock F (3,176) = 2.8571, $p < .04$ such that the magnitude of heart rate change increased as manipulated effectiveness increased in the no shock condition but peaked at an uncertainty level in the threat of shock condition (see Figure 14).

Actual guilt of subjects interacted with trials F (1, 176) = 4.4428, $p < .05$ such that the magnitude of heart decreases increased for innocent subjects over trials while remaining the same for guilty subjects (see Figure 15).

Finally, the responses to question types were different such that the magnitude of decrease to innocent questions was less than that to guilty questions F (1, 176) = 5.2957, $p < .05$ ($\bar{X}_I$ =-7.303, $\bar{X}_G$ = -7.941).

Anxiety Scale Data

The state-trait anxiety measure was used at various points to assess levels of anxiety. A four factor analysis of variance with trait anxiety as a covariate was conducted. Of the between factors(manipulated effectiveness, shock and guilt) only guilt was significant as a main effect F (1, 175) = 8.82, $p < .03$ ($\bar{X}_I$ = 38.1, $\bar{X}_G$ = 40.9). Analysis of the within factor (three repeated assessments) showed that anxiety declined over the course of the experiment F (2, 352) = 50.70, $p < .001$ ($\bar{X}_1$ = 41.91, $\bar{X}_2$ = 40.49, $\bar{X}_3$ = 36.89). Guilt interacted with the repeated factor such that anxiety for innocent subjects declined at a

Figure 12

Significant Guilt by Question Type Interaction

on Guilty Knowledge Heart Rate

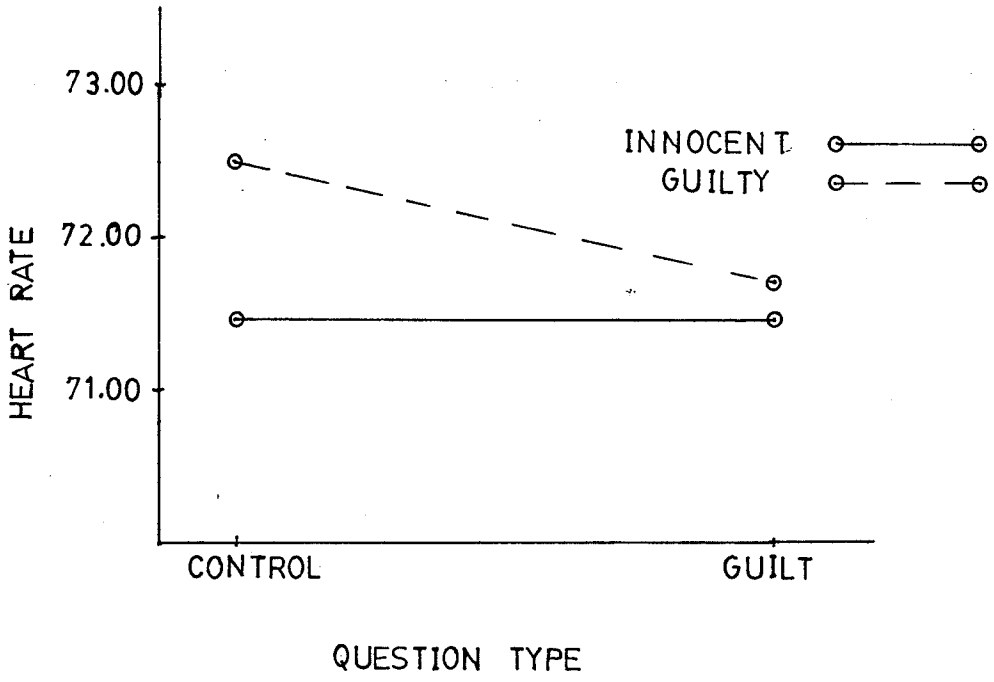

73.00
72.00
71.00
HEART RATE
INNOCENT
GUILTY
CONTROL
GUILT
QUESTION TYPE

Figure 13

Significant Question Type by Shock Interaction

on Guilty Knowledge Heart Rate

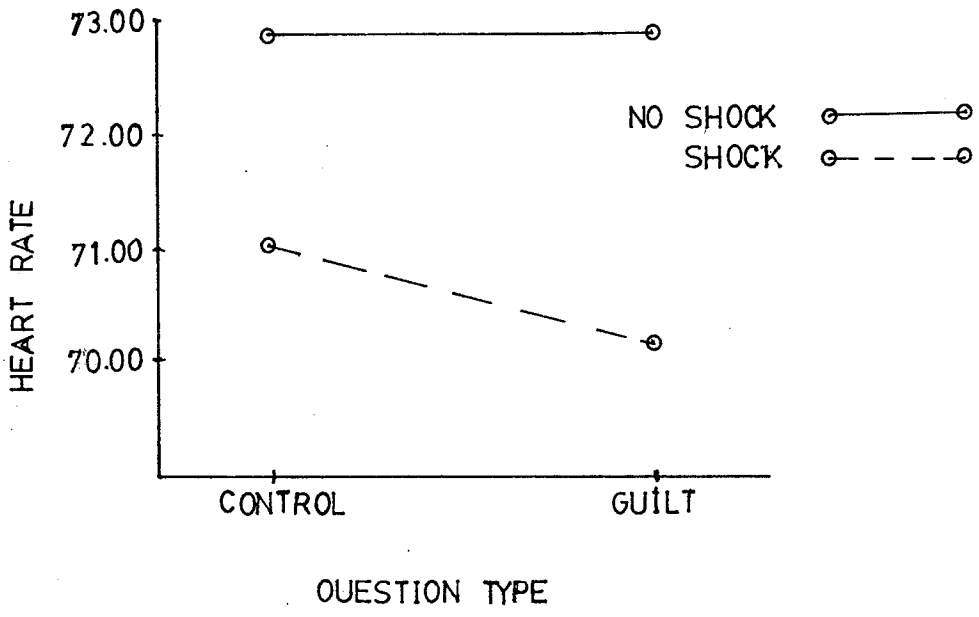

73.00
72.00
71.00
70.00
HEART RATE
NO SHOCK
SHOCK
CONTROL
GUILT
QUESTION TYPE

Figure 14

Significant Effectiveness by Shock Interaction

on Guilty Knowledge Heart Rate

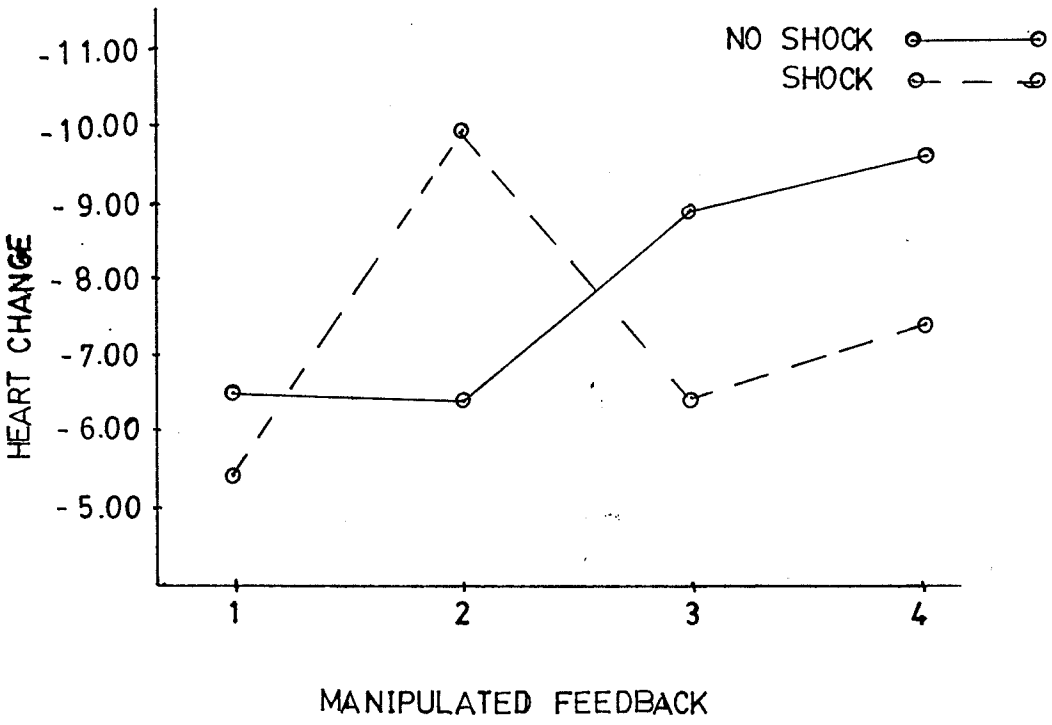
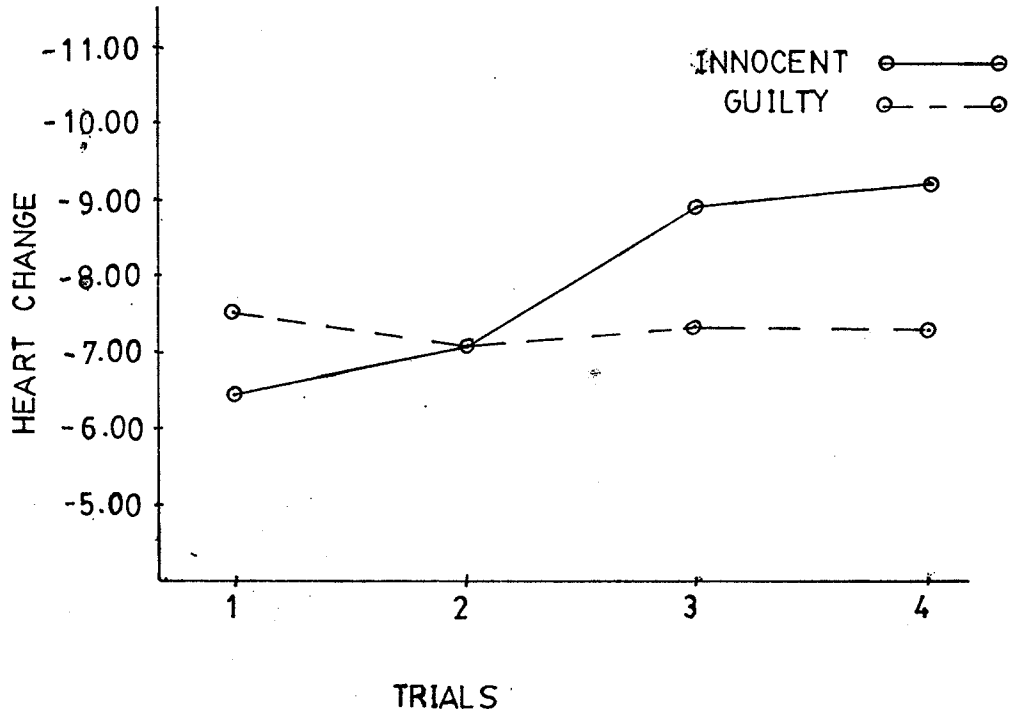

