THE PUPILLARY RESPONSE AND THE DETECTION OF DECEPTION

BY

Michael T. Bradley

A Thesis Presented to the Faculty of Graduate Studies

University of Manitoba

In Partial Fulfillment of the Requirements for the

Degree of Master of Arts

Department of Psychology

April, 1976

"THE PUPILLARY RESPONSE AND THE DETECTION

OF DECEPTION"

by

MICHAEL T. BRADLEY

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

MASTER OF ARTS

ACKNOWLEDGEMENTS

I was especialy fortunate in having had the direction of Doctor Michel Pierre Janisse. His professionalism, enthusiasm and his dedication to knowledge have made him a sincere teacher and a conscientious scientist. I wish to thank Doctor Janisse for being a humanitarian as well as a scientist, for having patience and understanding and for his encouragement when it was most needed.

Similarly I would like to thank Doctors M. Singer and J. Whiteley for their helpful advice coupled with their intelligent and thoughtful criticisms. Further I would like to thank Doctor B. Sneideman for the donation of his time and energies and for his expression of interest in this thesis. I am indebted to M. Shooter for his technical expertise and knowledge which proved indispensable.

Finally I would like to thank my loving wife Yvonne and my young son Michael, Yvonne for her loyalty, devotion and support expressed in so many ways, Michael for his love and for the time we could not spend together.

Abstract

The pupillary response was measured in a "lie detection" or detection of deception paradigm in an attempt to: a) determine the utility of this measure, and b) to assess some theoretical assumptions about the detection of deception paradigm. All subjects selected a number from one to five. They were then asked, in random order, questions of the nature, "Is it one?", "Is it two?"... (etc.) ... Depending on which group, they either answered "No" verbally to all five questions (overt), remained silent but said "No" to themselves in response to all questions (covert), or simply listened to each question (control). In Experiment 1 no differences were found between the groups but pupillary responses to the number selected (critical stimulus) were larger than the average of responses to noncritical stimuli over groups and significant detection rates were found in each group. In Experiment Two these results were replicated. In addition two blocks of five trials were presented to each subject. Differential responding to the critical question was evidenced over the first block but not the last block in all three groups. No differences emerged between the groups on habituation trials. The fact that the control group evidenced differential responding to critical stimuli suggested that a "lie" was not a necessary event in the detection of deception paradigm. It was concluded that " short term attention" is a sufficient condition to evoke

differential responding in this paradigm. It was also concluded that detection was most likely to occur on early trials, rather than later ones.

Table of Contents

List of Tables

List of Figures

Figure

page

Responses of physiolcgical 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 tc control when a subject is exposed to meaningful stimuli and to the degree that this is true represent objective physiological manifestations of on going 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 pclygraph. 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. Over all, 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. Further work is also 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 perceptable physiological or behavioural alterations has been prevalent. This is especially true with regard to the detection of deception or "lie detection" (Bering, 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. Witchdoctors sometimes leapt at suspects smelling them feverishly. Distinctive odours indicated guilt and it is possible that the fear of being caught produced such an odour. 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 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 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 behavioural and physiological differences accompanying deception began around the turn of the century. In 1906 Jung (see Orne, Thackray and Paskewitz, 1972) studied differential reaction times to stimuli on which subjects hoped to deceive the experimenter. Inbau and Reid (1953) report that Ceasare Lombroso used a "hydrosphygmograph" to record blood pressure changes during interogation. 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. These changes were recorded on a revolving drum. 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 and finally Keeler (1930) developed the polygraph. His device measured respiration, relative blood pressure and the GSR. These have remained the major physiological measures in "lie detection" since that time (Davis, 1961).

Field Work

Field work or the practical application of lie detection in criminal investigations has dominated much of the work in the detection of deception. Orne, Thackray and Paskewitz (1972) mention that little systematic scientific investigation has been done. 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, there have been few reports that deal with attempts to validate findings. Field work is characteristically more of an art than a science. Inbau, Moenssen and Vitullo (1972) stress 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 recording of autonomic responses on the polygraph represents only part of a structured 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 the scoring of responses has not been specified in quantitative terms. Inbau, Moenssens 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 much 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).

Questioning Techniques in the Field

Practioners in criminal lie detection have develcped and come to rely on certain techniques and procedures to present questions during an interrogation.

The guilty person technique The guilty person technique (Lykken, 1960) or "undisguised question method" (Burack, 1955) is a modification of a direct confrontation questions 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. Reid has formulated this procedure, labelled the Reid Control Questioning Technique, such that four relevant questions are asked along with four irrelevant questions (Reid and Inbau, 1966). Unfortunately, questioning a suspect about whom he may have killed can yield responses interpretable as deception even though the suspect is innocent (Burack, 1955). Orne, Thackray and Paskewitz (1972) have called for more investigation into what is termed "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. These could be questions such as "did you ever steal anything else"? The particular questions are worked out for each individual suspect by the interrogator in a pre-lie detection test interview to insure a lie response. If the lie reaction to the irrelevant question is the same as or greater than the response to a question relevant to the crime then the suspect is considered innocent. An additional 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.

Lee (1953) has added "secondary relevant questions" which concern themselves with details only a guilty suspect could know. 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 nonthreatening and thus would not be expected to create strong reactions.

The above techniques have been based on the assumption that a "lie response" is being measured rather than an emotional reaction to the content and implications of such questions. But, the literature does not support the conception of a "lie response" (Kugelmass, Lieblich, and Bergman, 1967). Day (1972) states 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. 755).

The disguised question technique Another technique, actually anticipated by Lee (1953) 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. One variation of the technique presents stimuli in serial order and the interrogator looks for a gradual rise (or fall) of baseline readings which reach a peak at the item. Thus a series of questions could be asked all in the form of "did you steal $100.00? " and continue through to the actual amount stolen. Day (1972) points out that the disguised question or guilty knowledge technique has a fairly solid rationale in that there may be some involuntary physiological response to remembered details of a crime.

Measures used in the Field

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 (Inbau and Reid, 1966).

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. 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 of the 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 is attributed to 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.

<u>Cardiovascular responses</u> 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) criticize the GSR because they are 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.

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 nonsystematic sources of variance. Previously it was mentioned how various factors relegate the field interrogation procedures to an art. In spite of this, the detection of deception situation or paradigm is appropriate for laboratory investigation.

Questioning Techniques in Laboratory Research

Orne, Thackray and Paskewitz (1972) have discussed the major experimental designs used in the detection of deception paradigm. These involve the guilty information technique, the guilty person technique and the mock crime situation.

The guilty information technique The guilty information situation 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 fairly common (Alpert, Kurtzberg, and Friedhaff, 1963; Block, 1957;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).

The guilty person technique The guilty information situation is contrasted with the guilty person procedure in that the experimenter attempts to discriminate between a guilty or innocent person, which more closely resembles the situation in the field. Gustafson and Orne (1964) found that it was easier to detect subjects who thought the experiment was an attempt to detect guilty subjects among the innocent than to detect information (the number or card) that the subject was attempting to conceal.

<u>The mock crime situation</u> The mock crime situation has been used extensively in laboratory investigations, (for example, Berrien, 1942; Berrien and Huntington, 1943; Lykken, 1959). The attempt has been to achieve realism and approximate a field situation, with the subject usually being an observer or participant in a "crime". Here again the subject can either view himself as having "guilty information" or as being a "guilty person" (Berrien and Huntington, 1943).

Measures used in Laboratory Research

Measures used in laboratory detection of deception situations have not been limited to those popular in the field. A much wider array of measures both autonomic and nonautonomic have been used. Cutrow, Parks, Lucas and Thomas, (1972) have used eye blink rate, 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 (Berkhout, Walter and Adey, 1970). Oberman (1939) has used the electroencephalogram and pupillary dilation was investigated by Berrien and Huntington, (1943).

Cardiovascular responses 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 (1968) have not been able to detect deception using blood pressure measures. Orne, Thackray and Paskewitz (1972) have suggested that since the 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.

Other cardiovascular measures have not been very satisfactory in discriminating deception responses in the laboratory. Davis (1961), based on his preliminary studies, had predicted that a pulse rate change index would be moderately successful but Kugelmass and Lieblich (1966) found that it gave detection results no better than chance. Ellson, Davis, Saltzman and Burke (1952) cited in Orne, Thackray and Paskewitz (1972) found heart rate to be a more successful discriminator than pulse volume but both these measures were inferior to the GSR and systolic blood pressure. Kugelmass and Lieblich (1966) and Kugelmass, Lieblich, Ben-Ishai, Opatowski and Kaplan (1968) have found that heart rate discriminates no better than chance. Cutrow, Parks, Lucas and Thomas (1972) used several measures and although heart rate and finger pulse volume discriminated deception better than chance, with finger pulse volume superior to heart rate, the measures were considerably poorer than the GSR.

