In presenting this thesis in partial fulfilment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it
freely available for reference and study. I further agree that permission for extensive
copying of this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that copying or
publication of this thesis for financial gain shall not be allowed without my written
permission.

Department of Psychology
The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3
Date April 13, 1988
Abstract

Eighty male undergraduate students were randomly assigned to one of five groups in an analogue investigation of crime-intoxication on the physiological detection of deception. Sixty-four of the subjects committed a mock crime and half of these were legally intoxicated during the crime. Sixteen subjects committed no crime and served as innocent controls. Results only partially replicated those of Bradley and Ainsworth (1984). Whereas they found crime-intoxication diminished the effectiveness of both the control question test (CQT) and the guilty knowledge test (GKT), the present study found crime-intoxication diminished the accuracy of the CQT only for certain subjects; those who reported high subjective arousal during the crime. Results showed no alcohol effect on the GKT. In light of their results Bradley and Ainsworth suggested that alcohol may act through emotional or memory processes important to polygraphic examination. In a fully factorial design, the present study investigated the effects of threat during the crime and memory for crime details on polygraph outcome. As well, the effect of alcohol on these "emotion" and memory variables was examined. Memory was found to be an important variable in GKT accuracy but not important to CQT accuracy. Threat, as operationalized for the present investigation, had no effect on either the CQT or the GKT but a component of the threat variable, subjective arousal, was found to affect GKT accuracy but not that of the CQT. Raskin's (1979) two-response model of detection of deception is used to explain the results of this study although
The relationship of subjective arousal to polygraph outcome is unclear and requires examination in future studies.
Table of Contents

Abstract ................................................................. ii
List of Tables ............................................................. vi
List of Figures ............................................................. vii
Acknowledgement ........................................................... viii
Introduction ............................................................... 1
  History of Detection of Deception ................................. 3
  Polygraph Testing Procedures ..................................... 10
  Detection of a Lie Response: Polygraph Validity ............... 19
  Field vs Laboratory Studies of Polygraph Validity .......... 27
Countermeasures .......................................................... 36
Study Rationale ........................................................... 42
Method ..................................................................... 46
  Subjects ................................................................. 46
  Apparatus ............................................................... 47
Study Design .............................................................. 48
  Procedure ............................................................... 50
Polygraph Chart Analysis ............................................. 59
Results .................................................................... 63
  Independent Variable Manipulations .............................. 65
  Accuracy of Polygraph Tests ....................................... 71
Effect of Independent Variables on
  Polygraph Outcome .................................................. 80
  Post hoc Analysis of Effects due to Arousal ................... 85
Memory Test Outcome ................................................... 100
Additional Analyses .................................................... 104
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion</td>
<td>109</td>
</tr>
<tr>
<td>Effects of Alcohol</td>
<td>109</td>
</tr>
<tr>
<td>Effects of Threat</td>
<td>112</td>
</tr>
<tr>
<td>Effects of Memory Priming</td>
<td>123</td>
</tr>
<tr>
<td>Summary</td>
<td>129</td>
</tr>
<tr>
<td>References</td>
<td>131</td>
</tr>
<tr>
<td>Appendix I</td>
<td>136</td>
</tr>
<tr>
<td>Appendix II</td>
<td>138</td>
</tr>
<tr>
<td>Appendix III</td>
<td>140</td>
</tr>
<tr>
<td>Appendix IV</td>
<td>141</td>
</tr>
<tr>
<td>Appendix V</td>
<td>143</td>
</tr>
<tr>
<td>Appendix VI</td>
<td>148</td>
</tr>
<tr>
<td>Appendix VII</td>
<td>149</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Classification Accuracy for Various Cutoff Limits</td>
<td>75</td>
</tr>
<tr>
<td>Table 2</td>
<td>Cell Means - CQT Composite Scores</td>
<td>82</td>
</tr>
<tr>
<td>Table 3</td>
<td>CQT Hit Rates</td>
<td>83</td>
</tr>
<tr>
<td>Table 4</td>
<td>Cell Means - GKT Composite Scores</td>
<td>86</td>
</tr>
<tr>
<td>Table 5</td>
<td>GKT Hit Rates</td>
<td>87</td>
</tr>
<tr>
<td>Table 6</td>
<td>GKT Numerical Scores</td>
<td>91</td>
</tr>
<tr>
<td>Table 7</td>
<td>GKT Categorical Scores</td>
<td>92</td>
</tr>
<tr>
<td>Table 8</td>
<td>CQT Numerical Scores</td>
<td>94</td>
</tr>
<tr>
<td>Table 9</td>
<td>CQT Categorical Scores</td>
<td>98</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1  Study Design ........................................... 49
Figure 2  Mean Systolic Pressures for High and Low
Threat Groups .................................................... 69
Figure 3  Mean Systolic Pressure for Alcohol and
No Alcohol Groups .............................................. 70
Figure 4  Mean STAI Scores for High and Low
Threat Groups .................................................... 72
Figure 5  Mean SCR Scores for Subjective Arousal x
Alcohol Interaction ............................................. 95
Figure 6  Mean Respiration Scores for Subjective
Arousal x Alcohol Interaction .............................. 96
Figure 7  Mean STAI Scores for Subjective Arousal
Groups ............................................................. 106
Figure 8  Mean Blood Pressure Scores for High and
Low Arousal Groups ............................................ 107
Acknowledgements

I would like to thank my thesis committee for their guidance and suggestions: Dr. John Yuille (Supervisor), Dr. Eric Eich, and Dr. Demetrios Papageorgis.

I would also like to thank Dr. William Iacono and Dr. Chris Patrick for their valuable assistance in planning this study.

Since this was a large study involving many subjects it could not have been done without the capable help of Dr. Yuille’s research assistants. Thanks to Vanessa Farr, Doreen Kum, and Mark Turner, and a special thanks to Evelyn Tan who was often cool under pressure and provided an essential amount of humour.
Introduction

In an age of technology, with increasing reliance on technological devices to assist us, it is not surprising that great interest has been aroused by the polygraph or "lie detector". Such a device offers many advantages, especially to law enforcement agencies, not the least being the elimination of human bias. However, as appealing as the polygraph's potential benefits might be, and although proponents claim it to be a scientific and objective method, its use should not be accepted uncritically. As with any medical or psychological test, the polygraph test should be subjected to rigorous evaluation of its validity and reliability. Given the widespread use and acceptance of the polygraph one would assume a sturdy foundation of evidence was supporting its validity, but a review of the relevant literature shows that the question of validity is still not settled. There remain important questions to answer and procedural problems to solve, many of which are complex and difficult.