NO SHOCK
SHOCK
-11.00
-10.00
- 9.00
- 8.00
- 7.00
- 6.00
- 5.00
HEART CHANGE
1
2
3
4
MANIPULATED FEEDBACK

Figure 15

Significant Guilt by Trials Interaction on

Guilty Knowledge Heart Change


-11.00
-10.00
-9.00
-8.00
-7.00
-6.00
-5.00
HEART CHANGE
INNOCENT
GUILTY
1
2
3
4
TRIALS

much faster rate than for guilty subjects (see Figure 16).

Post Experimental Questionnaire

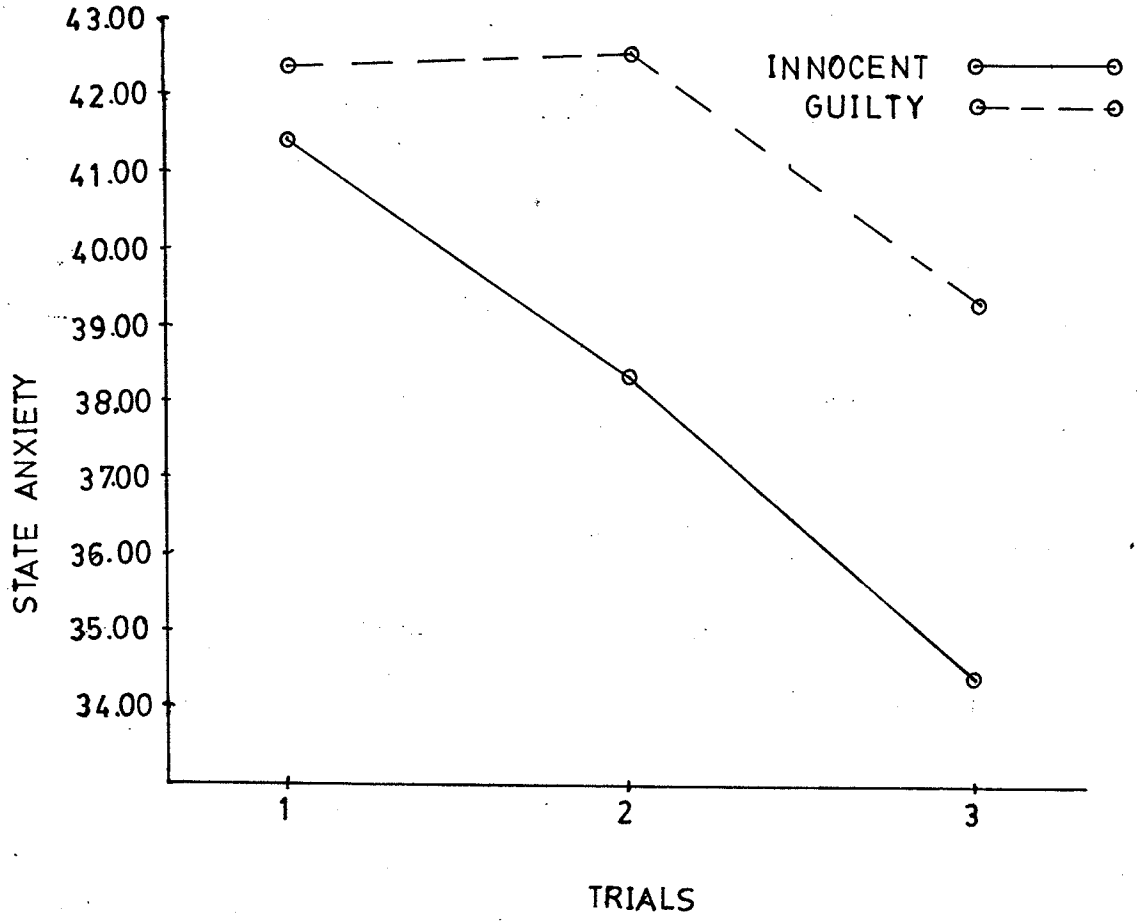
The data from the post experimental questionnaire was treated in the standard fashion as frequency data and also analyzed in a 4 x 2 x 2 analysis of variance format with the possible responses of yes-no or Backster - Guilty Knowledge treated as dichotomous.

Examination of the frequency data indicated that although subjects felt they were most responsive to the Backster test $\chi^2_1 = 11.75$ $p < .05$ they did not favor one test over the other as more effective in a police interrogation $\chi^2 = 3.60$ N.S. The analysis of variance also revealed that more innocent than guilty subjects rated themselves as most responsive to the Backster test F (1, 172) = 12.546, $p < .001$. No other factors were significant. Most subjects would not take the money given the opportunity $\chi^2_1 = 7.13$, $p < .01$ (would take = 77, would not take = 115) and very few (n = 30) felt they were stealing if instructed to take the money $\chi^2 = 74.26$, $p < .01$. Also, fewer subjects (n=73) felt they would be detected than those who felt they would not $\chi^2 = 7.08$, $p < .01$ and 126 indicated that they had a chance of getting off undetected $\chi = 25.26$, $p < .01$. The analysis of variance revealed that more of the no shock group than the shock group expected to be detected. A manipulated effectiveness by guilt interaction revealed that while more innocent subjects in the uncertainty than certainty conditions indicated they expected to get off undetected, the guilty subjects were distributed evenly over effectiveness levels. Subjects were evenly divided between those who accepted the accurate description of the experiment that the interrogator did not know the actual guilt or innocence of the suspects (n=106) and

Figure 16

State Anxiety Interaction of Guilt by

Repeated Assessments


43.00
42.00
41.00
40.00
39.00
38.00
37.00
36.00
35.00
34.00
STATE ANXIETY
INNOCENT
GUILTY
1
2
3
TRIALS

those who were suspicious (n = 83) $\chi^2 = 2.56$ N.S. There were no analysis of variance differences.

Correlated Data

Physiological ranks. The correlations among the ranks of the physiological responses for each measure and question technique are displayed in Table 4. Other than positive relationships between pupil measures and the galvanic skin response specifically in the Backster test there appears to be little relation across different response organs. It is of interest to note that while pupil measures performed poorly in the Backster test they were the measures related to the best performer, the GSR. Within each response organ, that is, pupil size related to pupil differences and heart rate related to heart change, the correlations were significant across both tests suggesting that the measures are somewhat redundant.

Subjective ranks. The estimates of physiological responsiveness made by each subject to questions in the Backster test and topics in the Guilty Knowledge test were correlated with the ranks of actual physiological responses. In Table 5 these correlations and significance values are displayed for both testing techniques.

Three aspects of the results are notable. The correlations are uniformly very low and very few are significant. None would be significant at the family wise error rates appropriate for 30 and 24 comparisons for the Backster and Guilty Knowledge tests, respectively. There appears to be some relation between the GSR and subjective ranks in the Backster test from the consistency of significant results.