<u>Respiration</u> 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. 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, quantifiable features of respiration such as breathing rate (cycles per second) and amplitude have been compared in the laboratory. Cutrow, Parks, Lucas and Thomas (1972) found that these measures discriminated deception better than chance when the criteria for a lie response was the smallest breath amplitude and the longest time for three inhalation - exhalation cycles.

<u>The galvanic skin response</u> In contrast to the GSR 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). Laboratory

investigators (Cutrow, Parks, Lucas and Thomas, 1972; Thackray and Orne, 1968) who have used multiple ANS indicies agree that the GSR is superior to other variables in the detection of deception. Orne, Thackray and Paskewitz (1972) report that field investigators think the GSR 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. Some investigators have carried out laboratory studies where an attempt was made to create strong emotional or motivational involvement, for example Kubis (Orne, Thackray, and Paskewitz, 1972) employed mock crimes, Violante and Ross (1964) used aversive noise, Kugelmass and Lieblich (1966) used policemen who were told that their career future may depend on the experimental results. Even with these arousing manipulations the GSR was still found to be highly effective.

<u>Ocular responses</u> 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. The pupillary response of a slow negatively accelerated dilation, lasting 1-5 seconds, followed by a rapid constriction yielded correct discrimination in 70% of the cases. Generally the pupillary

response in lie detection has been neglected because of measurement problems. However, Orne, Thackray and Paskewitz (1972) predict that more recently developed television pupillometers will facilitate further research with this measure.

In summary there is much need of improvement in measures used in the detection of deception. The GSR, the most accurate measure in the laboratory situation does not seem to perform as well in the field. Cardiovascular measures except for blood pressure (Davis,1961) have not been very accurate in detecting deception. 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. Work on the pupillary response in lie detection has been delayed because of the equipment involved and the lack of instant feedback on the response (developing film and measuring) (Day, 1972).

Laboratory Research and Field Work Compared

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

The degree to which the laboratory and field situations may differ is considerable and these differences may have great effect on the results of investigations. One 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). One laboratory study, using college students who were told that only those with superior intelligence and emotional control could avoid detection (Gustafson and Orne, 1966) supports this contention, but neither Day (1972) nor Lieblich, Naftali, Shmueli and Kugelmass (1974), in their laboratory investigations have found any effect due
to varying levels of motivation. Lieblich, Naftali, Shmueli and Kugelmass (1974) used the same manipulation to increase motivation as Gustafson and Orne (1963). 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. 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 elimimation 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 occured in relation to instrumentation used. Laboratory investigators have had much success with the GSR (eg., Kubis, 1962), whereas some field investigators have found the GSR to be a poor discriminator (Reid and Inbau, 1966). 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.

<u>Theoretical bases for physiological responses in deception.</u>

Several theories have been advanced to explain why 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 is based on the threat of punishment and 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 physiological responses then explanations based on fear of punishment or motivations to deceive the experimenter are not necessary. Day and Rourke (1974) included conditions where 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.

This result fits well with data suggesting that differential physiological responses in detection experiments are not contingent upon overt or 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 pays attention to a given stimulus but how much attention 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.

Habituation of responses during deception.

One method of increasing the probabilty of the correct detection of deception is to increase the number of trials available to compare critical stimuli to control stimuli (Lieblich, Naftali, Shmueli and Kugelmass, 1974). The simplest way to accomplish this is to repeat the critical and control questions several times. However, since ANS responses habituate with stimulus repetition the utility of this approach depends on the rates of habituation of responses to the control and critical stimuli. Differences between the critical and neutral stimuli may only occur on the first few trials because the response to the critical stimulus may habituate rapidly. A confusing situation could arise if the response to the critical stimulus habituated more rapidly than responses to one or some of the control stimuli. Little or nothing would be gained by extensive repetition of the critical stimulus if the above held true.

Differential habituation to stimuli Ellson, Davis, Saltzman and Burke (1952) cited in Orne, Thackray and Paskewitz (1972) found that the rate of detecting critical responses declined over trials. Ten subjects were given two series of five trials with each trial consisting of the names of six months, one of which the subject had selected previously. Detection of the selected month names went from 80% in the first series of trials to 70% in the second series. On the other hand Thackray and Orne (1967) found that detection became possible only after many presentations of the stimuli. This experiment was notable for a great deal of initial responsivity to control stimuli, which then seemed to habituate more rapidly than responsivity to critical stimuli. Peterson and Jung (1907) also report that responses to control stimuli habituated more rapidly than responses to critical stimuli over three trials. Jones and Wechsler (1968) using a list of emotional and neutral words found response habituation to neutral words but not to emotional words as the end of the list was neared.

Geldreich (1941) habituated responses to control stimuli while minimizing or preventing habituation to the critical stimulus. He presented a set of five cards to one group of subjects requesting that they chose one. Then he asked a single series of questions about which card they had chosen. For this group he detected 74% of the chosen cards correctly. The second group was also requested to pick a card but they were asked a series of twenty to fifty questions about cards, none of which included the original five cards from which they had chosen. After this they were asked about the five cards. The GSR's to non-critical questions were substantially reduced while the GSR's to the critical stimulus remained large. The detection rate for this group was 100%. This approach was effective for increasing detection and may hold promise as a methodolgy in future.

Another attempt at improving detection rates involved the combination of trials comparing the critical and control stimuli. If the response to the critical stimulus is slow to habituate, then combining trials should lead to more accurate detection. Gustafson and Orne in an unpublished study, cited in Orne, Thackray, and Paskewitz (1972), combined trials one and two of a simple card test that had been repeated five times . This lead to an increased accuracy over the first trial alone. However, adding the results of the third trial did not augment detection and combining all five trials decreased the detection rate from that of the first trial. Lieblich, Naftali, Shmueli, and Kugelmass (1974) demonstrated a marked improvement in detection rates by repeating the stimuli and combining all the trials. The response to the critical stimulus did habituate over trials such that in the first trial it was largest 60% of the time but declined by the tenth to where it was largest 48% of the time in the low motivation group. In the high motivation group the critical response went from being largest in 55% to 50% of the trials. Combining all trials yielded detection rates of 96% in the low motivation group and 93% in the high motivation group. Repetition aided in this instance but there is a very important element to be noted . The subject's name was used as the critical stimulus and responses to a person's own name do not readily extinguish (Lynn, 1968).

Ben Shakhar, Lieblich, and Kugelmass (1970) used an approach which involved the presentation of many critical items. With this method the subject attempts deception in response to different specific questions and therefore cannot habituate to a specific repeated stimulus but has to adapt to repeated lying about many different questions . The experimenters reported a detection rate of 77% using this method. It is impossible to conclude from the report of Ben Shakhar, Lieblich and Kugelmass (1970) at what rate responses may habituate but since detection rates remain high habituation should be studied in that situation.

ANS measures and habituation Habituation may occur at greater or lesser rates depending on the measures used (Davis, 1961). All of the studies on habituation cited in the section above have used the GSR, which appears to take the longest of the commonly used measures to habituate. Solomon, Black, Watson, Huttenlocker, Turner, and Westcott (1958) cited in Orne, Thackray, and Paskewitz (1972) found that human heart rate, and respiration responses extinguish very rapidly in a conditioning paradigm whereas the GSR persisted over a long period of time. Seward and Seward (1934) cited in Orne , Thackray , and Paskewitz (1972) found that both body movement and respiration habituated more rapidly to shock than the GSR.

Studies designed to systematically investigate the habituation characteristics of different autonomic responses in various detection of deception situations would be a great asset in understanding lie detection phenomena.

The pupil response in the detection of deception

The present study will focus on the pupillary response. Berrien and Huntington (1943) have shown the pupil to be an indicant of deception but beyond that study the measure has been ignored in the detection of deception literature.

The pupillary response has not been neglected in other areas of psychology. Many researchers (e.g., Hess, 1972) have contended that there is an intimate relation between the activities of the pupil and brain activities. The number of possible influences on the pupil from various brain areas has made this seem likely. Both sympathetic and parasympathetic divisions of the ANS innervate the pupillary response with the sympathetic division controlling the dilator muscles of the iris and the parasympathetic division controlling the sphincter muscles of the iris (Lowenstein and Loewenfield, 1962). In addition, hypothalmo-thalmo-cortical activity effects pupil size. Dilation of the pupil can be evoked by stimulation of cortical areas around the orbital cortex, temporal tip, cingulate gyrus, insula, rhinal fissure and hippocampus in addition to the basal telencaphalon, hypothalmus, septum and thalamic nuclei (Delgado,1966). Hakerem (1970) suggests that the pupil is such a precise indicant of cerebral function that he refers to it as as "a permanently implanted electrode in man" (1970 p.59).