What is known as the lie detector test is a technique that measures psychophysiological responsivity. It detects deception in a suspect on the basis of certain physiological changes reflected in cardiovascular, respiratory, and electrodermal responses. Research suggests that polygraph techniques which utilize such physiological changes can be highly accurate in criminal investigations when conducted by well trained and skilled examiners. Just how accurate the polygraph is in actual practice has yet to be unequivocally determined, and
more field research is needed to provide a good estimate. One of the obvious problems associated with the polygraph technique, besides errors in identification of guilty suspects, is errors in identification of the innocent suspects. Estimates of the false positive rate range from a low of 8% in laboratory studies to a high of 55% in field studies. Even low rates of false positive errors can have substantial effects if the base rate of innocent suspects is high in the population being tested, a situation that likely occurs in government and employment screening programs. Indiscriminate use of the polygraph in screening procedures is inappropriate but even in appropriate applications the practice of polygraphy is "often plagued by less than optimal levels of training, competence and integrity among polygraph practitioners" (Raskin, 1986, p. 73).

Recognizing that the investigation of polygraph validity is still incomplete, the American Psychological Association (APA) drafted a resolution that was made public in an APA Press Release on February 1, 1986. The resolution said in part: "Polygraph tests used in all applied settings should be based on adequate psychological and psychophysiological training and sophistication. Their use by psychologists must be consistent with the (APA) Standards for Educational and Psychological Testing and the (APA) Ethical Principles of Psychologists. They should be used only when such use is justified by the existence of sufficient data on their reliability and validity for the particular population, context and specific purpose." The present study was designed to look at one particular issue relevant to polygraph test validity: the effect of alcohol intoxication at
the time of the crime on subsequent polygraph test results.

History of Detection of Deception

Shortly after people learned to communicate with language, they probably learned to use it to deceive. On the heels of early attempts to deceive with lies came attempts to detect such deception. Ancient societies developed many exotic methods to unmask potential or suspected liars. Torture was, and continues to be, a widely used method to extract the truth. The theory behind this practice seems to be a crude application of the negative reinforcement paradigm: tell us what we want to hear and the pain will stop. Lykken (1981) outlines several other modes of ascertaining truth that have been used throughout history. Trials by ordeal, such as holding one's arm in boiling water or bleeding from an incision, were predicated on religious or magical beliefs that the innocent person would somehow be empowered to endure the ordeal longer than a guilty person. In medieval Germany an interesting method was used to settle marital disputes involving infidelity. The husband's left arm was bound behind him while he held a short club in his right hand, and he stood in a tub buried waist deep in the ground. The wife wore a blouse which had one sleeve longer than the other, and the sleeve had a rock sewn into it. The wife was then required to dance around the tub, attempting to smite the husband, who was himself busy trying to club the wife. Unfortunately, how this process was expected to decide the truth of the matter is not recorded. The ancient Hindus employed a process that was more a psychological
than a physical ordeal. A suspect was required to chew a mouthful of rice and then spit it out on to a leaf. The person who successfully spat out the rice was truthful while the person who was left with grains of rice sticking to his tongue or palate was considered guilty.

This rice test was a primitive precursor of modern physiological tests of deception. The fear of being found out, in the guilty party, would lead to sympathetic arousal in the autonomic nervous system and in turn salivation would decrease and saliva would become more viscid, causing the rice to stick in the mouth. There was, of course, no way to guarantee that innocent parties would not be as fearful in this situation as the guilty. As investigators became more sophisticated in their attempts to detect physiological changes in the guilty, more direct measures of fear and anxiety were used. Such signs as sweating, blushing, trembling, unsteady gaze, and a racing pulse were considered to be indications of lying. Perhaps the first attempt to objectively measure these physiological indicants was made by the 19th century Italian criminologist, Lombroso. He and his student, Mosso, adapted a plethysmograph, which measured changes in volume of a limb, to produce a continuous record of pulse and blood pressure (Lombroso, 1895). A.R. Luria is reported to have measured finger tremor and reaction time of criminals while they were being interrogated, and Francis Galton found that guilty suspects showed disturbances in word association tasks involving crime relevant words (Lykken, 1981).

The development of the modern polygraphic "lie detector" did not occur until early this century. W.M. Marston, a psychologist,
is generally credited with being one of the earliest founders of polygraphy because of his reports, published prior to 1921, outlining a specific lie response which, in his immodest words, "marked the end of man's long, futile striving for a means of distinguishing truth-telling from deception" (Marston, 1938, p. 45). Marston felt this specific lie response was physiological in nature and could be detected as a transitory increase in systolic blood pressure when the person was fearful of detection but trying to maintain an outward composure. Although Marston made extravagant claims for his new methodology, his early work did have an influence on John A. Larson, a California police officer who later became a forensic psychiatrist. Larson began experimenting with measurement of a suspect's blood pressure, pulse, and respiration during questioning. He modified Marston's discontinuous method of recording blood pressure by recording a continuous measure which could be examined after the interrogation was complete. Larson seems to have had some doubts about the validity of his technique because he concluded a 1938 paper with the warning, "Because of the errors of interpretation, and these have been found to be large, a deception test alone should never be used as court evidence" (Larson, 1938, p. 895). Larson also wrote, "I originally hoped that instrumental lie detection would become a legitimate part of professional police science. It is little more than a racket. The lie detector, as used in many places, is nothing more than a psychological third-degree aimed at extorting confessions as the old physical beatings were. At times I'm sorry I ever had any part in its
development" (Larson, 1955, p. 714). It should be noted that the procedure used at the time of Larson's disillusionment was one called the Relevant/Irrelevant format, which consists of a mixture of irrelevant questions, such as: "Is your name John Doe?" and questions related to the crime: "Did you kill Mr. Smith?"