The mean subjective ratings of the six crucial questions, three

Table 4

Correlations Among Physiological Measures for Each Testing Technique

	A	B	C	D	E
Backster Test					
A (GSR)	1.00				
B(Pupil Size)	.16*	1.00			
C (Pupil Change)	.23*	.38*	1.00		
D (Heart Rate)	.05	.11	.03	1.00	
E (Heart Change)	-.06	.06	.03	.43*	1.00
Guilty Knowledge Test					
A	1.00				
B	.05	1.00			
C	-.05	.38*	1.00		
D	.03	.09	.11	1.00	
E	.09	.03	-.04	.38*	1.00

*$p < .05$

Table 5

Subjective Ranks and Physiological Ranks Correlated

	1	2	3	4	5	6
Backster Test Questions						
GSR	.12	.14*	.24*	.15*	.10	.06
Pupil Size	-.05	-.13	.03	-.07	.01	-.04
Heart Rate	-.02	-.09	.07	-.01	.05	.06
Pupil Change	.34*	-.14*	.04	-.06	-.03	-.04
Heart Change	.09	-.08	.15*	-.01	.09	.01
Guilty Knowledge Topics						
	1	2	3	4		
GSR	-.08	.14*	.05	.08		
Pupil Size	-.02	-.14*	-.05	-.13		
Heart Rate	.01	-.02	.02	.07		
Pupil Change	-.10	-.11	-.14*	.04		
Heart Change	-.04	.06	-.01	.10		

*$p < .05$

control and three crime relevant, in the Backster test are displayed in Table 6. The differences between the guilty and innocent subjects were significant beyond the .001 level on the multivariate package of six crucial questions and on all of the individual univariate tests for each question. The $\underline{F}$ (1, 176) values ranged from a low of 68.52 to a high of 144.77. Interestingly, the accuracy of judging guilt or innocence from the subjective rank orders (164 correct, 1 inconclusive, 27 incorrect) treated in the same manner as physiological rank orders actually exceeded the accuracy obtained with the GSR (114 correct, 60 inconclusive, 22 incorrect).

Table 6

Mean Subjective Ratings of Crucial Questions in the Backster Test

	Questions					
	Control			Crime Relevant		
	4	6	8	5	7	9
Subjects						
Innocent	2.53	3.29	2.87	5.26	5.21	6.00
Guilty	5.01	5.37	5.41	2.37	2.77	3.56

CHAPTER IV

DISCUSSION

Manipulated Effectiveness

Clear support for the hypothesis proposed by field investigators (Reid and Inbau, 1966) was found for the Backster Control test using the galvanic skin response measure. The rank scores diverged in the appropriate directions (positive for innocent, negative for guilty) as demonstrated effectiveness improved. From a technical standpoint employers of this test would be justified in fostering the belief that the detection equipment is highly effective.

This experiment did little to clarify the relationship between effectiveness and the guilty knowledge test. The significant quadratic trends with both the galvanic skin response and pupil size scores did show that detection as predicted by Davis (1961) was maximized in uncertainty conditions. However, this appeared as a main effect rather than in a guilt by effectiveness interaction which suggested that the manipulation was affecting not only guilty subjects but also innocent subjects. Theoretically, the scores of innocent subjects should have resulted in a horizontal function because such subjects were unaware of the critical items and thus could not be differentially influenced on such items by the effectiveness demonstrations. Simple main effects tests actually revealed this to be the case by showing a relationship between guilt and effectiveness and no relation between innocence and effectiveness (see Table 7). But, the innocent means, although not significantly different from each other, do mirror the pattern of the significant quadratic guilty function.

Table 7

Simple Main Effects, Effectiveness and Guilt with GSR Data

Source of Variation	df	MS	F
Effectiveness and Guilt	3	15.5	5.60*
Effectiveness and Innocence	3	1.46	0.53
Error	176		

*$p < .05$

The possibility exists that some bits of guilty information became available to innocent subjects over the course of the experiment and the innocent subjects with this information were affected appropriately by the manipulation. For example, if a classmate or friend was told by a previously guilty subject that he had been instructed to hide money in his right front pocket the innocent subject could be responsive to that particular item simply because, as Lykken (1974) has emphasized, he recognizes that as a crucial bit of information. This has been a problem with the practical application of the guilty knowledge test since news reports may reveal most or all of the items of information suitable for the test and thus sensitize innocent subjects prior to the interrogation. In the present experiment an honor system of having the subjects promise not to reveal information was designed to prevent such an occurrence. Asking subjects after the experiment if they were aware of information did not at the time lead the experimenter to believe that information was available. Overall, the results are promising but because of the above mentioned problems the hypothesis could be pursued again. An innocent group aware of information could be included in a new study to find out if they are affected as predicted by the effectiveness manipulation.

Although for the reasons outlined above, the point of maximum detection efficiency could not be identified, the experiment showed that a demonstration of perfect detection does not enhance detection rates over other levels with the guilty knowledge test. Thus, in considering both tests the advisability of providing a demonstration of perfect effectiveness depends on the test used.

The discomfort of obtaining different results with two tests

designed to achieve the same ends, detection, may be alleviated by noting that the results of each test conformed to the hypothesis generated by the most frequent users of each test. Examination of the tests may lead to the conclusion that there are crucial differences. For example, resignation with consequent inattentiveness in the certainty of detection condition was part of the reason Davis (1961) suggested that detection rates would not rise in comparison to those in the uncertainty conditions. It has some face validity to suggest that it is easier to be inattentive in the guilty knowledge test than the Backster test because critical items are embedded in a list of similar items (for example, a list of different types of envelopes). Inattentiveness could result in confusion about the items or actually missing the critical item in the list. Questions in the Backster test are directly accusatory in nature and may be more difficult to ignore. For example, in the post experimental questionnaire, 69% (122) of the subjects reported that they were physiologically more responsive to questions in the Backster test, although there were no particular physiological differences found between the two tests to correspond with the self-report differences.

If, as the above reasoning suggests, resignation in the certainty condition begins to interfere with recognition of critical items, methods which increase their salience might prevent this. Of course, in an actual crime situation the interrogator has little control over the material to which the criminal may have attended. Possibly safe combinations, names and addresses may be well learned but whether the money was in a blue pouch or a green pouch might just not have been noticed. In the laboratory several manipulations are available to test the above hypothesis on salience and certainty. Attention can be drawn to the

critical items in the instructions or subjects could be made to learn and rehearse the items (Balloun and Holmes, 1979) or finally, the experimenter could let the subject know that a reward will be paid for each critical item correctly identified. Another technique, untried, could involve repetition of the item in the question. For example, "I am going to ask you about the safe combination, 32 right, 2 left, 6 right (pause). Did you open the safe with a combination of 32 right, 2 left, 6 right?" Such a strategy would increase the probability of correct recognition of the critical item by the guilty subject while also doubling the number of irrelevant items presented. Thus, if Ben Shakhar's (1977) dichotomization theory is correct detection should be enhanced. The purpose of the pause after the first presentation would be to allow a response to occur because subjects have been known to react even when an answer was not required (e.g., Bradley and Janisse, 1976; Gustafson and Orne, 1965; Horvath and Reid, 1972). Details as to whether only the largest of the two responses for each item should be compared or whether they should be summed and compared would have to be worked out empirically.

Threat of Shock

The present experiment is an attempt to follow the recommendations of Podlesny and Raskin (1977) who involved the subject in a simulation of a real crime and attempted to provide meaningful motivation by threatening half of the innocent and guilty subjects with shock. This form of motivation did not prove effective on the rank detection scores except with heart rate change in the Backster test. The finding of a single measure affected by threat of shock should not be considered as anomolous

since such a result fits well with recent studies in the literature using similar measures. Stanners, Coulter, Sweet and Murphy (1979) found heart rate sensitive to shock threat while pupil size and GSR did not evidence such effects. Dumoff (1979), using an ego threat rather than a shock threat to increase arousal, found that heart rate was affected while pupil size was not. In both of the above studies pupillary responses were sensitive to varied levels of mental effort. The continuous data in the present study revealed shock threat related differences in heart rate change for both tests and for heart rate in the Guilty Knowledge test while revealing no shock related pupil differences. Such results suggest that investigations of threat of punishment in relation to cardiovascular measues may be a profitable avenue of research in psychophysiological detection of deception.

One caveat for the interpretation of shock results is that while state anxiety scores were higher for guilty subjects than innocent subjects, there was no extra increment in anxiety for those under threat of shock. Possibly the shock threat manipulation was not effective. Interestingly, other factors could have been at work since the post experimental questionnaire revealed that fewer subjects in the shock group expected to be detected. One can only speculate but subjects may have ameliorated shock threat effects by not admitting or denying the possibility of detection.