Goldwater (1972), Hess (1972), Janisse (1973) and Janisse and Peaver (1974) have reviewed literature showing the utility of the pupillary response as a dependent variable when investigating phenomena such as mental effort, arousal or emotion. Hess had hoped to demonstrate that the pupil responded by constricting to aversive stimuli, however, Janisse (1973) questioned this hypothesis in a review of the literature and found that any constriction effects were probably due to inadequate controls when using visual stimuli. Janisse (1974) did find that the greater the intensity of a stimulus whether positive or negative the greater the dilation response. Thus, the pupillary response is unidirectional no matter what the valence of the stimulus and it is intensity that determines the magnitude of the response.

Voluntary Control of the Pupil

An important question especially in the detection of deception concerns voluntary control of ANS response systems. Prather and Berry (1973) demonstrated that the pupillary response can be shaped by instrumental conditioning. If ANS responses were easily controlled a subject could avoid detection. Bumke (1911) cited by Hess (1972) considered that the pupil can be influenced indirectly by changing visual accomodation from near to far, holding one's breath, exerting one's muscles, self infliction of pain or performing mental calculation, but that it could not be controlled directly. Not only does the performance of such tasks result in pupillary dilation but Simpson and Paivio (1968) have shown that this dilation can be enhanced when an overt verbal or motor response is required. They used a cognitive imaging task. Thus the pupillary response is an involuntary concomitant of the above mentioned physical and cognitive acts and can be additionally influenced during a cognitive task by the performance of an overt response. When subjects were asked directly to inhibit dilation in response to stimuli they did not succeed in doing so (Krueger, 1967). Clark and Johnson (1970) informed, misinformed and did not inform different groups about the dilation and constriction effects expected as a result of mental effort. This manipulation did not result in differential pupillary responses during the tasks. Overall it appears that the pupillary response can be controlled indirectly, that is made to dilate as a result of engaging in some motor or cognitive activity. However suppression of dilation has not been found except in the one experiment by Prather and Berry (1973) where real time feedback of the pupillary activity was necessary for the subject to gain the control.

Habituation of the pupillary response.

An important topic for the research in this paper is the habituation of the pupillary response to psychosensory stimuli. Unfortunately there has been no work directly concerned with habituation of the pupillary response in a detection of deception paradigm. Reliance must be placed on data from habituation studies extinction in conditioning studies.

The pupillary response is a component of the generalized orientation reaction and is typically obtained with decreasing intensity for something like 10-15 stimulations (Lynn,1968). Clynes (1962), cited in Goldwater (1972), found that the pupillary response did not habituate to tones or clicks over several hundred trials. This lack of habituation may not be surprising since Goldwater (1972) noted that the later responses were generally small and he suggested that the method of evaluating habituation may give the impression of lack of habituation. Goldwater (1972) thought that a reanalysis of the data would reveal both diminution of response and spontaneous recovery within the blocks of trials evaluated by Clynes (1962). Nunally, Knot, Duchnowski, and Parker (1967) found that a differentially large pupil response habituated to novel stimuli. Within two trials the responses were of the same magnitude as responses to control stimuli.

Extinction data from experiments of Baker (1938), Cason (1922), and Hudgins (1932) suggested that the conditioned pupillary response did not extinguish. However, all these experiments attempted to condition the pupillary reflex using light as well as an unconditioned stimulus and it is now believed that the experimenters did not obtain successful conditioning since all replications have failed and the original studies were done with very crude instruments (Young,1958). Successful conditioning of pupillary dilation with shock (Harlow, 1940) and auditory stimulation (Kugelmass, Hakerem and Montgiaris, 1969) has been obtained. Goldwater (1972) states that the weight of evidence suggests that some type of motivational component is necessary for classical conditioning. Once conditioning had been obtained with shock, Gerall and Obrist (1962) found no tendency for the pupillary response to habituate over fourteen trials.

Experiment 1

The first experiment was designed to assess the effectiveness of the pupillary response for the detection of deception. The experiment included three groups. The first group attempted to "lie" verbally, members of the second group were told to "lie" silently to themselves. The third group also remained silent, but unlike the other groups made no attempt to lie. It was predicted that the magnitude of the pupillary response to questions concerning the critical item would be larger than responses to the noncritical questions in both the lie groups and control group. Detection rates for all groups including the control group were predicted to be greater than chance. This prediction was based on Day and Rourke's (1974) finding that even mere exposure to a stimulus is sufficient condition for subsequent detection. Any enhancement of detectability due to lying (Davis, 1961) would be evidenced on a groups X question interaction. If overt responding (Simpson and Paivio, 1968) further increased the difference this would be evidenced on the same interaction. These hypotheses, that lying and overt responding enhance the pupil response, would be substantiated if post-hoc tests following a significant groups by questions interaction revealed significant differences between each group on the critical question with the means of group A the largest and group C the smallest.

Method

Subjects

Subjects were 60 male university students who took part in the experiment to fulfill an introductory psychology course requirement. Twenty served in each of the three conditions which were Overt, Covert and Control.

Apparatus

A Whittaker Space Sciences Eye View Monitor and Television Pupillometer was used. This instrument was designed to provide an accurate assessment of pupil diameter and record this on FM tape. In addition, eye movements within the range of 30° horizontal and 25° vertical can be monitored. The apparatus has two television cameras, one to monitor the pupil, the other to monitor the areas of the visual stimulus upon which the subject has focused.

The derived data was recorded in digital form on a Kennedy incremental magnetic tape recorder, Model 1600/360. Verbal stimuli were presented via a two channel Sony tape recorder. The onset of stimuli can be marked on the digit tape by pressing a connected hand button.

The apparatus was set up in an all white experimental room illuminated by three 100 watt bulbs which were placed directly above the subject and approximately 3 1/2 metres from the visual target. The bulbs were positioned to provide uniform illumination of the target area. Under these conditions the illuminance was 72 lux at the target surface and luminance, the amount of visually effective light from an extended source, was 25 nits at the pupillometer headrest. These bulbs were connected to a 25 ampere constant voltage transformer to provide a steady nonfluctuating power source to prevent change in illumination due to power surges in the external electrical supply.

The pupillometer cameras and a head rest for subjects were positioned on a moveable tray mounted on a swivel stand by the subject's chair. The chair was adjustable so it could be raised or lowered. When the subject was seated, the experimenter could move the tray towards the subject so that a comfortable position could be obtained.

Numerical stimuli were stenciled on heavy white index cards and served as the stimuli in the experiment.

Procedure

Prior to the arrival of a subject at the laboratory the experimenter determined the order of conditions and numbers assigned to each condition using a random numbers table (Runyon and Haber, 1967). The restrictions were such that there were to be twenty subjects in each condition and that each of the five numbers from one to five served as the critical stimulus four times in each condition. Thus as each subject entered the laboratory he was assigned the next condition and number in a randomly determined sequence.

All subjects were informed that there would be a 10 minute wait while their pupils adapted to the room lighting. During this time they sat at a table beside the experimental chair and read their instructions.

All subjects, whether they were in the "lie" groups or the control group read that they were participating in an experiment concerning physiological responses used in lie detection. The instructions said that the pupillometer measures changes in the size of the pupil and by these changes the experimenter can tell if the subject is lying. Subjects in the first group, the "Overt lie" condition read that they were to answer "no" to all questions in an attempt to conceal the number they had chosen. The second group, the "Covert lie" group read that they were to remain silent during the questioning but to "think no" in response to the questions, in an attempt to conceal the chosen number. The third group, the Control group, were told to remain silent and just listen to the questions since they were not attempting to conceal the number.

Once seated in the pupillometer chair subjects picked one card from a set of five, numbered individually from one to five. The choice of the card was made while the experimenter held them face down so that neither the subject nor the experimenter could see the number chosen. The experimenter insured that each subject chose the correct number for his particular condition by having five decks of five cards with each of the cards in a deck having the same number. Since the experimenter held the deck in such a way as to conceal the numbers from the subject they did not suspect this ruse.

After choosing the card the subjects memorized the number. The Overt and Covert groups had previously read that they were to conceal the number whereas the control group had been instructed not to hide it.