A junior collaborator of Larson's, Leonarde Keeler, was probably the first to add galvanic skin response to the three measures of blood pressure, pulse, and respiration. He also developed and manufactured a portable field polygraph, and this is the machine used, with few changes, by many polygraph firms today. Keeler eventually moved to Chicago and opened one of the first schools of polygraphy, where examiners were trained in his newly developed technique. In Chicago Keeler met John Reid, another polygraph enthusiast. Reid also opened a school in Chicago, but more importantly to the field he was responsible for developing the Control Question Test (CQT), the most widely used technique today. The CQT includes questions directly relevant to the matter under investigation and control questions that are not related but are nevertheless intended to cause an emotional reaction. Ideally, these control questions are ambiguous and phrased in a way that would lead the suspect to lie or at least to be uncertain about the truthfulness of his response. Such a question might be: "Before the age of 18 did you ever steal anything of value?" The control questions are constructed by the examiner after a pretest interview with the suspect. According to Control Question Test theory, an innocent suspect should be more concerned with the content of the control
questions and will react more strongly to these than to relevant questions such as: "Did you steal the money?" Conversely, the guilty subject will react more strongly to the relevant questions. A decision about the suspect's truthfulness is made by comparing his responses to the control and relevant question pairs. Variations of this technique will be discussed later.

A Reid student, Cleve Backster, advanced the concept of the Control Question Test and developed a variation known as the Zone of Comparison format (ZOC). The ZOC question sequence is divided into three comparison zones and a score for responses in each zone is determined. The question sequence does not change on each chart except that the control question eliciting the greatest response is placed in the middle position on the following chart. Under the ZOC scoring system this gives an advantage to the innocent suspect. Backster was also responsible for implementing numerical scoring, as an alternative to the global scoring system which had been in use until then. In a global scoring system the polygrapher makes subjective judgments of the polygraph test results, and these may be influenced by impressions of the suspect or other information relevant to the case. Although Backster's original numerical scoring procedure had many shortcomings, it was the precursor of more reliable systems such as those developed at the University of Utah and by the United States Army (cf. Raskin, 1986) which have systematized analysis and greatly reduced the subjective element in chart interpretation. Future use of computer programs, now being developed, will further reduce a polygrapher's role in the
In the numerical scoring procedure responses to questions are scored on a scale of +1 to +3 if the response to the control question is larger than that to the relevant question, and scored -1 to -3 if the control question response is smaller. The sum of these scores in each polygraph channel, and for all repetitions of the question pairs, yields a total score. If the total score is sufficiently large and negative the suspect is judged guilty, and if the score is large and positive the suspect is considered to be innocent. Any score between these two cutoff points is considered as inconclusive. The significance of this numerical scoring innovation was that it conferred, for the first time, the status of a "test" to the polygraph examination. Although standardization of scoring does not make a test valid and reliable, it at least allows for evaluation of these parameters. Interrogation with the polygraph using a global scoring method cannot be considered a true test because the global scoring method depends heavily on subjective judgments of the examiner for results, making it difficult to assess the accuracy of the psychophysiological technique itself.

Since the 1950s the use of the polygraph examination, in settings ranging from screening of potential employees to criminal investigations, has been increasing steadily. It is known to be used in many countries of the world including the United States, Canada, Mexico, Brazil, Argentina, Puerto Rico, France, Israel, Iran, Japan, China, Thailand, and the Philippines. Although this cannot be verified, its use is likely even more widespread than the above list suggests (Barland &
Despite its widespread use, the present-day polygraph has not changed much since the 1930s. The major changes have been a reduction of size and weight, the replacement of vacuum tubes by transistors (Barland and Raskin, 1973), and electronic enhancement of the physiological signals (Iacono & Patrick, 1987). A typical field polygraph will record measures of respiration, galvanic skin response, and cardiovascular changes. One additional improvement that can be found on some machines is a measure of muscular activity (EMG) that can be used to detect muscular flexing that might be used by a suspect to distort other polygraph measures.

Although there has been little change in the polygraph machinery, there has been a recent innovation in testing procedure. In a departure from the traditional objective of trying to detect when a person is being deceptive in answering questions, a technique has been developed which attempts to assess whether or not a suspect has knowledge of crime details. This method, called the Guilty Knowledge Test (GKT), was originally proposed by Hugo Munsterberg and more recently advanced by David Lykken (see Lykken, 1981). In this new approach the examiner must first establish a number of facts about the incident under investigation that only a guilty person would know. The facts are then presented in a multiple choice format and are embedded in a set of alternatives that seem equally plausible. The basic assumption of this test is that the guilty subject will show greater autonomic arousal to the significant alternative. Those who are innocent would respond randomly to the
alternatives since, to them, all have similar stimulus value. Empirical evidence to date suggests that the Guilty Knowledge Test could be a reasonable alternative or addition to some of the current polygraph practices (see Iacono & Patrick, 1987; Kleinmuntz & Szucko, 1982; Lykken, 1981). However, none of the studies have been done under field conditions. While field studies await to be done, there are still important questions that can be answered in the laboratory, and some of these questions have been addressed in this thesis.

Polygraph Testing Procedures

In an ideal situation the traditional field polygraph examination is conducted by an experienced examiner in an environment that is free of distractions and potential interruptions. Typically, the client is seen in a relatively quiet office that is plainly furnished but comfortable (Orne, Thackray & Paskewitz, 1972). The field polygraph examination is basically of two types: specific incident investigation or personnel screening (Barland & Raskin, 1973). In the first instance a person is suspected of having committed a specific crime, or of being involved in a related incident, and is given a polygraph test to determine if he or she is indeed guilty. This type of examination is usually done by police examiners in law enforcement agencies, but in some cases private examiners will conduct such an examination for companies as part of an internal investigation. The second type of polygraph examination is that used in personnel screening. In many situations where employees have access to money, classified information, or are in positions
that could be compromised by blackmail, employers make use of the polygraph to screen applicants before they are hired. The screening test may involve asking applicants questions about their sexual preferences, use of illegal drugs, or past criminal activities. The issues of interest to employers are generally much broader than the specific questions asked in a criminal investigation.