Detection

Of major interest in any detection experiment is the performance of the measures in discriminating between the guilty and innocent subjects. Detection frequencies revealed that with the exception of pupil size and pupil change in the Backster test, all measures in both tests,discrimin-

ated beyond chance levels. Of course, even with significant detection, accuracy levels may be far short of the levels desired for practical application. With the GSR, 84% of the judgments made in the Backster test were correct while 74% were correct in the guilty knowledge test. Other measures ranged on down from these levels. In the Backster test, excluding inconclusives, heart rate correctly classified 70%, and heart change 65% of the subjects. In the guilty knowledge test pupil size and heart change correctly classified 64% while heart rate classified 61% and pupil change 57% of the subjects. A discriminant analysis using the ranks upon which the detection scores were based from all of the measures combined failed to improve detection frequencies beyond those found with the GSR. Such a finding suggests that the other measures accounted for the same subjects as the GSR only less accurately.

One possible conclusion from the above is that, of the measures used, the GSR is the only one worth serious study for lie detection because the others are just less adequate relections of the same attendant autonomic activity. However, other considerations may lead one to entertain the hypothesis that the particular experimental situation was less than adequate for maximizing the performance of other measures. For example, although the GSR is highly successful in laboratory studies (Cutrow, Parks, Lucas, and Thomas, 1972) it is a poor indicant of deception in field work (Inbau and Reid, 1966). According to Davis (1966), cardiovascular measures are considered quite effective in the field. The hypothesis could be entertained that intense emotion is needed to maximize differential responding in the cardiovascular system. It may be of interest that in this experiment heart measures outperformed pupil measures in the Backster test, the test rated as most arousing. In a

similar manner it should be noted that the pupil measures only discriminated successfully in the guilty knowledge test. If, as Janisse (1977) has concluded, the pupil is particularly sensitive to mental effort, it would be expected to perform well in this task which requires recall and recognition of critical items. In a guilty information paradigm where memorization of a code number was required, Clark (1975) detected 80% of the subjects with the pupil response. Heart rate in Clark's study was actually the poorest measure with an accuracy rate of 63.33%. Bradley and Janisse (1979) had success with the pupillary response in a guilty knowledge test but no other measure had been included so no comparisons were possible. The design of studies specifically emphasizing emotional or cognitive factors when comparing these measures would enhance our understanding of their relative effectiveness.

Beyond comparing the specific measures to each other it is necessary to know how detection frequencies in this experiment compared with those in similar studies. The accuracy of the Backster test was not as high as the 98% correct level found by Raskin and Hare (1978) in their mock crime study. This may be, at least in part, a function of the scoring methods employed. Raskin and Hare (1978) used a "numerical evaluation technique" that included "on the spot evaluation" of the records to estimate the degree of deceptiveness. The present data gathering was limited to objective quantification based on the physical measurement of the designated response. Raskin and Hare (1978) also objectively quantified the specific response of interest, but unfortunately, failed to report accuracy levels. One serious problem related to reporting only the results of a more subjective evaluation is that the reader unacquainted with the techniques cannot surmise how advances in scoring

could be made. For example, Raskin and Hare (1978) must have improved the subjective technique incredibly from a similar study four years earlier (Barland and Raskin, 1975) where, including inconclusiveness, a 53% correct level was achieved. Lykken (1978) has complained specifically about the subjective nature of Raskin's scoring technique.

The present study included levels of a factor, effectiveness, which in theory should and did in practice suppress detection. For example, only two of twelve guilty subjects were detected on the Backster test in the no-shock condition after a demonstration of ineffectiveness whereas eleven of twelve guilty subjects were detected after a demonstration of effectiveness. In an earlier study by Bradley and Janisse (1979), most subjects with no "lie detection" experience rated the lie detector as very effective. Thus studies without any demonstration may be dealing with subjects already convinced of the equipment's effectiveness and such subjects should, therefore, be easy to detect. Further studies could look at this person variable to discover if it is easy to manipulate. Bradley and Janisse (1979) found the subjective estimate of effectiveness not related to detection but surmised that the few subjects who rated the technique as ineffective may have become uncomfortable with the guess when exposed to the elaborate equipment (pupillometer) and procedures in the experiment.

The detection results with the guilty knowledge test were a respectable 74%, but not as high as the 98% correct level reported by Davidson (1968) or the 90% levels achieved by Podlesny, Raskin and Barland (1976). These studies used six and five guilty items, respectively, rather than just four items as used in this experiment. The number of false positives decreases directly as a function of the number of items.

In addition, of course, the major manipulation in this study was designed to suppress detection in some conditions.

Comparing detection accuracies of the two tests in this experiment is a complex enterprise involving, as was discussed above, the physiological measures used and, as will be discussed here, some of the parameters of the tests. The following examples will help clarify this point.

Geldreich (1941) presented a series of 20 to 50 buffer items before any guilty information and found 100% detection accuracy. Ben Shakhar, Lieblich and Kugelmass (1975) improved detection by increasing the set size in which a single critical item was embedded. Increasing the number of critical items and question sets improves effectiveness in the guilty knowledge test (Lykken, 1974). Even though detection accuracy declines over each repeated administration of a guilty knowledge test (Balloun and Holmes, 1979), Bradley and Janisse (Note 1), found that comparing the average critical rank to the average of the non-critical ranks over trials results in higher levels of detection than obtained on the first trial alone. An interesting note is that the present experiment found that detection rates for one type of test did not decline if it followed the administration of another type of test. That is, 57 subjects were detected when the Backster test came first and the same number were detected when the Backster test followed the guilty knowledge test. In a similar manner virtually the same number of subjects were classified correctly when the guilty knowledge test was first as when it was second, that is, 75 and 67, respectively. Podlesny, Raskin and Barland (1978) found a more exact specification of control questions, as to the period of time referred to, did not increase the accuracy of control

question tests. Two, untested to date, procedures included in the Backster test may alter its effectiveness. Subjects are given additional trials if the initial three tests are inconclusive and second, subjects are asked between trials about the subjective response to control questions. Podlesny, Raskin and Barland (1978) made adjustments as they thought necessary in attempts to increase the salience of control questions. Such procedures should be specified and studied as to their effects. In summary, a perusal of the literature reveals a see-saw battle between the two techniques as to which is more effective. A head count of which is efficacious more often would not increase our knowledge. Only a systematic study of the above mentioned parameters would yield answers and the answers would not reveal which technique was more effective but would reveal the conditions under which one technique would perform at optimal levels.

Subjective Ratings

A controversy over the utility of control question tests in the detection of deception has now raged for some years across several journals (for example, Lykken, 1974; Lykken, 1978; Podlesny and Raskin, 1978; Raskin and Podlesny, 1979). At issue is whether the control question tests can, as they were designed to do, evoke the attention and thus the autonomic response of innocent subjects to the control questions while focussing the guilty suspect's attention on the "crime relevant" questions. The rationale is that control questions in the test are of an emotional nature because they ask suspects about possible misdemeanors or crimes in which they could have been involved in the past. The innocent person, it is hypothesized, will be responsive because the implications of such questions reflect ill on his character. In effect, not only is he accused of the particular crime but now he is being regarded as a highly suspicious

character. An aggravating factor is that the questions are of such a general nature that it is easy to feel uncomfortable with one's answer. In comparison the guilty questions are easier to answer because on these the innocent suspect knows he is truthful. The guilty person, ignores the control questions because he realizes he is being interrogated for the specific crime and questions about his past life are irrelevant. Lykken (1979) has decided it is naive to think control questions would elicit the attention of, or cause much concern for, innocent suspects whereas Raskin and Podlesny (1979) have decided that control questions should cause great concern for innocent suspects. The controversy has not been easy to resolve because of the somewhat confusing story on detection accuracy. For instance, field workers who claim high rates of success for the control question test also in general claim much higher accuracy rates than found by scientific assessment of field results (Horvath, 1977, Bersh, 1969). In a similar manner Lykken (1979) has questioned Podlesny and Raskin's (1977) high rates of success in the laboratory because of their failure to specify their scoring technique in an objective manner. Of course, such knowledge is fundamental to encourage replications of results. Neither side has searched for other types of evidence to bolster their arguments. Fortunately, in the present experiment data were collected which bear directly on how the suspects perceived the control and guilty questions. After the interrogation but before they received the judgment of guilt or innocence, subjects rank ordered the series of questions in the Backster test assigning a one to the question they believed evoked the largest autonomic response and a nine to the least autonomically evocative question. This rating was made with no knowledge of the interrogation results or actual autonomic responsiveness

and, very importantly, under the assurance that the ratings would not be used in the judgment process. In fact, the suspect did not physically yield the rating sheet until the judgment was delivered and in the case of shock threat the shock electrodes were removed.

The display of mean ratings for the six crucial questions, three control and three crime relevant, in the results section show that innocent subjects perceived themselves as responding more to questions about their past life while guilty suspects perceived themselves as responding more to questions about the crime. The finding that these data (scored for detection by the same method used for the autonomic data) yielded an accuracy rate exceeding the GSR results ($\chi^2 = 8.6$ $p < .01$), suggests that subjects regard their responses to questions in a manner congruent to the intent of the test. Further, this state of affairs is not always expressed autonomically. Indeed, the results (Table 5) showed that correlations between estimates of responsivity and actual responsivity were so low as to suggest subjects were reacting less to their autonomic responsivity than to the cognitive impact of the questions. In sum subjects who were ignorant of the rationale behind the test saw themselves as responding autonomically in the manner predicted by the theory and thus it may not at all be as Lykken (1978) states naive to assume innocent suspects will be more concerned with control questions while guilty suspects will be more concerned with crime relevant questions.