Next the subjects placed their heads on the head rest and focused on the middle of a target positioned on the wall 3.5 metres in front of them. The experimenter then made adjustments to the chair and the head rest to insure their comfort, and then adjusted the pupillometer. The subjects were informed that there would be a fifteen second silent period from the time a tape recording of the questions was turned on to the first question and thereafter a seven second interval between each question. When the questioning was finished subjects would be told to rest. The order of presentation of the five tape recorded questions had been randomly determined before the experiment. The tape questions were of the form "Is it one?", "Is it two?", etc.

As a final instruction all subjects were told not to let their focus of vision wander off the target, keep their eyes open wide and try not to blink.

Analysis

A critical response was defined as the pupillary response occurring after a subject was asked about the number he had actually chosen. A noncritical pupillary response followed questions about any other number in the questioning series.

The basic unit of analysis was the per second average of pupil size calculated from pupil measurements recorded every 1/60 of a second. There were 7 seconds following each question. The first 4 seconds were designated the response period. The remaining 3 seconds comprised the baseline period. The 3 seconds following a response on one question were sequentially prior to the next question and as such served as a prequestion baseline for that following question. Prior to the first question the last 3 seconds of the initial baseline were used as the first pre-question baseline. The distinction between the 4 second response period and the 3 second baseline was primarily for the detection rate analysis as described below.

The data analysis was carried out in two ways. The first used the basic unit, actual pupil size, in analysis of variance procedures. Since the experimental design involved repeated measures and it was not known if the equality of variance-covariance matricies assumption would be violated conservative degrees of freedom, adjusted by the Geisser-Greenhouse (1958) technique, were used. The second part of the analysis involved tabulation of the number of times that critical "peak scores", "mean scores", and "difference scores" exceeded all noncritical scores of their respective type in an appropriate set. "Peak scores" used the largest, or peak, of the 4 actual pupil size values (one per second) in a response period. "Mean scores", involved the average of the four actual pupil size points in a response period to form one value for that period. "Difference scores" were calculated from the mean of the response period subtracted from the mean of the appropriate pre and post baseline periods. Successful detection was scored if the magnitude of the critical response was larger than all of the four noncritical responses in a set of questions.

Results

The means of actual pupil size during the three seconds of initial baseline prior to any questioning were 4.57, 4.66, and 4.64 millimetres respectively for Overt, Covert and Control. These values were not different from each other $F < 1.0$.

Figure 1 shows actual pupil size data collapsed over the 20 subjects in each group for the critical and noncritical questions across the seven second post question period.

Insert Figure 1 about here.

An analysis of variance on this data showed that the pupil size following a critical question was found to be greater than the average of pupil sizes following noncritical questions $\underline{F}(1,57) = 19.2$ $\underline{p} < .01$. Differences in the 7 consecutive values in the post response period were found $\underline{F}(1,57) = 36.1$, $\underline{p} < .01$. No significant effects were found for the groups by question interaction $\underline{F} < 1.2$. (A summary table of the above analysis is presented in Appendix A.)

The pupil size values, with a mean of 4.92 millimetres, in the four second response part of the post question period were found to be larger than the three values, with a mean of 4.81 millimetres, in the three second baseline when Scheffe's post hoc technique was used $\underline{F}(1,57) = 38.7$ $\underline{p} < .05$ (A summary table of the above analysis is contained in Appendix B).

Frequency

The number of times the pupillary response following the critical questions exceeded all responses in the appropriate set following noncritical questions was counted. The binominal test was used to detect deviation greater than the expected chance proportion of .2 or 4 successful detections out of 20. Any number of successful detections equal to or more than 8 was greater than chance beyond the .035 level. Table 1 shows detection results using analysis of mean scores, peak scores and difference scores for each of the three groups.

Insert Table 1 about here.

In Table 1 it can be seen that detection rates were significant using mean scores in groups B and C, for peak scores in group A and C, but were not significant for any group using difference scores. An overall χ^2 using the average detection rate for all measures failed to reveal superiority of any one measure.

Discussion

There are several points to note in this experiment. Overall at the group level the pupillary response was differentially sensitive to numbers brought to the attention of the subject prior to questioning. The experiment does not offer support that the lie detection paradigm enhances detectability since the detection rate data fails to reveal differences between the control and lie groups. This conclusion is supported by the parametric analysis in which the group by question interaction does not obtain significance. A hypothesis proposed by Day and Rourke (1974) that "short term familiarity" is a sufficient condition for differential responses seems more appropriate. At a more general level this finding implies that procedures designed to differentiate a stimulus from a set of stimuli will result in an augmented response to that stimulus. Whether detection rates using the pupillary response could be differentially enhanced for the lie groups with the addition of stressful or motivating factors remains an open question.

The lack of a significant group by question interaction precludes finding differential enhancement of the pupillary response due to overt responding. This is contrary to the findings of Gustafson and Orne (1965) who found subjects in the verbal response group easier to detect than subjects in the mute group. However Orne, Thackray and Paskewitz (1972) suggest that the role of verbal responding in lie detection remains unclear. Simpson and Paivio (1968) have concluded that overt responding whether verbal or motor, augments the pupillary response but their studies have concentrated on cognitive imaging tasks and not lie detection.

Although significant detection rates were found in each of the three groups this was only when all three methods of measuring a response were considered. No method was significantly better than another so there is no way to suggest which of the three measures would be most fruitful for further research.

Detection rates were low compared to those found in other laboratory experiments using comparable stimuli, numbers, letters or geometric designs and different physiological measures, predominantly the galvanic skin response. The detection rates reported in the literature ranged from 40% to 83% with a median value of 73% (Orne, Thackray and Paskewitz, 1972). In this experiment detection rates ranged from a low of 25% to a high of 50% with a mean rate of 36% (Table 1).

Comparison with other experiments may be misleading since no buffer item was used to attenuate the orienting reflex. The orienting reflex is a physiological response that may occur solely as a result of exposure to a novel stimulus regardless of the significance or lack of significance that stimulus has to the subject (Sokolov, 1963). This response to novelty may actually be larger than responses to critical and noncritical questions. The buffer item is usually in the form of an initial question that allows the subject to become accustomed to hearing the questions thus reducing the novelty.

Overall it may be concluded that measurement of the pupillary response is effective in the detection of numbers to which the subject has been pre-exposed. This provides added support for the attention hypothesis of Day and Rourke (1974) which states that mere pre-exposure to a number before questioning is an adequate condition for detection .

Experiment 2

Experiment 2 was designed both as a replication and extention of Experiment 1 and to study the habituation of the pupillary response over repeated trials. The effect of repeated presentations of stimuli remains an important topic in the detection of deception since the probability of successful detection may be raised by increasing the number of trials available to compare the critical and neutral responses.

Optimal detection would occur if habituation to the critical stimulus occurred less rapidly than to neutral stimuli. Repetition would be valuable even if habituation occurred at the same rate for critical and noncritical items if the general relationship of the critical response being larger than the mean of the noncritical responses held. Any one of the noncritical responses may be larger in a given trial but it would not be expected that any noncritical response would be systematically larger over several trials. If the critical response habituated at a more rapid rate than the noncritical response, repeated trials would only obscure detection. Neither lying nor overt responding emerged as important factors in Experiment 1. If either of these factors do have an enhancing effect that emerges over trials then a group by question by trials interaction would be expected.

Method

Subjects

Subjects were 60 male university students who took part in the experiment to fulfill an introductory psychology course requirement. They were not allowed to serve in this experiment if they had served in the first experiment in this report. 20 subjects served in each of the three groups Overt, Covert and Control.

Procedure

Procedural details concerning general instructions given to subjects, and the method of card choosing remain the same as those reported in Experiment 1. In addition to receiving the same instructions, as reported in Experiment 1, subjects were instructed that sets of questions would be repeated five times and that they were to remain in the head rest apparatus looking at the target until they heard a taped voice say "rest". The rest would be 30 seconds long. During this time subjects were to withdraw their heads from the head rest apparatus and look at the number on their chosen card. They were to continue to conceal the number from the experimenter. At the end of the rest period the experimenter would again set them up in the head rest apparatus ready for another five trials.

The tape recorded questions for this experiment, consisted of 10 sets of five questions about the numbers presented in random order. There was a 15 second baseline period before and after the first five sets of questions and the second five sets of questions. There was a 7 second period between each question, the initial 4 seconds of which comprised the response period and the following 3 seconds the baseline. This distinction was most important for the detection rate data.

No buffer item was presented in this experiment since the first trial was to serve as a replication of Experiment 1. After the first trial novelty effects should have worn off.

Results

Replication Data

The means of the actual pupil size during the three seconds of initial baseline before questioning were 4.88, 5.00 and 5.01 millimetres respectively for Overt, Covert and Control. These were not different from each other $F<1$.

A three factor analysis of variance was carried out on the pupil sizes measured during the first trial. These means collapsed across groups and subjects are presented in Table 2.