Administration of the polygraph examination is divided into three stages: 1. the pretest interview, 2. the testing stage, and 3. the posttest interrogation (Barland & Raskin, 1973). The examination generally lasts 2 hours or more, and about half of this time is devoted to the pretest interview. Before the pretest interview the examiner becomes acquainted with the suspect's dossier and the facts of the case, and may be briefed by an investigator on the topics to be discussed in the examination. At this point the polygraph equipment is set up and calibrated.

The pretest interview opens with the examiner meeting the suspect and explaining that he will be conducting the polygraph examination. The suspect is then asked to sign a statement of consent to take the polygraph examination. Once these details are taken care of, the examiner explains that the polygraph is a scientific instrument which will indicate when the suspect is lying. The suspect is led to believe that the detection of lying can be done with certainty, therefore any attempts at deception are pointless because they will be revealed (cf. Orne et al., 1972). The tone of the initial interview is meant to maximize the suspect's belief in the machine's ability to detect lies. The
questions to be asked are devised with the cooperation of the suspect and thoroughly reviewed with him so that there will be no surprises. The examiner emphasizes at this point that all questions must be answered truthfully since any doubts or reservations about the answers will be reflected on the polygraph charts. The suspect is encouraged to discuss any misgivings he may have about the questions with the examiner before the polygraph examination has begun. The examiner does not want the suspect to feel he is trying to obtain a confession, but merely attempting to obtain all the necessary information to conduct an adequate examination. If the suspect is bothered by any particular question, it can be reworded. This is often done when the suspect reveals some previously unmentioned information. For example, if the question is, "Did you ever steal anything of value from an employer?" the suspect might reveal that he had stolen a package of cigarettes when he was working in a grocery store. The question can then be prefaced with the phrase, "Except for what you have told me, did you ever........?" This process of clarifying the meaning of each question, and suggesting that it is vital that there be no confusion about the answers, is designed to illustrate that truthfulness is essential for the suspect's own good (Orne et al., 1972). The essence of the pretest interview is to focus the subject's "psychological set" on the appropriate set of questions; control questions for the innocent suspect, and relevant questions for the guilty suspect.

Only after all the questions have been carefully formulated and clarified does the polygraph examination begin. Polygraph
procedure emphasizes that the examiner establish rapport with the
suspect so that anxious suspects will relax before testing. The
pretest interview also serves to reassure the innocent suspects
about the infallibility of the polygraph and hopefully reduce
their anxiety. The guilty suspect, on the other hand, should
become maximally responsive under the same circumstances. Once
the examiner is satisfied that the suspect has understood the
meaning of all the questions, and is physically and mentally fit
to take the polygraph examination, the suspect is attached to the
instrument (Barland and Raskin, 1972).

There are some variations in the above pretest procedure
when conducted by polygraphers using the Keeler technique
(Harrelson, 1964), or the Reid technique (Reid and Inbau, 1966).
Both the Reid and Keeler techniques employ a shorter interview of
20-30 minutes as compared to a minimum of 1 hour for the pretest
interview outlined above. Users of the Reid technique also
advocate attaching the electrodes, and other polygraph sensors,
immediately after the suspect's arrival. A third, and perhaps
more important, difference in the Reid technique is that specific
questions are asked for the purpose of observing and recording
behavioral changes, which can be used to supplement chart
interpretation.

In field polygraphy, the subject is seated in a plain,
wooden chair facing away from the polygraph, so that the chart is
not visible. The polygraph is put into operation by engaging the
respiration measure first, then the skin conductance measure, and
finally the cardiovascular measure. The reason for this order.
according to Barland and Raskin (1973), is twofold. By starting with respiration and skin conductance the polygrapher obtains about 15 seconds of base-line data that may prove useful in chart interpretation. As well, by inflating the cardio cuff last the length of time it is inflated is minimized, and this relieves the subject of added discomfort.

The typical commercial polygraph equipment measures only the three physiological responses mentioned above: skin conductance, respiration, and cardiovascular changes. Skin conductance is a measure of autonomic activity that reflects the amount of sweat exuded by the skin. This response, also called the electrodermal response (EDR), is recorded by placing electrodes on a suitable area of skin, usually the distal portion of two fingers. The respiration measure generally involves a system of tubes and bellows that are attached around both the chest and the abdomen. The polygraph records the cycles of respiration as the subject inhales and exhales. The cardiovascular measure, often referred to as "relative blood pressure," is not strictly a measure of blood pressure. A pneumatic blood pressure cuff is used, but it is inflated to an arbitrary point somewhere between systolic and diastolic pressure. The behavior of the "cardio" channel appears to depend on whether the pneumatic cuff is inflated above or below the subject's mean arterial pressure. Geddes and Newberg (1977) showed that three changes in the chart amplitude of pulsatile oscillations can occur with an increase of blood pressure. When the pressure cuff is inflated above mean arterial pressure (point of maximum oscillations) an increase in blood pressure will cause an increase in amplitude of oscillations.
Conversely, if cuff pressure is set below mean arterial pressure, an increase in blood pressure results in a decrease in oscillation amplitude. A cuff pressure that is equal to the mean arterial pressure will result in no change in amplitude of oscillations when blood pressure increases. The above study suggests that chart changes due to blood pressure increase or decrease are not uniform, but depend largely on the level of inflation of the pressure cuff and should only be considered as a relative measure.