Legal Issues

This study is directed at one of the more sensitive issues in the practice of the psychophysiological detection of deception. Field investigators have felt the necessity of suggesting that lie detectors are infallible in order to obtain high detection rates. In a limited technical

sense, results with the Backster test, in this experiment, supported that position. However, there is a moral and ethical dilemma with a delicious twist of irony that the present author sees no evidence of being addressed. How can those engaged in the business of lie detection lie about or at least misrepresent the lie detector's capabilities. It borders on the ludicrous that those portraying themselves on the side of "truth" and on the leading edge of scientific endeavor should foster this myth. The results in neither the laboratory, nor the field offer perfect levels of detection.

The present study suggests that debunking the myth would reduce the Backster test's accuracy. However, no misrepresentation would be necessary for the guilty knowledge test since demonstrations of fallibility did not detract from detection. It would be an interesting strategy to suggest that polygraphers use the guilty knowledge test exclusively since they could accurately represent it as a moderately effective technique without diminishing that effectiveness.

Summary and Recommendations

Manipulation of demonstrated effectiveness altered detection levels in both the Backster and guilty knowledge tests. Threat of shock had an effect on continuous heart data but had no effect on detection ranks other than with heart rate in the Backster test. The galvanic skin response was the most effective measure in both tests and while other measures were significant in distinguishing between the guilty and innocent, discriminant analyses revealed that as an aggregate of measures they did not contribute to greater accuracy than that of the GSR alone.

The results of this study support two assumptions commonly made by field workers in employing the Backster Control test. Demonstrations of

perfect effectiveness enhance accuracy and innocent suspects perceive themselves as more responsive to control questions than guilt questions. However, because it appears necessary to factually misrepresent accuracy levels to obtain optimal performance from the Backster test, polygraphers have an ethical problem to resolve. Should they continue to claim that detection is foolproof or should they accept the lower accuracy rates that result from claims of the correct but more modest levels of accuracy? It is beyond the scope of this paper to deal with such a question but the guilty knowledge test, as an alternative, can be recommended. Accurate demonstrations of modest success did not detract from its effectiveness. The present study did not isolate, in any specific way, the range of demonstrated effectiveness that promotes optimal detection for the guilty knowledge test. One possibility is that other than for demonstrations of complete ineffectiveness the guilty knowledge test is impervious to such manipulations. Another possibility is that there is an interaction between effectiveness and salience of items. The prediction to be tested is that low salience items would be affected very much by effectiveness manipulations while high salience items would not.

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

Threat of shock had little effect on detection accuracy. Such a result could encourage the laboratory study of detection since such threat was designed to be a step on a continuum shared with legal

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

Reference Notes

1. Bradley, M.T. and Janisse, M.P. Pupillary responsivity in the detection of deception paradigm. Paper presented at the Ninth Pupil Colloquium, Iowa City, May 30 1975.

2. Bradley, M.T., and Janisse, M.P. Machiavellianism and the pupillary response in the detection of deception. Paper presented at the annual meeting of the Canadian Psychological Association, Toronto, 1976.

3. Bradley, M.T. and Janisse, M.P. Pupil size and lie detection: The effect of confidence on detection rates. Paper presented at the annual meeting of the American Psychological Association, San Francisco, 1977.

4. Clark, W.R. A comparison of pupillary response, heart rate, and GSR during deception. Unpublished manuscript. 1975.

5. Day, D.A. Lie detection. Unpublished dissertation, University of Windsor, 1972.

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

References

Alpert, M., Kuritzberg, R.L., and Friedhoff, A.J. Transient voice changes associated with emotional stimuli. Archives of General Psychiatry, 1963, 8, 362-365.

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

Backster, C. Technique fundamentals of the tri-zone polygraph test. New York: Backster Research Foundation, Inc., 1969.

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

Barland, G.H. Reliability of polygraph chart interpretation. Polygraph, 1973, 1, 192-206.

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

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

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

Berkhout, J., Walter, D.O. and Adey, W.R. Autonomic responses during a replicable interrogation. Journal of Applied Psychology, 1970, 54, 316-325.

Berrien, F.K. Ocular stability in deception. Journal of Applied Psychology, 1942, 26, 55-63.

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

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

Block, J.D. A study of affective responsiveness in a lie-detection situation. Journal of Abnormal and Social Psychology, 1957, 55 11-15.

Block, J.D., Rouke, F.L., Salpeter, M.M., Tobach, E., Kubis, J.F., and Welch, L. An attempt at reversal of the truth-lie relationship as measured by the psychogalvanic response. Journal of Psychology, 1952, 34, 55-66.

Bradley, M.T., and Janisse, M.P. Pupil size and lie detection: The effect of certainty on detection. Psychology (in press).

Brown, C.C. The techniques of plethysmography. In C.C. Brown (Ed.), Methods in psychophysiology (Ch.2). Baltimore: Williams and Wilkins, 1967.

Boring, E.G. Sensation and perception in the history of experimental psychology. New York: Appleton-Century-Crofts, 1942.

Burak, B.A. Critical analysis of the theory, method and limitations of the "lie detector." Journal of Criminal Law and Criminology, 1955, 46, 414-426.

Burtt, H.E. Motor concomitants of the association reaction. Journal of Experimental Psychology, 1936, 19, 51-63.

Chappell, M.N. Blood pressure changes in deception. Archives of Psychology, New York, 1929, 17, 1-39.

Cutrow, R.J., Parks, A., Lucas, N., and Thomas, K. The objective use of

multiple physiological indices in the detection of deception. Psychophysiology, 1972, 9, 578-588.

Darrow, C.W. The galvanic skin reflex (sweating) and blood pressure as preparatory and facilitative functions. Psychological Bulletin, 1936, 33, 73-94.

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

Davis, R.C. Physiological responses as a means of evaluating information. In A.D. Biderman and H. Simmer (Eds.), The manipulation of human behavior. (Ch.4). New York: Wiley, 1961.

Day, D.A., and Rourke, B.P. The role of attention in "lie-detection." Canadian Journal of Behavioral Science, 1964, 6, 270-276.

Edel, E.C., and Jacoby, J. Examiner reliability in polygraph chart analysis. Identification of physiological responses. Journal of Applied Psychology, 1975, 6, 632-634.

Edelberg, R. Electrical activity of the skin. Its measurement and uses in psychophysiology. In N.S. Greenfield and H.A. Sternbach (Eds.), Handbook of psychophysiology. New York: Holt, Rinehart and Winston, 1972.

Ferguson, R.J. The polygraph in private industry. Springfield, Ill.: Charles C. Thomas, 1966.

Geldreich, E.W. Studies of the galvanic skin response as a deception indicator. Transactions of the Kansas Academy of Science, 1941, 44, 346-351.

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

Gustafson, L.A., and Orne, M.T. The effects of task and method of stimulus presentation on the detection of deception. Journal of Applied Psychology, 1964, 48, 383-387.

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

Hakerem, G. Pupillography as a tool in the assistance of CNS functions and drug eggects. In W.L. Smith (Ed.), Drugs and cerebral function. Springfield, Ill.: Thomas, 1970.

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

Horvath, F.S. and Reid, J.S. The polygraph silent answer test. The Journal of Criminal Law, Criminology and Police Science, 1972, 63, 285-293.

Inbau, F., Moenssens, A.A., and Vitullo, L.R. Scientific police investigation. Philadephia: Chilton Book Co., 1972.

Inbau, F.E. and Reid, J.E. Lie detection and criminal interrogation. (3rd. ed.) Baltimore: Williams and Wilkins, 1953.

Inbau, F.E., and Reid, J.E. Truth and deception. Baltimore: Williams and Wilkins, 1966.

Janisse, M.P. Pupillometry: The psychology of the pupillary response. Washington, D.C.: Hemisphere, 1977.

Keeler, L. A method for detecting deception. American Journal of Police Science, 1930, 1, 38-51.

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

Kugelmass, S., Lieblich, I., and Bergman, Z. The role of "lying" in psychophysiological detection. Psychophysiology, 1967, 3, 312-315.

Kugelmass, S., Lieblich, I., Ben-Ishai, A., Opatowski, A., and Kaplan, M. Experimental evaluation of galvanic skin response and blood pressure change indices during criminal interrogation. Journal of Criminal Law, Criminology, and Police Science, 1968, 59, 632-635.

Landis, C. and Gullette, R. Studies of emotional reactions: III. Systolic blood pressure and inspiration-expiration ratios. Journal of Comparative Psychology, 1925, 5, 221-253.

Landis, C. and Wiley, L.E. Changes of blood pressure and respiration during deception. Journal of Comparative Psychology, 1926, 6, 1-19.