Insert Table 2 about here.

Differences between groups were not significant $F<1$. Pupil size following a critical question was larger than pupil size following a noncritical question $F(1,57) = 18.1$, $p < .01$. A significant difference between the values in the 7 consecutive seconds following a question was also found $F(1,57) = 48.7$, $p < .01$. As in Experiment 1 no significant groups by question interaction was found. (A summary table of this analysis is contained in Appendix C).

A post hoc analysis using Scheffe's technique on the four values of the respose period, with a mean of 5.16 millimetres, revealed that they were larger than the three values in the baseline period, with a mean of 4.98 millimetres, $F(1,57) = 55.4$, $p < .05$ (A summary of this analysis is contained in Appendix D).

As in Experiment 1 the number of times the pupillary response following the critical questions exceeded all responses in the appropriate set following noncritical questions was counted. Any number of successful detections equal to or more than 8 was greater than chance beyond the .035 level (See Table 3). Detection rates greater than chance were found only in the covert group and these occurred with each of the three methods of analysis.

Insert Table 3 about here.

An overall comparison of the detection rates in Experiment 1 and 2 showed them not to be significant $\chi^2 = .1$.

Habituation Data

This section of the results examines the change in pupil size over the course of ten repeated trials. Figure 2 presents this data collapsed over subjects and trials.

Figure 2 about here.

A five factor repeated measures analysis of variance carried out on this actual pupil size data revealed that there were no significant differences among the levels, overt, covert, and control of the only between factor, $F < 1$. Regarding the repeated factors, differences between the first and second block were found, $F(1,57) = 78.7$, $p < .01$, as were differences between trials within these blocks, $F(1,57) = 128.6$, $p < .01$. Responses to critical questions were greater than to noncritical questions $F(1,57) = 18.7$, $p < .01$. It can be seen from Figure 1 that pupil size declined over the seven second post question interval and this was significant $F(1,57) = 112.0$, $p < .01$. The predicted groups by trials by questions interaction was not found $F < 1$. (A summary table of this analysis is contained in Appendix E).

It was necessary to carry out simple main effects analysis where main effects were involved in significant interactions (Kirk, 1968). There were several such interactions, blocks x trials, blocks x questions, blocks x post question interval, trials x post question interval, group x post question interval, and finally question x post question interval. Simple main effects revealed that pupil size became smaller over successive trials in each of the two blocks, $F(1,57) = 107.5$, $p<.01$, $F(1,57) = 70.7$, $p<.01$.

The difference between critical and noncritical questions was found to be significant in the first block but not in the second. The respective values being, $F(1,57) = 22.1$, $p<.01$, $F(1,57) = 2.7$, $p<.11$.

The difference among pupil size values in the post question interval was significant for each block, $F(1,57) = 90.5$, $p<.01$, $F(1,57) = 50.5$ $p<.01$.

The post question interval difference held across each of the five trial levels collapsed across blocks (A summary of this analysis is presented in Appendix F).

In each of the groups actual pupil size differed through the post-response period, $F(1,57) = 66.2$ $p<.01$, $F(1,57) = 38.6$, $p<.01$, $F(1,57) = 16.4$, $p<.01$. Differences between critical and noncritical questions were found to hold only over the first four values of actual pupil size in the post question interval (A summary table of this analysis is presented in Appendix G).

Tukey's honestly significant difference test (Kirk, 1968) was used to examine the difference between critical and noncritical questions with each individual group and block. These differences ranged from 0.001 millimeters to 0.082 milimetres and are presented along with their significance values in Table 4.

Insert Table 4 about here.

Overall the differences in the first block of each group were significant but these differences failed to obtain significance in any of the groups in the second block.

Frequency Data

The number of times the magnitude of the pupillary response following the critical questions exceeded the magnitude of all pupillary responses following noncritical questions appropriate for a particular trial was counted for each trial. The binomial test was used to detect deviation greater than the expected chance proportion of .2 or 4 successful detections out of 20. Any number of successful detections equal to or more than 8 was greater than chance beyond the .035 level. Table 5 shows detection results for mean scores, peak scores and difference scores over the 10 trials in each of the three groups.

Insert Table 5 about here.

Even though the data was dichotomus in nature analyses of variance were carried out on the detection rates for means, peaks and difference scores. Lunney (1970) has shown the analysis of variance with dichotomous data to be robust if the proportion of the smaller response category is equal to or greater than .2 and there are at least 20 or more degrees of freedom for the error term

The analysis of variance for mean scores was carried out with the three groups as levels of the between factor, and trials 5 per block as the second within factor. No significant differences were found.

The same format analysis failed to yield significant differences for peak scores and difference scores. (Summary tables of these analysis are presented in Appendicies H, I and J respectively).

Looking at the total detection rates for each block it can be seen that only difference scores yielded significance for all three groups in the first block. Peak scores yielded significance only for the covert group and mean scores yielded significance for both the covert and control groups.

Table 6 presents detection rates derived by averaging the response magnitudes for the critical questions over each of a subject's five trials in a block and comparing them with the average of each of the noncritical numbers. If the mean of the critical questions was larger than the means appropriate to each of the noncritical questions then detection was successful. The binomial test was used revealing that any value over 8 was significant beyond the .0325 level.

Insert Table 6 about here.

Discussion

The first trial in Experiment 2 served as a replication of Experiment 1. The parametric analysis of the appropriate data revealed fundamentally the same results as found in Experiment 1: no differences in pupil size between the three groups; differences between critical and noncritical questions; differences between the designated response period and its subsequent baseline period; no differences between the groups on initial baselines. In addition, there were no differences in overall detection rates between Experiment 1 and the first trial of Experiment 2 for all groups, on all methods of scoring (peaks, means or difference scores) when an expected value derived from the mean of detection rates appropriate to the particular χ^2 analysis was carried out. Thus the findings of Experiment 1 have been successfully replicated. However the rearrangements of detection rates different from chance between groups from Experiment 1 and Experiment 2 suggests that for any one group and measure the obtaining of significant detection rates is not a highly reliable phenomena even though the underlying parametric data is replicable.

In Experiment 2, no differences were found between any of the three groups in detection rates over repeated trials. Again the findings by Gustafson and Orne (1965) that, lying overtly led to higher detection rates than lying covertly, were not supported. Day and Rourke's (1974) finding that mere pre-exposure to a number yields differential responding receives support from the detection rates during habituation.

In spite of the probable lack of relevance the numerical stimulus had to the subjects in this experiment, the overall detection results are very close to those reported by Shakhar, Lieblich and Kugelmass (1975) in their study of habituation. Using the GSR they observed a detection rate of 34%. In Experiment 2, a detection rate of 31% was found. Shakhar, Lieblich and Kugelmass (1975) used the subject's own name or the names of family and friends as the critical stimuli. These stimuli as Oswald , Taylor and Treisman (1960) have found are particularly relevant to the subject and continue to evoke physiological responses for many trials. Thus the slightly lower detection rates in the present experiment may actually be quite impressive considering the stimuli used.

Even though a habituation paradigm does not enhance detection rates on any given trial, it may be used to advantage by combining the results of trials. Combining the results of trials is done by finding the average response to the critical question over several trials and comparing that average with the average of responses to noncritical questions calculated across the same number of trials. This was done and detection rates in one group were as high as 75% in block one when the average of the responses to the critical questions was calculated from the first 5 trials and compared to the averages derived for the noncritical questions. The detection rates diminished in block 2.

If only detection rates are considered the differential pupillary response to critical stimuli does not appear that impressive. However, the more powerful parametric statistical analyses on the actual pupil size data revealed that the pupillary response, in the first block, was larger to the critical question than to the mean of the noncritical questions. It was not until the second block that the pupil response became the same for all questions.

The theories of lie detection described by Davis (1961) and discussed earlier in this paper would have difficulty in accounting for the data, especially since differential responding was evidenced in the control group. In none of the three groups were the stimuli likely to evoke a conditioned emotional response, and in the control group specifically there was neither need to deceive nor conflicts between lying and telling the truth. Again the hypothesis proposed by Day and Rourke (1974), that "short term familiarity" is a sufficient condition for differential responding is supported. Speculation on this finding carried to its ultimate extent would suggest that a suspect would merely have to attend to a list of stimuli to evidence detectably different physiological responses.

Establishing that "short term familiarity" is a sufficient condition for detection does not preclude the possibilty that stressful or motivating factors associated with lying about a specific item could enhance detection rates. However, little comment can be made on this issue here since this experiment was not designed to account for motivational factors and the results in the literature remain mixed (Day,1972; Gustafson,and Orne, 1963).