Once the sensors are attached and the cardio-cuff is inflated, the subject is informed that testing is about to begin. Questions are asked in a low, audible voice, without any inflections which may influence a subject's response. The questions are those that have been formulated and discussed with the subject during the pretest interview and usually require a simple yes or no answer. Between each question the interrogator waits 15-20 seconds to allow each of the three measures to return to a reasonably stable baseline. During questioning the interrogator makes notations on the chart of relevant information such as sensitivity of polygraph amplifiers, when questions begin and end, observations of movement, and so on. After completion of the first set of questions it is common practice to perform a card test (also called a "stim" test) with the subject (Barland & Raskin, 1973). Some polygraphers may use the card test before the first set of questions is asked (Orne et al., 1972). The purpose of the card test is to increase the subject's confidence in the accuracy of the polygraph. The subject is asked to pick one of a
number of playing cards and conceal it, then the polygrapher names each card and asks if it was the one chosen; the subject replies "no" each time. The ostensible reason for this test is to calibrate and adjust the polygraph, but in fact some polygraphers arrange things so that they know which card the subject picked regardless of what the polygraph says. The polygrapher then tells the subject which card was picked and says the equipment is operating well and testing can begin. This demonstration is seen as a way to enhance the subject's confidence in the polygraph, and consequently makes guilty subjects more reactive and innocent subjects less fearful of false detection. Following the card test, polygraph testing is continued (or begun, if the card test is done first). The question set is generally asked three times, yielding three "charts", but a fourth or fifth chart may sometimes be necessary to make a decision as to guilt or innocence. Between each test the interrogator may allow the subject to reconcile or explain any unusual or incriminating physiological responses.

The treatment of the subject after completion of testing depends upon the results. If the subject is felt to be innocent, he may be told the outcome and informed that the results will be given to the investigators. If the examination suggests that the suspect was being deceptive, he may be immediately interrogated so as to exploit any psychological advantage inherent in the testing situation. Guidelines for conducting posttest interrogations vary among polygraph schools, but detailed sources are available (cf. Barland & Raskin, 1973).

There are several variations of the CQT (see Lykken, 1981)
each with its own sequence of question presentation. In some, the sequence is predetermined and inflexible, in others the sequence is altered from chart to chart to avoid habituation. One commonly used format is the Zone of Comparison (ZOC) test developed by Backster (1962). The ZOC procedure utilizes a number of safeguards designed to compensate for variations in subject reactivity during the polygraph examination. The ZOC question format also includes irrelevant, non-threatening questions to provide a rest stop between more arousing questions, and outside issue questions designed for the situation in which the subject might be concerned about other disturbing information. Such a question might be, "Do you understand you will only be asked the questions we have discussed?" If the subject appears disturbed by this question, testing can be postponed until he is convinced that only the specified questions will be asked. The questions in the ZOC test that are actually used to determine guilt or innocence are the lie control and crime relevant questions. The control questions are designed to act as a baseline of the subject's physiological reactions when he is lying. Reid and Inbau (1977) explain that the polygrapher should choose as controls only questions to which the suspect shows symptoms of deception, such as hesitation, breaking eye contact or squirming. Some examiners choose questions that refer to things they feel everyone has done, so any "no" answer is automatically assumed to be deceptive. An example of a lie control question is, "Before age 19 did you ever lie to get out of trouble?"

In the current investigation, a CQT format was employed that
consisted of 9 questions; 3 questions irrelevant to the crime, and 3 control/relevant question pairs. To maintain consistency, the questions were predetermined and identical for all subjects. Similarly, subjects were administered 3 charts, regardless of whether their charts appeared to be inconclusive or not. The full CQT question format used in this study, along with the designation of each question (eg. lie control), can be seen in Appendix V.

An alternative to the CQT, which does not depend on the examiner's skill in manipulating a subject's expectations and beliefs, is the Guilty Knowledge Test. The GKT consists of questions to which only guilty subjects could know the answer. The questions are posed in a multiple choice format, with only one correct answer, and the subject must either repeat each alternative or answer "no" to each one. A typical GKT question might be, "If you murdered Mary Jones, you would know what the murder weapon was. Was it a gun?.... a hammer?.... a knife?.... a rope?.... a bottle?" When confronted with such information the guilty person is expected to give the largest physiological response to the correct answer, if not out of guilt or fear of detection then possibly out of recognition. An innocent person, knowing none of the specifics of the crime, would yield random physiological responses to the same questions. The advantage of the GKT is that it offers protection to the innocent. In a series of 10 questions, with 5 alternative choices in each, and assuming each alternative is equally arousing, the chance of an innocent suspect giving his largest response to the correct alternative on each question is $1/5$ raised to the power of 10, which is one
chance in 10 million. The accuracy of the GKT can be compromised in favor of the guilty suspects if too few questions are asked or if the questions are not carefully formulated. Criticisms leveled at the GKT are that it is often impractical to use in criminal investigations, and it is not compatible with the task of personnel screening. Podlesny and Raskin (1978) also argue that both innocent and guilty criminal suspects have typically been exposed to much of the crime relevant information before testing. They state further that, "Time consuming field work and considerable ingenuity are required to construct a valid GK test, while a CQ test may be used immediately" (p. 345). On the other hand, Reid and Inbau (1977) describe some well conceived and successful uses of the single-item, guilty knowledge method in the field. Regardless of the arguments mounted for or against the GKT, its usefulness in the field has not been systematically studied and it remains to be validated in the context of real criminal investigations. This study will use a GKT procedure that asks 5 questions about crime relevant items, and each question will have 5 alternative choices. The GKT format is outlined in Appendix V.

Detection of a Lie Response: Polygraph Validity

What has become known as the lie detection test is, in fact, simply a physiological measure of reactivity to questioning. The cornerstone assumption of a testing procedure deemed to be a detector of lying must be that a specific, physiological reaction is associated with attempts to deceive. It follows that any
physiological response to deception must be characteristically different from other responses that can be expected of a person during questioning, such as anxiety, stress, or arousal. It is plausible that lying, for many people, is accompanied by some physiological reactions; after all most people have the subjective experience of "guilt" when lying. However, can we reasonably assume that there is a specific lie response, some involuntary but distinctive psychophysiological reaction that is produced by people only when they are lying and never when they are being truthful? Contrary to the belief of many, there is no evidence to support the idea of such a specific lie response.

Since we can distinguish our own various emotional states, such as delight, fear, anger, or sadness, as well as guilt, it is natural to assume that each of these emotions is expressed in its own particular physiological pattern. If this were true, the response associated with guilt might be a good candidate for the specific lie response. Unfortunately, there is no evidence that physiological patterns of response to various emotions are consistent across individuals. Albert Ax (1953) measured many different physiological variables, including blood pressure, heart rate, stroke volume, respiration, face and hand temperature, sweating, and muscle tension, while subjects were in a state of fear or anger. He concluded that the two stimulus conditions of fear and anger produced physiological reaction patterns that were significantly different within individuals. However, he also found that for each emotion there was a general lack of correlation between the physiological variables across individuals. This lack of correlation supported the hypothesis
that people are unique in the physiological expression of their emotions. This hypothesis was further supported by the finding that the between subject variance of the physiological measures was significantly larger than the within-subject variance.