Langfeld, H.S. Psychophysical symptoms of deception. Journal of Abnormal Psychology, 1921, 15, 319-328.

Larson, J.A. Lying and its detection: A study of deception and deception tests. Chicago: University of Chicago Press, 1932.

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

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

Luria, A.R. The nature of human conflicts: Or emotion conflict and will. Translated by W.H. Gantt. New York: Liveright, 1932.

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

Lykken, D.T. The validity of the guilty knowledge technique. The effects of faking. Journal of Applied Psychology, 1960, 44, 258-262.

Lykken, D.T. Psychology and the lie detector industory. American Psychologist, 1974, 29, 724-739.

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

Lyon, V.W. Deception tests with juvenile delinquents. Journal of Genetic Psychology, 1936, 48, 494-497.

Marston, W.M. Psychological possibilities in the deception tests. Journal of Criminal Law and Criminology, 1921, 11, 551-570.

Obermann, G.E. The effect on the Berger rhythm of mild affective states. Journal of Abnormal and Social Psychology, 1939, 34, 84-95.

Orlansky, J. An assessment of lie detection capabilities. Tech. Rep. 62-16, prepared for Institute for Defense Analysis, Research and Engineering Suport Division, AD 603860. Springfield, Va. Clearing House of Scientific and Technical Information, 1962.

Orne, M.T. Implications of laboratory research for the detection of deception. Polygraph, 1973, 2, 169-199.

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

Podlesny, J.A., and Raskin, D.C. Physiological measures and the detection of deception. Psychological Bulletin, (in press).

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-99-0001, U.S. Department of Justice). Salt Lake City, Utah: Department of Psychology,

August, 1976.

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

Raskin, D.C. Psychopathology and detection of deception in a prison population. (Report No. 75-1, Contract 75-NI-99-0001, U.S. Department of Justice). Salt Lake City, Utah: Department of Psychology, August, 1976.

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

Reid, J.E. and Inbau, F.E. Truth and deception: The polygraph ("lie-detector") technique. Baltimore: Williams and Wilkins, 1966.

Sokolov, Y.N. Perception and the conditioned reflex. New York: Pergamon PRess, 1963.

Spielberger, C.D., Gorsuch, A.L., and Lushene, R.E. Manual for the State-Trait Anxiety Inventory. Palo Alto: Consulting Psycholigsts Press, 1970.

Stanners, R.F., Coulter, M., Sweet, A.W., and Murphy, P. The pupillary response as an indicator of arousal. Motivation and Emotion, (in press).

Sternbach, R.A. Principles and psychophysiology: An introductory text and readings. New York: Academic Press, 1966.

Thackray, R.I., and Orne, M.T. Effects of the type of stimulus employed and the level of subject awareness on the detection of deception. Journal of Applied Psychology, 1960, 52, 234-239. (b)

Thackray, R.E., and Orne, M.T. A comparison of physiological indices in detection of deception. Psychophysiology, 1968, 4, 329-339.(a)

Trovillo, P.V. A History of lie detection. Journal of Criminal Law and Criminology, 1939, 29, 848-881.

Van Buskirk, D., & Marcuse, F.L. The nature of errors in experimental lie detection. Journal of Experimental Psychology, 1954, 47, 187-190.

Waid, W.M., Orne, M.T., and Wilson, S.K. Effects of level of socialization on electrodermal detection of deception. Psychophysiology, 1979, 16, 15-22.

Woodworth, R.S., and Schlosberg, H. Experimental psychology, (Rev. ed.). New York: Holt, Rinehart and Winston, 1965.

Young, F.A. Studies of pupillary conditioning. Journal of Experimental Psychology, 1958, 55, 97-110.

APPENDIX A

1. Is your name _____?
2. Regarding the ring and the watch, do you intend to answer the questions about them truthfully?
3. Are you convinced that I will only ask questions on this test that you have already okayed?
4. Before the age of 18 did you ever steal any money?
5. Did you take that ring?
6. While you were in school did you ever take anything of value?
7. Did you take that ring from the desk?
8. Did you take that watch from room 702?
9. Before the age of 18 did you ever cheat anyone who trusted you?
10. Do you have that ring with you now?

APPENDIX B

Please close the door, set the timer located on the desk for three minutes, sit down and relax. At the end of the three minutes open the door and wait for the experimenter to arrive and escort you to the laboratory. At some point he will accuse you of committing a crime while you were in the room. You, of course, are innocent and must deny committing the crime. To prove your innocence you must submit to an interrogation procedure using physiological measures much as would be done in a police investigation. *(If you are judged guilty by these measures, even though innocent, you will receive a painful but not permanently damaging electric shock. If you are judged innocent you will not be shocked.) Of course you are free to withdraw from the experiment with no penalty if you so desire. Once you have read these instructions conceal them on your person.

*Inserted in the threat of shock condition.

APPENDIX C

Please close the door, set the timer located on the desk for three minutes, read and carry out these instructions carefully.

The room you are in is part of a store. In the top drawer of the desk you will find a white envelope. Open it. Steal the one dollar from the envelope and put it in your right front pants' pocket. Conceal these instructions on your person and wait for the duration of the three minutes. Once the time is up open the door, sit down, relax and wait for the experimenter to come and escort you to the laboratory.

At some point the experimenter will accuse you of committing a crime while you were in the room. Even though you are guilty of this crime deny any involvement or knowledge of such an act. You may give your name and answer freely questions about matters not directly relevant to the crime.

To find out if you are guilty you will be subjected to an interrogation procedure using physiological measures such as would be done in a police interrogation. If you are judged guilty by the measures you must forfeit the money *(and you will receive a painful but not permanently damaging electric shock). If you are judged innocent you will be allowed to keep the money *(and will not receive shock). Of course you are free to withdraw from the experiment with no penalty if you so desire.

*Inserted in the threat of shock condition.

APPENDIX D

A) Is your name _____?

B) Regarding the money, do you intend to answer all questions about it truthfully?

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

D) Before the age of 18 did you ever steal any money?

E) Did you take the money?

F) While you were in school did you ever take anything of value?

G) Did you steal the money from the room?

H) Before the age of 18 did you ever cheat anyone who trusted you?

I) Do you know what happened to the money?

APPENDIX E

A) Regarding the envelope:

a) was it a large sized brown envelope?
b) was it a small sized blue envelope?
c) was it a large sized white envelope?
d) was it a small sized white envelope?
e) was it a medium sized white envelope?
f) was it a large sized blue envelope?

B) Regarding the room:

a) was it in an office?
b) was it in a store?
c) was it in an hotel?
d) was it in a home?
e) was it in a bank?
f) was it in a tavern?

C) Regarding where the money is now:

a) is it in your shirt's right front pocket?
b) is it in your pants' left front pocket?
c) is it in your pants' right front pocket?
d) is it in your pants' left rear pocket?
e) is it in your pants' right rear pocket?
f) is it in your shirt's left front pocket?

D) Regarding the money:

a) was it 30 dollars?
b) was it 5 dollars?
c) was it 10 dollars?
d) was it 1 dollar?
f) was it 2 dollars.

APPENDIX F

Below is the set of questions from the Backster Control Question Test. Whether you were guilty or not you may have felt that you responded physiologically more to some questions than to other questions. Rank order by putting a 1 beside the question that you think you responded to physiologically the most, 2 beside the next and so on until you rank as 9 the question you felt you responded to the least physiologically.

A) Is your name _____?

B) Regarding the money, do you intend to answer all questions about it truthfully?

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

D) Before the age of 18 did you ever steal any money?

E) Did you take the money?

F) While you were in school did you ever take anything of value?

D) Did you steal the money from the room?

H) Before the age of 18 did you ever cheat anyone who trusted you?

I) Do you know what happened to the money?

Write out your answers:

1. Which of the two techniques, the Backster Control Test or the Guilty Knowledge Test do you think you responded the most to physiologically?

2. Which of the two techniques do you think would be the most effective for detecting individuals guilty of a crime?

Below are the questions that were asked in the Guilty Knowledge Test. If you are guilty circle the correct alternative as you remember it.

A) Regarding the envelope:

a) was it a large sized brown envelope?

b) was it a small sized blue envelope?

c) was it a large sized white envelope?

d) was it a small sized white envelope?

e) was it a medium sized white envelope?

f) was it a large sized blue envelope?

B) Regarding the room:

a) was it in an office?

b) was it in a store?

c) was it in a hotel?

d) was it in a home?

e) was it in a bank?

f) was it in a tavern?

C) Regarding where the money is now:

a) is it in your shirt's right front pocket?

b) is it in your pants' left front pocket?

c) is it in your pants' right front pocket?

d) is it in your pants' left rear pocket?

e) is it in your pants' left front pocket?

f) is it in your shirt's left front pocket?

Continued...

D) Regarding the money:

a) was it 30 dollars?

b) was it 5 dollars?

c) was it 20 dollars?

d) was it 10 dollars?

e) was it 1 dollar?

f) was it 2 dollars?