The results of this experiment show that an "overt lie" is not only not necessary for detection but also fails to enhance the physiological response. This finding is in opposition to that of Gustafson and Orne (1965). They used the GSR as the physiological measure and found that a verbal lie facilitated detection.

Experiment 2 also provides information on habituation of the pupil response in the lie detection paradigm. Differential responding to the critical stimulus was found in the first block but dissappeared by the second block. This occurred even with the short rest interval between blocks. Ben Shaker, Lieblich and Kugelmass (1975) found that responses to critical and noncritical items exhibited habituation curves that were similar in form. To account for differential responses to critical and noncritical items they formulated the notion of dichotomization. Items are not differentiated on an individual basis by the subject but are put into one of two categories, the critical item in the relevant cateogory and the noncritical items in the irrelevant category. In any situation where there are more noncritical items than critical, the irrelevant category will be presented more often and will be habituated sooner by virtue of greater frequency of occurrance. Their data appears to support this notion and presents intriguing possibilities of improving detection rates by increasing the ratio of noncritical to critical items. Evidence to this effect already exists. Geldreich (1941) acheived 100% detection rates when the responses to noncritical items were habituated over a series of twenty to fifty trials before questions about critical items were asked.

With the above qualifications in mind, the data in this experiment show that there is little to gained, on a per trial basis, by presenting more than five trials in the hope of improving detection rates when the ratio of the noncritical to critical items is 4:1 and the critical item is presented on every trial.

Theoretical approaches to lie detection postulating concepts such as guilt, lying, conflict, fear, etc. as psychological states necessary for differential responses to critical questions may merely represent extreme cases of a more general theory. Detection could fit into a broader paradigm where any operation designed to bring an item to the attention of a subject (Day and Rourke,1974) will result in that item being classed into a "relevant" cateogory for that subject. Once an item has been classed as "relevant" the efficiency of detection depends on how many "irrelevant" items it is presented with in a trial (Ben Shakhar, Lieblich, and Kugelmass, 1975). Thus a differential response can be explained simply by the notion that the subject has classed an item as "relevant" and has less habituation trials on "relevant" items than on the greater number of subjectively undifferentiated "irrelevant" items. One interesting prediction for the data collected in Experiment 2 would be that habituation curves generated from each critical stimulus in five consecutive trials would be most like habituation curves generated from the first four noncritical stimuli in the first trial and the first noncritical stimulus in the second trial.

Overall it may be concluded that the pupillary response can be used to measure differential effects of psychosensory stimuli. This was demonstrated in two experiments. Repeating trials was not found to be an effective method of increasing detection rates on a per trial basis. However, comparing the mean of the critical responses over a block of trials to the means of the noncritical responses does result in improved detection rates. Overt lying did not facilitate detection over covert lying and lying either overtly or covertly was not more readily detectable than having the number merely brought to the attention of the subject. Thus, traditional theories of lie detection have been brought into question. An explaination of differential responding was offered considering differential opportunities for habituation of categories of "relevant" and "irrelevant" stimuli.

References

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

Baker, L.E. The pupillary response conditioned to subliminal stimuli. Psychological Monographs, 1938, 50, No.223.

Berkhout, J., Walter, D.O., Adey, W.R. Autonomic responses during a replicable interrogation. Journal of Applied Psychology, 1970, 54, 4, 316-325.

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

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

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., & Welch, L. An attempt at reversal of the truth - lie relationship as measured by the psychogalvanic response. Journal of Psychology, 1952, 34, 55-66.

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

Brener, J., & Hotherall, D. Heart rate control under conditions of augmented sensory feedback. Psychophysiology, 1966, 3, 23-28.

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

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.

Cason, H. The conditioned pupillary reaction. Journal of Experimental Psychology, 1922, 5, 108-146.

Chapman, L.J., Chapman, J.P., & Brelje, T. Influence of the experimenter of pupillary dilation to sexually provocative pictures. Journal of Abnormal Psychology, 1969, 74,396-400.

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

Clark, W.R., & Johnson, D.A. Effects of instructional set on pupillary dilation to sexually provocative pictures. Journal of Experimental Psychology, 1970, 85, 315-317.

Cutrow, R.J., Parks, A., Lucas, N., & Thomas, K. The objective use of multiple physiological indicies in the detection of deception. Psychophysiology. 1972, 9, 578-588.

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

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

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

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

Delago, J.M.R., & Diego, M. Infagatability of pupillary constriction evoked by hypothalmic stimulation in monkeys. Neurology, 1966, 16, 939-950.

Edelberg, R. Electrical activity of the skin. Its measurement and uses in psychophysiology, in N.S. Greenfield and R.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.

Floch, M. Limitations of the lie detector. *Journal of Criminal Law and Criminology*, 1950, *40*, 651-653.

Geisser, S., and Greenhouse, S.W., An extension of Box's results on the use of the F ditribution in multivariate analysis. *Annals of Mathematical Statistics*, 1958, *29*, 885-891.

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

Goldwater, B.C. Psychological significance of pupillary movements, *Psychological Bulletin*, 1972, *77*, 340-355.

Gerall, A.A., and Obrist, P.A. Classical conditioning of the pupillary dilation response of normal and curarized cats. *Journal of Comparative and Physiological Psychology*, 1962, *55*, 486-491.

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

Gustafson, L.A., & 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., & 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 assessment of CNS functions and drug affects. In W.L. Smith (Ed.) Drugs and cerebral function, Springfield Ill., Thomas, 1970.

Harlow, H.F. The effects of incomplete curare paralysis upon formation and elicitation of conditioned responses in cats. The Journal of Geneteic Psychology, 1940, 56, 273-282.

Hess, E.H. Pupillometrics: A method of studying mental, emotional, and sensory processes. In N.S. Greenfield and R.A. Sternbach (eds.) Handbook of Psychophysiology. New York: Holt, Rinehart and Winston, 1972.

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

Hudgins, C.V. Conditioning and the voluntary control of the pupillary light reflex. Journal of General Psychology, 1933, 8, 3-51.

Inbau, F., Moenssens, A.A., & Vitullo, L.R. Scientific Police investigation. Philadelphia, Chilton Book Co., 1972.

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

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

Iwahara, S., Miseki, H., Shiokawa, N., & Yoshida, R. GSR to forced lying and personality traits. Psychological Abstracts, 1960, 6, 269. (Abstract)

Janisse, M.P. Pupil size and affect: A critical review of the literature since 1960. Canadian Psychologist, 1973, 14, 4, 311-329.

Janisse, M.P. Pupil size, affect, and exposure frequency. Social Behavior and Personality, 1974, 2(2), 125-146.

Janisse, M.P. and Peavler, W.S. Pupillary response today: Emotion in the eye. Psychology Today, 1974, 7, 60-63.

Kahneman, D., & Peavler, W.S. Incentive effects and pupillary changes in association learning. Journal of Experimental Psychology, 1969, 79, 312-318.

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

Kimmel, H.D. Instrumental conditioning of autonomically mediated responses in human beings. American Psychologist, 1974, 29, 325-335.

Kirk, R.E. _Experimental design procedures for the behavioral sciences_. California, Brooks/Cole, 1968.

Krueger, L.M. Voluntary control of pupillary responses to visual stimuli. _Dissertation Abstracts_, 1967, _28_, 6-3, 2644.

Kubis, J.G. Instrumental, chemical and psychological aids in the interrogation of witnesses. _Journal of Social Issues_, 1957, _13_, (2), 40-49.

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., Bergman, Z. The role of "Lying" in psychophysiological detection. _Phychophysiology_, 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 indicies during criminal interrogation. _Journal of Criminal Law, Criminology and Police Science_, 1968, _59_, 632-635.

Landis, C., & 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, s., 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.

Lowenstein, O., & Loewenfeld, I.E. The pupil, In H. Dawson (Ed.), *The eye.* Vol. 3. New York: Academic, 1962.

Lunney, G.H. Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement*, 1970, *7*, 263-269.

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.

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 Support Division, AD 603860. Springfield, Va: Clearing House of Scientific and Technical Information, 1962.

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

Prather, D.C. & Berry, G.A. Differential reinforcement of the human pupillary response. Behavioral Science, 1973, 18, 420-423.

Recent trends in criminal law- polygraph examinations. Comment The Journal of Criminal Law and Criminolcgy, 1973, 64, 87-89.

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

Runyon, R.P., & Haber, A. Fundamentals of behavioral statistics. London: Addison-Wesley Publishing Company, 1967.

Senter, R.J., & Hummel, W.F. Suppression of an Autonomic response through operant conditioning. Psychological Record, 1965, 15, 1-5.