Schachter (1957) conducted a study similar to that of Ax. The Schachter study had groups experiencing fear and anger but it also included a third group which experienced "cold pressor" pain, and five physiological variables not used by Ax were recorded: systolic pressure, diastolic pressure, peripheral skin resistance, cardiac output, and inspiratory index. Results of the Schachter (1957) study are consistent with those of Ax. Considering all subjects, the average response patterns were different for the pain, anger, and fear groups, but the response patterns of individual subjects were variable. This suggests that using a specific physiological pattern to distinguish when an individual is in pain, as opposed to being fearful or angry, would not be highly successful, regardless of the accuracy of measurement.

The conclusions from these and other studies of autonomic response to emotions converge on the same point, that generalized and undifferentiated arousal is easily detected, but as we try and distinguish between more distinct emotions such as guilt, shame, or anger, detection becomes increasingly unsound. Since the polygraph is typically used to measure peripheral, autonomic responses (i.e. cardiovascular, electrodermal, and respiration), it is highly unlikely that it will ever be able to accurately detect qualitatively different emotional responses in the
individual. The polygraph can tell the experienced operator if a subject is more aroused by one stimulus than another, but it can not reveal with certainty if that arousal is due to pride, guilt, anger, or unbridled fear.

In addition to studies of the physiology of emotion there have been investigations that approach the lie response issue from a different perspective. These studies were aimed at understanding the cognitive processes involved in the physiological response to deception. Kugelmass, Lieblich, and Bergman (1967), employing a procedure similar to the card test discussed above, found that subjects could be detected even when they had been instructed to answer "yes" rather than "no" to each question. In answering "yes" they were technically telling the truth when asked about the critical item but were nevertheless found to be deceptive. Such findings suggest that the verbal act of lying itself may not be as important as the inherently arousing nature of the critical questions. A study by Thackray and Orne (1968) illustrates that the nature of the questions can affect detectability. Subjects memorized three critical "code" words that had been randomly chosen by the experimenter, then they were told that they would be interrogated about these code words and about their first names, last names, and date of birth. The polygrapher was blind to the subject's personal information and to the code words. This procedure allowed the authors to compare the rate of detection of the subject's personal information (which inherently has more meaning to the subject than randomly chosen words) with the detection rate of the word list. The detection rate of the personal material was found to be
higher than that of the over-learned words. This finding supports the idea that critical questions acquire arousal properties that contribute to detection apart from fear or effort to deceive. Like a person's name, crime related details may evoke a conditioned physiological response in the suspect that can be reliably detected. A polygraph method, such as the GKT, that utilizes the conditioned properties of crime-relevant details avoids the problem of having to distinguish a "lie" response, a problem inherent in traditional detection techniques. However, the strength of the GKT is also its weakness because many situations such as employment screening, or interrogations that are of a general nature, do not lend themselves to the specific detail analysis necessary for construction of the GKT.

Although the polygraph cannot be considered a true "lie detector," for the reasons outlined above, it can be considered as a diagnostic tool. The validity and reliability of such a tool depends heavily on how it is used and how the results are interpreted. The following discussion will focus on some methods that are used in the field of polygraphy, the logic of each approach, and a critical analysis of how these methods work.

The idea that there is a specific response to lying, and that the polygraph can infallibly detect it, is very important to current lie detection procedures. The importance of a subject's belief in the infallibility of the test is reflected in a comment by Waid and Orne (1981):

The ostensible purpose of the [card] test is to assess the individual's typical response to deception, but the actual purpose is to demonstrate to the subject that the polygraph
can, indeed, distinguish between truthful and deceptive answers. Such a demonstration is viewed as enhancing the belief in the accuracy of the test and, consequently, enhancing the physiological responses of the deceptive subject while also reducing the fear of false detections among truthful subjects (p. 66-67).

Another perceived advantage of the subject's belief in the polygraph's infallibility is that this may help elicit behavioral reactions that can be used by the polygrapher in his determination of guilt or innocence. For example, if the subject believes in the accuracy of the polygraph and his reaction to testing is hostility, annoyance, or some other unsympathetic response, guilt is indicated. A statement by Reid and Arther (1953) is revealing:

This belief that the innocent have in the accuracy of the lie-detector, and that they will be exonerated, is usually shown by their attitude. This attitude is one of genuine confidence in both the machine and the examiner. Because of this confidence they regard the examination as an experience they will want to relate to their family and friends (p. 106).

Both aspects of the polygraph test; that the subject be led to believe in the procedure, and that there are behavioral indices of deception, raise an important question. Is the actual lie detector the machine or the operator?

Some polygraphers believe that the examiner is the real detector of deception, as indicated by Barland and Raskin (1973):

Among field examiners there are two schools of thought regarding the information to be consciously used in making the [truthful or untruthful] decision. One school advocates that only by utilizing all possible information available to him from every source does the examiner optimize the probability of making the correct
decision............ The other school advocates the position that all extraneous sources of data, while sought out before and during the pretest interview, must be rigorously excluded from consideration when the decision is made (p. 436).

Polygraphers of the first school use polygraph recordings only as an adjunct in reaching their diagnosis. These practitioners think of themselves as experienced diagnosticians who rely heavily on the feel, or gestalt, of all the information (i.e., police records, observed behavior, and court records) when making their decision. In contrast, some of the more recently trained polygraphers place more emphasis on the polygraph charts. Although the subjective, all-inclusive approach to detection is the most widely used, it is the most difficult to evaluate. The interrogation procedure is as much art as science, and outcome depends a great deal on how the polygrapher synthesizes all available data. The dangers of intuitively integrating data have been outlined in the context of psychometric testing (see Wiggins, 1973), as well as with polygraph testing itself. Szuko and Kleinmuntz (1981) analyzed the judgments of polygraph experts who blindly evaluated the polygraph charts of subjects who were guilty or innocent of committing a mock crime. The authors found that statistical analysis of the charts significantly outperformed the judges. They conclude that statistical methods of interpreting polygraph data should be developed and used. However, before satisfactory statistical models can be developed, the subjective approach to scoring will have to be abandoned in favor of a standardized objective scoring system. Such objective systems, although they exist (cf. Raskin, 1986), are often not
used by practicing field polygraphers (Ben-Shakhar, Lieblich, & Bar-Hillel, 1982). The present study employed an objective scoring system (described in detail in the Method section) in the analysis of both the CQT and GKT techniques.