When you were being questioned, whether you were guilty or innocent you may have felt that you gave a larger physiological response (that is, a response more likely to be judged as deceptive) to some sets of questions more than to others. Rank order, by writing the number 1 beside the set of alternatives you responded the most to physiologically, 2 besides the second most, then 3 and finally 4 beside the questions to which you think you responded the least to physiologically.

Circle the answer of your choice:

1. Would you have taken the money if you knew where it was and were given the choice to take or not take it in the instructions?

YES NO

2. Did you feel like you were really stealing when you took the money even though you had been instructed to?

YES NO

3. Did you expect to be detected prior to the interrogation?

YES NO

4. Did you think you had a chance of getting off undetected?

YES NO

5. Did you think that the interrogator might know who was guilty or innocent even before the interrogation began?

YES NO

6. What did you think of the experiment in general?
(Use the space below for comments).

APPENDIX G

Table 8

Backster Test Pupil Size ANOVA

Source	df	MS	F
A (Effectiveness)	3	6654.98	2.0095
B (Shock)	1	3542.23	1.0696
A B	3	891.22	0.2691
C (Guilt)	1	859.68	0.2596
A C	3	2333.07	0.7045
B C	1	240.78	0.0727
A B C	3	3491.93	1.0544
Error 1	176	3311.71	
D (Trials)	2	14177.99	95.4710*
D A	6	136.76	0.9209
D B	2	26.20	0.1764
D A B	6	226.23	1.5907
D C	2	36.13	0.2433
D A C	6	92.56	0.6233
D B C	2	92.75	0.6246
D A B C	6	204.29	1.3757
Error 2	352	148.51	
E (Question Set)	2	12739.31	224.2630*
E A	6	64.20	1.1301
E B	2	92.47	1.6279
E A B	6	55.62	0.9792
E C	2	8.06	0.1419
E A C	6	83.53	1.4704
E B C	2	63.17	1.1120
E A B C	6	64.61	1.1374
Error 3	352	56.81	

Continued...

Source	df	MS	F
F (Question Type)	1	1747.95	57.9712*
F A	3	14.58	0.4837
F B	1	6.40	0.2121
F A B	3	25.15	0.8342
F C	1	166.54	5.5232*
F A C	3	19.62	0.6508
F B C	1	0.01	0.0004
F A B C	3	51.20	1.6982
Error 4	176	30.15	1.6982
D E	4	1216.77	33.8565*
D E A	12	54.41	1.4600
D E B	4	29.74	0.7980
D E A B	12	39.41	1.0575
D E C	4	88.77	2.3820
D E A C	12	28.90	0.7755
D E B C	4	46.87	1.2577
D E A B C	12	46.50	1.2478
Error 5	704	37.27	
D F	2	213.38	12.6342*
D F A	6	11.84	0.7012
D F B	2	6.65	0.3937
D E A B	6	20.46	1.2115
D F C	2	8.94	0.5293
D F A C	6	8.33	0.4930
D F B C	2	32.72	1.9374
D F A B C	6	15.20	0.9001
Error 6	352	16.89	

Continued...

Source	df	MS	F
E F	2	796.75	32.1935*
E F A	6	11.29	0.4561
E F B	2	36.04	1.4562
E F A B	6	9.36	0.3784
E F C	2	60.93	2.4618
E F A C	6	12.57	0.5080
E F B C	2	0.38	0.0153
E F A B C	6	17.97	0.7262
Error 7	352	24.75	
D E F	4	26.76	1.3514
D F A	12	21.01	1.0610
D E F B	4	20.66	1.0435
D E F A B	12	19.80	1.0000
D E F C	4	24.86	1.2531
D E F A C	12	14.44	0.7290
D E F B C	4	9.74	0.4917
D E F A B C	12	6.36	0.3211
Error 8	704	19.80	

*$p < .05$ df corrected (Geisser Greenhouse, 1958)

Table 9

Backster Test Heart Rate ANOVA

Source	df	MS	F
A (Effectiveness)	3	2132.78	0.9337
B (Shock)	1	2851.84	1.2485
A B	3	1104.58	0.4836
C (Guilt)	1	577.96	0.2530
A C	3	2204.79	0.9652
B C	1	376.85	0.1650
A B C	3	1485.97	0.6505
Error 1	176	2284.29	
D (Trials)	2	73.56	2.3526
D A	6	38.72	1.2384
D B	2	42.63	1.3634
D A B	6	51.19	1.6373
D C	2	16.60	0.5308
D A C	6	18.08	0.5782
D B C	2	9.01	0.2881
D A B C	6	14.87	0.4757
Error 2	352	31.27	
E (Question Set)	2	6.66	0.3636
E A	6	9.13	0.4985
E B	2	1.06	0.0577
E A B	6	19.15	1.0455
E C	2	8.44	0.5157
E A C	6	8.75	0.4780
E B C	2	5.59	0.3052
E A B C	6	15.79	0.8619
Error 3	352	18.32	

Continued...

Source	df	MS	F
F (Question Type)	1	55.71	4.4299*
F A	3	26.06	2.0720
F B	1	19.94	1.5856
F A B	3	7.17	0.5698
F C	1	289.88	23.0492*
F A C	3	24.09	1.9152
F B C	1	2.04	0.1622
F A B C	3	11.94	0.9491
Error 4	176	12.58	
D E	4	39.01	3.0020
D E A	12	8.76	0.6744
D E B	4	11.13	0.8563
D E A B	12	9.42	0.7249
D E C	4	6.06	0.4666
D E A C	12	16.54	1.2731
D E B C	4	7.30	0.5620
D E A B C	12	8.77	0.6750
Error 5	704	12.99	
D F	2	20.16	1.9557
D F A	6	13.34	1.2939
D F B	2	0.32	0.0310
D F A B	6	6.80	0.6600
D F C	2	3.28	0.3179
D F A C	6	4.50	0.4361
D F B C	2	71.93	6.9770*
D F A B C	6	9.83	0.9539
Error 6	352	10.31	

Continued...

Source	df	MS	F
E F	2	24.29	2.6937
E F A	6	7.90	0.8755
E F B	2	0.27	0.0305
E F A B	6	5.25	0.5818
E F C	2	11.68	1.2955
E F A C	6	10.32	1.1447
E F B C	2	0.76	0.0840
E F A B C	6	10.83	1.2010
Error 7	352	9.02	
D E F	4	11.88	1.2485
D E F A	12	7.58	0.7970
D E F B	4	6.89	0.7243
D E F A B	12	11.46	1.2052
D E F C	4	13.13	1.3799
D E F A C	12	18.93	1.9904
D E F B C	4	15.70	1.6505
D E F A B C	12	7.68	0.8069
Error 8	704	9.51	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 10

Backster Test Pupil Change ANOVA

Source	df	MS	F
A (Effectiveness	3	542.78	4.0606*
B (Shock)	1	17.68	0.1323
A B	3	132.12	0.9884
C (Guilt)	1	65.75	0.4919
A C	3	164.40	1.2299
B C	1	0.06	0.0005
A B C	3	118.23	0.8845
Error 1	176	133.67	
D (Trials)	2	352.65	9.6324*
D A	6	47.72	1.3033
D B	2	24.60	0.6720
D A B	6	13.06	0.3568
D C	2	59.92	1.6367
D A C	6	37.42	1.0221
D B C	2	2.23	0.0608
D A B C	6	62.03	1.6943
Error 2	352	36.61	
E (Question Set)	2	217.17	6.2987*
E A	6	47.26	1.3707
E B	2	49.08	1.4236
E A B	6	27.43	0.7955
E C	2	28.15	0.8165
E A C	6	22.22	0.6444
E B C	2	9.13	0.2647
E A B C	6	34.11	0.9893
Error 3	352	34.48	

Continued...

Source	df	MS	F
F (Question Type)	1	185.63	4.1836
F A	3	31.29	0.7053
F B	1	25.71	0.5793
F A B	3	45.87	1.0338
F C	1	31.96	0.7204
F A C	3	15.59	0.3514
F B C	1	7.27	0.1638
F A B C	3	12.33	0.2779
Error 4	176	44.37	
D E	4	228.51	7.4839*
D E A	12	50.41	1.6508
D E B	4	44.93	1.4713
D E A B	12	34.11	1.1170
D E C	4	88.05	2.8837
D E B C	4	62.44	2.0449
D E A B C	12	30.53	0.9498
Error 5	704	3.53	
D F	2	22.05	0.5268
D F A	6	33.62	0.8033
D F B	2	36.88	0.8813
D F A B	6	89.09	2.1287
D F C	2	78.06	1.8652
D F A C	6	16.06	0.3838
D F B C	2	51.90	1.2401
D F A B C	6	56.73	1.3556
Error 6	352	41.85	

Continued...