Sokolov, E.N. Higher nervous functions: The orienting reflex. Annual Review of Physiology, 1963, 25, 545-580.

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

Thackray, R.I., & Orne, M.T. A comparison of physiological indicies 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.

Woodworth, R.S., & 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.

Table 1

Frequency of times subjects were detected in each group by different scoring methods.

	Overt	Covert	Control
Peaks	9*	7	8*
Means	6	10*	9*
Difference	5	6	7

*$p < .035$ binomial test

Table 2

Pupil size values for critical and noncritical questions over the post question period.

Post Qestion Period

Critical	5.24	5.38	5.23	5.15	5.08	5.05	4.99
Noncritical	5.11	5.15	5.04	4.99	4.94	4.90	4.90

Table 3

Frequency of times subjects were detected in each group by different scoring methods (trial 1).

	Overt	Covert	Control
Peaks	5	11*	6
Means	5	11*	7
Differences	5	8*	6

*p<.035 binomial test

Table 4

A comparison of questions(Q) in each of the groups(G) at each level of blocks using Tukey's HSD statistic.

Difference	q
Q at G1 in BL1 0.070	5.75*
Q at G1 in BL2 0.001	1.30
Q at G2 in BL1 0.082	7.39*
Q at G2 in BL2 0.003	2.63
Q at G3 in BL1 0.037	3.37*
Q at G3 in BL2 0.020	1.77

q(.05;1,57)

*p<.05

Table 5 :

Frequencies of detection over groups for each trial by the different scoring methods.

	BLOCK 1					BLOCK 2				
Trials	1	2	3	4	5	1	2	3	4	5
MEAN SCORES										
Group A	5	5	6	5	5	4	4	5	6	8*
Group B	11*	6	6	7	3	7	3	7	4	4
Group C	7	6	10*	5	3	4	7	5	5	5
PEAK SCORES										
Group A	5	3	7	8*	3	5	4	5	3	6
Group B	11*	6	5	6	3	6	4	5	7	4
Group C	6	4	5	7	4	3	2	4	6	7
DIFFERENCE SCORES										
Group A	5	9*	5	5	10*	5	6	6	5	5
Group B	8*	7	10*	9*	4	5	8*	7	4	6
GroupC	6	5	12*	6	3	3	6	6	4	4

*$p<.035$ binomial test

Table 6

Detection rates derived by the comparison of average responses to critical questions to the average responses to noncritical questions using mean and difference scores.

		Groups	
	A	B	C
		Means	
Block 1	15*	10*	6
2	9*	9*	6
		Differences	
Block 1	13*	11*	7
2	10*	6	7

*$p<.035$ binomial test

Figure Caption

Figure 1. Pupil size in response to critical and noncritical questions averaged across 20 subjects per group.

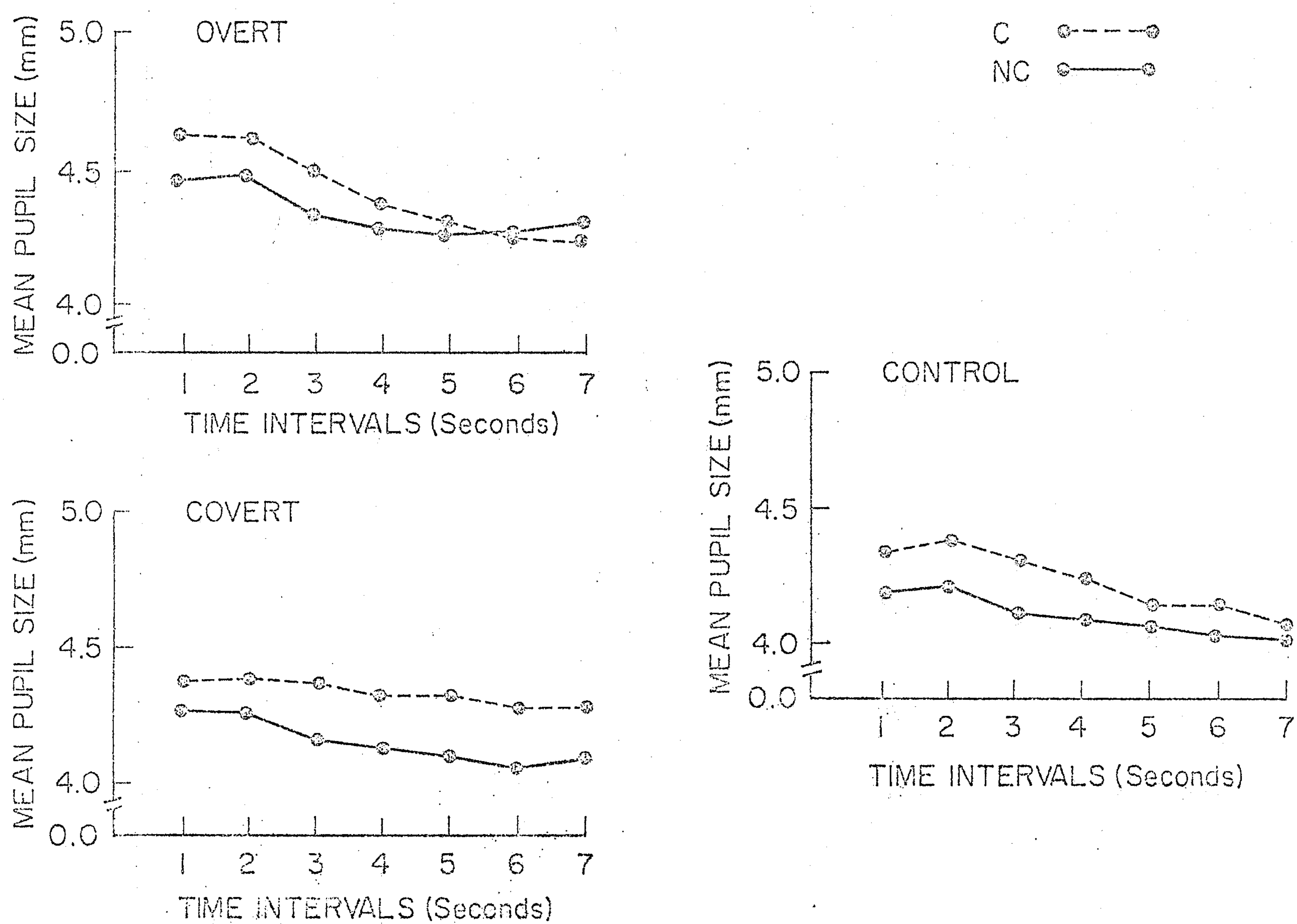
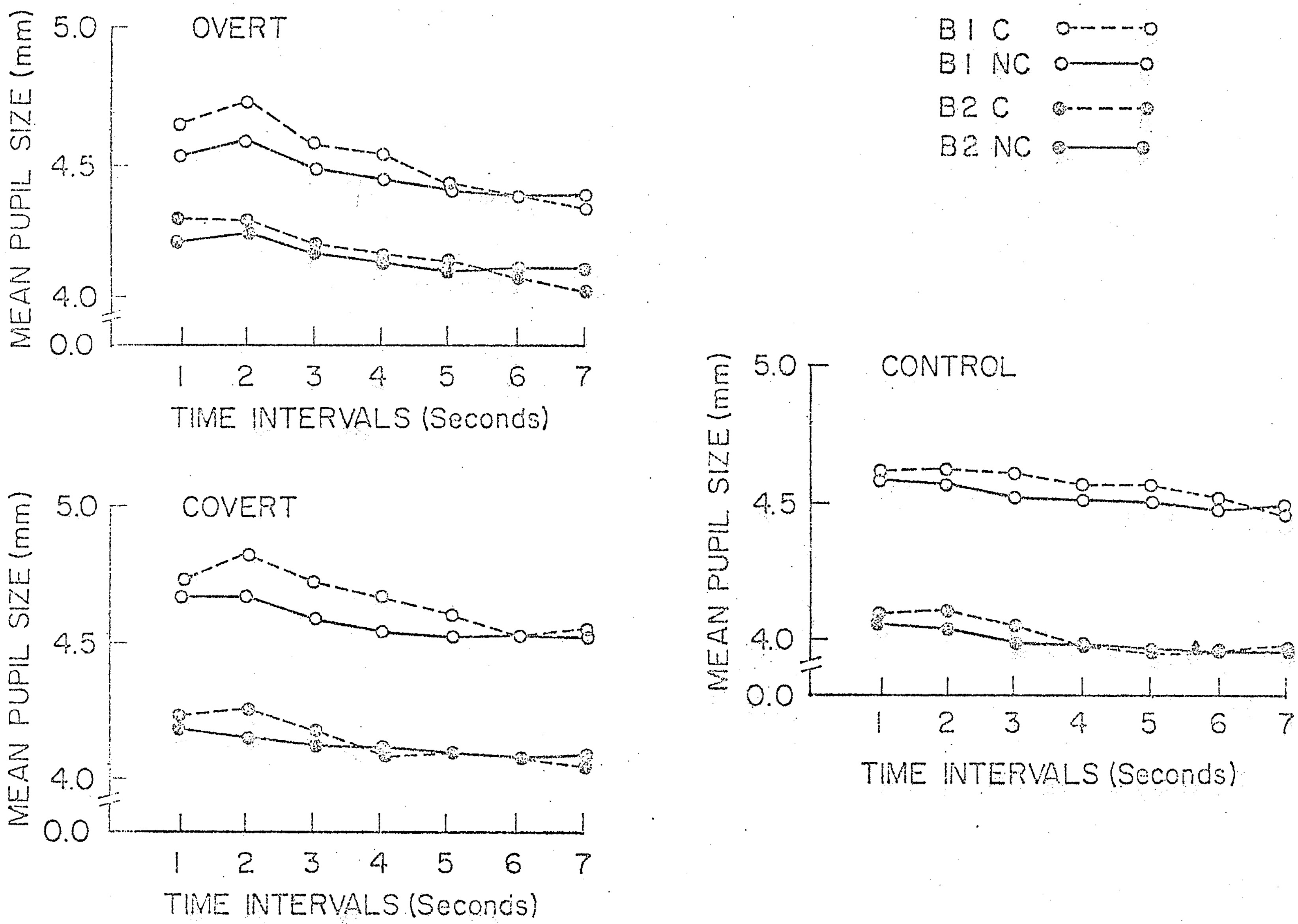