As may be apparent from the above discussion of the lie response, any practical realization of CQT theory rests heavily on the skill of the examiner in manipulating the subject's "psychological set" as is done when choosing control questions or convincing the subject of the machine's infallibility. Ignoring, for the moment, problems of scoring, the outcome of testing depends on the innocent subject's relative arousal to control questions and the guilty subject's arousal to the relevant questions. However, there is no good reason, physiological or otherwise, to believe that an innocent person would not be aroused by an emotionally laden question such as "Did you murder Mr. Smith?" Proponents argue that CQT theory can only be appreciated when viewed in the context of the whole testing procedure, including the pretest interview. Great care is taken, as pointed out in the above discussion of polygraph test procedure, to impress upon the subject that attempts at lying are useless and that he must be truthful at all times. Tactics like the "stim test" are often used to aid in convincing the subject that the polygraph is infallible in its detection of lying. Those who believe in the validity of the CQT are convinced that careful preparation of the subject will have different consequences for the guilty and the innocent. The innocent subject, who is convinced the machine is infallible, will have little worry about the relevant questions but will be aroused by the control
questions, which are designed to create doubt about the answer. The opposite is true of the guilty subjects, who focus on the relevant questions. They are aroused by their answers because they know they must lie, and have been led to believe that detection is inevitable. Whether these manipulations consistently have the desired effect or not is an empirical question and must be decided through carefully designed experimentation. An analysis of the validity of control question techniques, on the basis of existing data, is done in a later section.

Field vs Laboratory Studies of Polygraph Validity

Neither field studies nor laboratory studies are adequate, on their own, to provide a good estimate of polygraph validity. Although the literature contains studies of both sorts, each of these approaches has its own strengths and weaknesses. Many of the variables affecting polygraph outcome can be controlled in a laboratory setting, but the contrived nature of these studies makes the results questionable when generalized to a real-life situation. In the laboratory, subjects are often naive individuals who commit a mock crime, and although this crime provides an unambiguous criterion for determining guilt or innocence, it leaves in question the motivational and emotional concerns of the subject. In comparison to criminal situations, a subject in the laboratory often has little incentive to pass the polygraph test. Both guilty and innocent subjects have little at stake if they are found to be deceptive.

Although a field study provides the requisite emotional and
motivational factors, and involves actual criminal suspects, the issue of guilt or innocence is very difficult to determine. Reliance on judicial outcome is not a suitable recourse since some people will be falsely convicted of a crime, and probably a greater number will be absolved of crimes they did commit on technical grounds or lack of evidence. The matter is equally uncertain if confession is used as a standard, because confessing may not be independent of the polygraph interrogation. Those who do poorly on the polygraph test may be moved to confess in the belief that their deception has been exposed, and those who do well in testing may be less prone to confess. And, for a variety of reasons, some innocent victims may give a false confession. Any of the above scenarios would result in a biased estimate of polygraph validity. When reviewing the literature that addresses the issue of validity, the limitations of both the field and the laboratory study must be kept in mind.

In a review of lie detection literature, Iacono and Patrick (1987) point out that there have been 9 field studies conducted to assess CQT accuracy. These studies were conducted by social scientists trained in experimental methodology and by polygraphers associated with the Reid College for Detection of Deception. Two important procedural points can be used to distinguish among these studies: method of determining guilt or innocence and method of scoring charts. Establishing ground truth is obviously necessary when looking at validity, and the scoring method is an important variable when the role of the psychophysiological test is to be considered separately from the examiner's input in determining guilt or innocence. If a
subjective scoring system is used, there is no way to tell how variables such as the suspect's demeanor, police evidence, or personality interactions influenced the polygrapher's determination of deception. If a numerical scoring system is used the charts can be scored blindly so that variables other than chart scores do not enter into the final decision.

In the 9 studies, determination of guilt or innocence was made in two ways, by panel or by confession. Bersh (1969) conducted a study in which the polygraph examiner's decision was compared to the decision of a panel of four attorneys. The attorneys made their judgment on the basis of the suspect's legal file, which did not contain any polygraph evidence. Since the polygrapher also had access to case information prior to polygraph testing, the role played by the polygraph alone in his decision cannot be estimated. This study, therefore, cannot be used in assessing CQT accuracy. Barland and Raskin (1976) also used a panel, comprising 2 defense attorneys, 2 criminal prosecutors, and a judge, to establish guilt or innocence. A weakness of this study, according to the authors, is that suspect data used by the panel were compiled by "inexperienced" students. The resulting incompleteness of the suspect files may have led to criterion errors by the panel. A strength of this study is the use of a numerical scoring system in which the scores were summed over all of the suspect's charts. The charts were then analyzed by a person blind to the experimental groups. The results of this study indicated that 45% of the innocent suspects were correctly identified, and 98% of the guilty were identified,
for a total accuracy of 71%.

Three studies used confessions as the ground truth criterion. As mentioned previously, a confession-verified sample has several threats to its validity, not the least of which is a probable bias towards confirming the examiner's decision by those subject's deemed to be guilty. Horvath and Reid (1971) conducted a study in which they used case investigations that had been conducted by the senior author. In these cases 75 subjects had been originally tested but only 40 subjects were used in the study. The 35 subjects not used were eliminated because their charts were deemed, by the authors, to be "dramatically indicative of truth or deception" (p. 276). This subjective elimination of charts leaves an experimental sample of unknown composition. The charts were scored blindly by ten trained polygraph examiners, but a numerical scoring system was not used. The ten examiners achieved an average of 91% in correctly identifying innocent suspects and an 85% identification of guilty suspects, for an overall accuracy of 88%.