Source	df	MS	F
E F	2	34.64	0.7958
E F A	6	131.06	3.0112
E F B	2	31.91	0.7332
E F A B	6	35.12	0.8069
E F C	2	0.05	
E F A C	6	34.77	0.7988
E F B C	2	17.92	0.4117
E F A B C	6	78.37	1.8005
Error 7	352	43.53	
D E F	4	106.39	2.6282
D E F A	12	22.86	0.5647
D E F B	4	161.89	3.9994
D E F A B	12	39.48	0.9754
D E F C	4	99.97	2.4698
D E F A C	12	62.06	1.5331
D E F B C	4	59.54	1.4709
D E F A B C	12	28.60	0.7065
Error 8	704	40.48	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 11

Backster Test Heart Change ANOVA

Source	df	MS	F
A(Effectiveness)	3	714.49	1.6109
B (Shock)	1	3.72	0.0084
A B	3	1255.67	2.8310
C (Guilt)	1	183.74	0.4143
A C	3	609.24	1.3736
B C	1	112.52	0.2437
A B C	3	328.55	0.7407
Error 1	176	443.55	
D (Trials)	2	150.22	1.5093
D A	6	102.55	1.0304
D B	2	34.07	0.3424
D A B	6	178 41	1.7925
D C	2	128.02	1.2862
D A C	6	50.32	0.5056
D B C	2	155.41	1.5615
D A B C	6	137.67	1.3832
Error 2	352	99.53	
E (Question Set)	2	129.05	2.0719
E A	6	94.00	1.5090
E B	2	74.26	1.1923
E A B	6	89.54	1.4374
E C	2	55.22	0.8865
E A C	6	38.26	0.6143
E B C	2	13.25	0.2127
E A B C	6	34.11	0.5477
Error 3	352	62.29	

Continued...

Source	df	MS	F
F (Question Type)	1	4.65	0.0774
F A	3	27.63	0.4601
F B	1	180.47	3.0054
F A B	3	38.95	0.6487
F C	1	638.38	10.6314*
F A C	3	8.67	0.1444
F B C	1	14.39	0.2396
F A B C	3	15.23	0.2536
Error 4	176	60.05	
D E	4	58.80	1.2247
D E A	12	35.21	0.7334
D E B	4	204.78	4.2654*
D E A B	12	48.27	1.0054
D E C	4	42.72	0.8899
D E A C	12	48.10	1.0019
D E B C	4	96.36	2.0071
D E A B C	12	76.79	1.5995
Error 5	704	48.01	
D F	2	42.69	0.8162
D F A	6	60.99	1.1659
D F B	2	52.17	0.9974
D F A B	6	53.44	1.0217
D F C	2	17.53	0.3351
D F A C	6	37.61	0.7190
D F B C	2	141.36	2.7024
D F A B C	6	20.68	0.3953
Error 6	352	52.31	

Continued...

Source	df	MS	F
E F	2	608.29	9.9597*
E F A	6	94.61	1.5491
E F B	2	3.07	0.0503
E F A B	6	59.76	0.9785
E F C	2	206.14	3.3752
E F A C	6	79.47	1.3012
E F B C	2	94.20	1.5423
E F A B C	6	203.14	3.3261
Error 7	352	61.07	
D E F	4	35.57	0.7376
D E F A	12	44.44	0.9217
D E F B	4	80.90	1.6777
D E F A B	12	97.70	2.0260
D E F C	4	17.01	0.3527
D E F A C	12	20.65	0.4282
D E F B C	4	20.31	0.4212
D E F A B C	12	46.55	0.9653
Error 8	704	48.22	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 12

GKT Pupil Size

Source	df	MS	F
A (Effectiveness)	3	3031.37	2.2355
B (Shock)	1	1047.09	0.7791
AB	3	223.64	0.1664
C (Guilt)	1	2818.73	2.0973
A C	3	1142.39	0.8500
B C	1	49.79	0.0370
A B C	3	1044.60	0.7772
Error 1	176	1344.98	
D (Trials)	3	8357.02	121.9867*
D A	9	92.01	1.3430
D B	3	57.05	0.8400
D A B	9	58.27	0.8506
D C	3	17.78	0.2595
D A C	9	43.94	0.6415
D B C	3	26.18	0.3821
D A B C	9	39.48	0.5762
Error 2	528	68.51	
E (Question Type)	1	395.20	20.0626*
E A	3	38.93	1.9762
E B	1	12.98	0.6591
E A B	3	10.73	0.5446
E C	1	225.42	11.4434*
E A C	3	9.55	0.4849
E B C	1	12.25	0.6218
E A B C	3	41.12	2.0873
Error 3	176	19.70	

Continued...

Source	df	MS	F
E D	3	22.96	1.1614
E D A	9	8.23	0.4162
E D B	3	30.58	1.5471
E D A B	9	28.20	1.4266
E D C	3	26.34	1.3325
E D A C	9	11.45	0.5793
E D B C	3	20.86	1.0551
E D A B C	9	7.07	0.3577
Error 4	528	19.77	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 13

GKT Heart Rate

Source	df	MS	F
A (Effectiveness)	3	563.97	0.5526
B (Shock)	1	1918.43	1.8797
A B	3	627.62	0.6149
C (Guilt)	1	155.93	0.1528
A C	3	579.84	0.5681
B C	1	167.59	0.1642
A B C	3	794.93	0.7789
Error 1	176	1020.63	
D (Trials)	3	88.02	5.1029*
D A	9	39.72	2.3028
D B	3	0.31	0.0183
D A B	9	37.86	2.1951
D C	3	38.02	2.2044
D A C	9	26.38	1.5294
D B C	3	9.83	0.5698
D A B C	9	25.53	1.4802
Error 2	528	17.25	
E (Question Type)	1	60.16	5.9091*
E A	3	5.98	0.5871
E B	1	62.61	6.1495*
E A B	3	7.10	0.6979
E C	1	60.25	5.9185*
E A C	3	7.67	0.7534
E B C	1	1.20	0.1175
E A B C	3	3.05	0.2997
Error 3	176	10.18	

Continued...

Source	df	MS	F
E D	3	5.72	0.6527
E D A	9	7.69	0.8777
E D B	3	10.65	1.2152
E D A B	9	6.15	0.7015
E D C	3	8.50	0.9698
E D A C	9	13.98	1.5949
E D A B C	9	14.95	1.7076
Error 4	528	8.77	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 14

GKT Pupil Change ANOVA

Source	df	MS	F
A (Effectiveness)	3	366.84	5.2955*
B (Shock)	1	0.87	0.0125
A B	3	34.94	0.5044
C (Guilt)	1	403.33	5.8222*
A C	3	15.75	0.2273
B C	1	11.63	0.1678
A B C	3	63.28	0.9135
Error 1	176	69.27	
D (Trials)	3	202.22	9.4442*
D A	9	18.46	0.8619
D B	3	7.30	0.3410
D A B	9	16.28	0.7603
D C	3	21.33	0.9961
D A C	9	30.06	1.4038
D B C	3	4.30	0.2009
D A B C	9	23.62	1.1031
Error 2	528	21.41	
E (Question Type)	1	31.74	1.2907
E A	3	73.95	3.0072
E B	1	20.29	0.8251
E A B	3	28.97	1.1780
E C	1	118.55	4.8213*
E A C	3	17.00	0.6912
E B C	1	47.54	1.9334
E A B C	3	64.76	2.6336
Error 3	176	24.59	

Continued...

Source	df	MS	F
D E	3	10.42	0.4198
D E C	9	47.10	1.8976
D E B	3	18.13	0.7304
D E A B	9	11.80	0.4753
D E C	3	48.04	1.9354
D E A C	9	25.50	1.0275
D E B C	3	7.54	0.3039
D E A B C	9	24.10	0.9710
Error 4	528	24.82	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).

Table 15

GKT Heart Change ANOVA

Source	df	MS	F
A (Effectiveness)	3	477.88	1.8322
B (Shock)	1	109.26	0.4189
A B	3	745.19	2.8571*
C (Guilt)	1	134.01	0.5138
A C	3	329.29	1.2625
B C	1	700.18	2.6846
A B C	3	433.30	1.6613
Error 1	176	260.82	
D (Trials)	3	164.72	3.8533
D A	9	29.84	0.6980
D B	3	45.97	1.0753
D A B	9	83.45	1.9520
D C	3	189.92	4.4428*
D A C	9	52.22	1.2216
D B C	3	75.25	1.7602
D A B C	9	71.40	1.6702
Error 2	528	42.75	
E (Question Type)	1	156.65	5.2956*
E A	3	24.32	0.8221
E B	1	0.44	0.0137
E A B	3	12.76	0.4313
E C	1	73.98	2.5008
E A C	3	12.01	0.4060
E B C	1	15.45	0.5224
E A B C	3	10.81	0.3654
Error 3	176	29.58	

Continued...

Source	df	MS	F
E D	3	22.53	0.7193
E D A	9	13.85	0.4422
E D B	3	32.09	1.0245
E D A B	9	50.63	1.6164
E D C	3	42.62	1.3605
E D A C	9	47.61	1.5200
E D B C	3	17.20	0.5491
E D A B C	9	39.62	1.2649
Error 4	528	31.32	

*$p < .05$ df corrected (Geisser Greenhouse, 1958).