OVERT
COVERT
CONTROL
C
NC
MEAN PUPIL SIZE (mm)
5.0
4.5
4.0
0.0
1
2
3
4
5
6
7
TIME INTERVALS (Seconds)

Figure Caption

Figure 2. Pupil size in response to critical and noncritical questions averaged across 5 trials for each of 2 blocks.


OVERT
MEAN PUPIL SIZE (mm)
5.0
4.5
4.0
0.0
1 2 3 4 5 6 7
TIME INTERVALS (Seconds)
COVERT
MEAN PUPIL SIZE (mm)
5.0
4.5
4.0
0.0
1 2 3 4 5 6 7
TIME INTERVALS (Seconds)
B1 C
B1 NC
B2 C
B2 NC
CONTROL
MEAN PUPIL SIZE (mm)
5.0
4.5
4.0
0.0
1 2 3 4 5 6 7
TIME INTERVALS (Seconds)

Appendix A

A 3x2x7 Anova on actual pupil size for groups(G), questions(Q), and post-question period(P).

Source	df	MS	F
G	2	1197	0.310
ERROR1	57	3863	
Q	1	1687	19.150*
GxQ	2	115	1.311
ERROR2	57	88	
P	6	560	36.074*
GxP	12	18	1.188
ERROR3	342	15	
QxP	6	22	1.523
GxQxP	12	13	0.900
ERROR4	342	14	

*$P<.01$

Appendix B

Scheffe's analysis on the post-response period using the first 4 actual pupil size values as the response and the last 3 values as the baseline.

$$F = \frac{106.671 - 104.445}{18.430(\ .0047)} = 38.732^{*}$$

$^{*}F_c(.05,6,342) = (7-1)(2.10) = 12.60$

Appendix C

A 3x2x7 ANOVA on actual pupil size for groups(G), questions(Q), and post-response period(P).

Source	df	MS	F
G	2	3030	0.852
ERROR1	57	3375	
Q	1	2212	18.082*
GxQ	2	135	1.102
ERROR2	57	12	
P	6	710	48.710*
GxP	12	40	2.717
ERROR3	342	15	
QxP	6	25	2.048
GxQxP	12	6	0.512
ERROR4	342	12	

*$P<.01$

Appendix D

Scheffe's analysis on the post-response period using the first 4 actual pupil size values as the response and the last 3 values as the baseline.

$$F = \frac{108.79 - 104.99}{14.575(.0047)} = 55.432*$$

$$*F_c(.05,6,342) = (7-1)(2.10) = 12.60$$

Appendix E

A 3x2x5x2x7 ANOVA on actual pupil size for groups(G), blocks(BL), trials(T), questions(Q) and post-response period(P).

Source	df	MS	F
G	2	4540	0.190
ERROR1	57	23945	
BL	1	185820	78.740*
GxBL	2	3930	1.666
ERROR2	57	2359	
T	4	87239	128.577*
GxT	8	1692	2.495
ERROR3	228	678	
BLxT	4	3978	11.535*
GxBLxT	8	83	0.242
ERROR4	228	344	
Q	1	1571	18.718*
GxQ	2	59	0.703
ERROR5	57	84	
BLxQ	1	371	5.099*
GxBLxQ	2	28	0.390
ERROR6	57	72	
TxQ	4	269	2.336
GxTxQ	8	31	0.236
ERROR7	228	115	

Source	df	MS	F
TxBLxQ	4	84	1.002
GxBLxTxQ	8	76	0.918
ERRORS	8	84	
P	6	2625	112.039*
GxP	12	109	4.693*
ERROR9	342	23	
BLxP	6	82	5.493*
GxBLxP	12	32	2.170
ERROR10	342	15	
TxP	24	77	4.528*
GxTxP	48	21	1.245
ERROR11	1368	16	
BLxTxP	24	7	0.483
GxBLxTxP	48	9	0.668
ERROR12	1368	15	
QxP	6	232	13.521*
GxQxP	12	31	1.797
ERROR13	342	17	
BLxQxP	6	33	1.708
GxBLxQxP	12	17	0.891
ERROR14	342	19	
TxQxP	24	17	0.933
GxTxQxP	48	14	0.745
ERROR15	1368	19	
BLxTxQxP	24	14	0.868
GxBLxTxQxP	48	13	0.771
ERROR16	1368	16	

Appendix F

Post-response periods(P) at each trial(T).

Source	df	MS	F
P at T1	6	1154.45	63.22*
P at T2	6	700.94	38.39*
P at T3	6	547.57	29.99*
P at T4	6	284.57	15.58*
P at T5	6	236.93	12.97*
ERROR[a]	1710	18.26	

[a]pooled error term

*$p<.01$ (Geisser-Greenhouse df=1,57)

Appendix G

Simple main effects analysis of questions(Q) at each of the 7 consecutive seconds in the post-response period(P).

Source	df	MS	F
Q at P1	1	548.37	19.79*
Q at P2	1	1109.38	40.63*
Q at P3	1	821.71	29.65*
Q at P4	1	291.07	10.50*
Q at P5	1	104.44	3.95
Q at P6	1	0.75	0.270
Q at P7	1	79.57	2.87
ERROR[a]	399	27.71	

[a]pooled error term

*$p<.01$ (Geisser-Greenhouse $df=1,57$)

Appendix H

A 3x2x5 ANOVA with dichotomous data for frequencies of detection over groups(G), blocks(BL) and trials(T) using mean scores.

Source	df	MS	F
G	2	0.035	.149
ERROR1	57	0.234	
BL	1	0.240	1.430
GxBL	2	0.140	0.830
ERROR2	57	0.168	
T	4	0.186	0.843
GxT	8	0.256	1.116
ERROR3	228	0.220	
BLxT	4	0.229	1.220
GxBLxT	8	0.096	0.512
ERROR4	228	0.187	

Appendix I

A 3x2x5 ANOVA with dichotomous data for frequencies of detection over groups(G), blocks(BL), and trials(T) using peak scores.

Source	df	MS	F
G	2	0.305	1.312
ERROR1	57	0.234	
BL	1	0.082	0.387
GxBL	2	0.022	0.101
ERROR2	57	0.212	
T	4	0.193	0.894
GxT	8	0.113	0.523
ERROR3	228	0.216	
BLxT	4	0.373	1.640
GxBLxT	8	0.264	1.160
ERROR4	228	0.227	

Appendix J

A 3x2x5 ANOVA with dichotomous data for frequencies of detection over groups(G), blocks(BL) and trials(T) using difference scores.

Source	df	MS	F
G	2	0.522	2.020
ERROR1	57	0.259	
BL	1	0.963	3.346
GxBL	2	0.218	0.757
ERROR2	57	0.287	
T	4	0.340	2.050
GxT	8	0.140	0.846
ERROR3	228	0.166	
BLxT	4	0.083	0.366
GxBLxT	8	0.204	0.899
ERROR4	228	0.226	