Horvath (1977) used cases drawn from the files of a large police agency, and selected those from "verified investigations," verification being defined as a case in which the polygrapher's decision was confirmed by confession. Ten field-trained polygraph examiners evaluated the charts as indicating "truthfulness," "deception," or being "inconclusive." The evaluators were blind to the outcome of the original investigation, but they used the original charts and notations to arrive at their decision. Numerical scoring was not used. In this study it was found that 51% of innocent subjects and 77% of guilty subjects were
correctly identified, for an overall average of 64%.

The final study using confession as a determinant of ground truth was that of Kleinmuntz and Szucko (1984a). The subjects in this study were fifty confessed thieves and fifty innocent suspects who had been cleared by the confessions of others. There is no information given as to how or when confessions were made, leaving open the possibility that the original examiner obtained some or all confessions based on chart results that indicated deception. Another problem with the subject population is that all were suspects in cases of theft and such suspects may be harder to detect (Horvath, 1977). The charts in this study were scored blindly, using a global scoring system, by six polygraph examiners who were reported in a subsequent paper (Kleinmuntz & Szucko, 1984b) to be "polygraph trainees." In addition, only one chart was scored for each subject where it is standard practice to use three or more. These aspects of the Kleinmuntz and Szucko (1984a) investigation may have diminished accuracy rates. Of the innocent suspects 63% were correctly identified, and 75% of the guilty were identified, yielding an overall accuracy rate of 69%.

The remaining 3 studies (Hunter & Ash, 1973; Slowick & Buckley, 1975; Wicklander & Hunter, 1975), which were published in law enforcement journals, did not report the criteria that were used to establish ground truth. Although there were other weaknesses of these studies they are overshadowed by the absence of a clearly defined ground truth measure, without which there can be no confidence at all placed in the findings.

On the basis of the field studies reviewed above, it is
apparent that CQT validity has yet to be clearly determined. This evidence suggests that the CQT correctly determines guilt at above chance levels (range 75-98%) but it correctly identifies the innocent at a poorer rate (range 45-91%). This means that 9-55% of innocent suspects were being misclassified as deceptive in the above studies.

Raskin (1986) presents a somewhat brighter picture of CQT validity in his review of 5 laboratory studies conducted between 1978 and 1982. These 5 studies all employed non-student subjects, field polygraph techniques, trained polygraphers, and used a mock crime procedure. Classification of guilty subjects ranged from 89% to 100% correct, while classification of innocent subjects ranged between 82% and 100% correct. The combined accuracy rate for all subjects was 95% with an inconclusive rate of 8%. Of the misclassifications, 8% of the innocent subjects were determined to be guilty (false positives) and 3% of the guilty were classified as innocent (false negatives). Raskin states that these misclassifications "indicate the limitations of the control question technique, even when it is performed under carefully controlled conditions by highly skilled examiners with extensive psychological training and expertise" (p. 42).

It appears that the GKT is not used much in the field (Iacono & Patrick, 1987), so it is not surprising that there have been no field studies of GKT validity. The first laboratory study with the GKT was done by Lykken (1959) when 98 university students were assigned to one of four groups: a mock theft group, a mock murder group, a group that committed both mock crimes, and a control group that committed no crime. The test consisted of
six multiple-choice questions for each of the two crimes. Of 50 guilty suspects tested, 44 were correctly classified as guilty, while all 48 innocent suspects were correctly judged to be innocent, for an overall accuracy of 94%. These results have since been replicated in other studies. Geisen and Rollison (1980) had 20 subjects commit a mock crime and then tested them along with 20 innocent subjects. Like that of Lykken (1959) the GKT consisted of six items and was able to correctly identify all the innocent subjects and 95% of the guilty. Podlesny and Raskin (1978) used only five items in an experiment of similar design and correctly identified all innocent subjects and 90% of the guilty subjects. Although all laboratory studies have found false negative rates to be greater than false positive rates, the guilty "hit" rate has varied. However, a virtue of the GKT method, not shared by more traditional procedures, is that the discrimination of guilty from innocent subjects can be increased by increasing the number of critical items in the test. Lykken (1981, p. 300) has devised a table suggesting the theoretical accuracy of the GKT for tests of increasing length. According to the table, a GKT of five items would be expected to produce a true positive rate of 87.2% whereas a GKT of sixteen items should be 99.3% accurate. The rates predicted by the table are based on the assumption that all questions have been carefully constructed. Factors involved in construction of questions such as which details of the crime are likely to be remembered and what conditions might impair recollection of the crime are matters that require further research both in the laboratory and
ultimately in the field. The present study has examined some of the memory and affective issues that could have implications for test construction.

While some studies, for reasons of experimental control, are best conducted in the laboratory, steps can be taken to moderate the artificial situation. Several methodological issues important to laboratory studies are considered in this thesis. One criticism often leveled at laboratory studies is that subjects have little motivation to avoid detection. Gustafson and Orne (1963) demonstrated the importance of subject motivation. In that study one group listened to a tape that told them: a) the experiment was designed to see how well the subject could keep information from the experimenter, b) that this was extremely difficult and only persons with superior intelligence and great emotional control could succeed, c) they were to try as hard as they could to beat the experimenter and the equipment, and d) if they were successful, they would be paid. The second (control) group had no preexperiment instructions. Subjects picked a playing card and then the experimenter used the polygraph to determine which card had been chosen. Results showed that the control group had their card detected at rates no better than chance, while the motivated group were detected at rates significantly better than chance. In the present study a procedure similar to that of Gustafson and Orne (1963) was used to increase motivation in all groups. All subjects received the motivational instructions described above although payment did not depend on "beating" the polygraph.

Another important issue is that of crime salience. Many
laboratory studies use simple tasks, such as memorizing a list of words or choosing a card, which require little involvement or attention by the subject. Since the importance of attention to stimuli in physiological responding has been demonstrated (Corteen, 1969; McLean, 1969; Sampson, 1969) it is necessary to involve the subject as much as possible in the task, and presumably the more salient the task the more attention will be paid to relevant stimuli. Following the procedure of Bradley and Ainsworth (1984) each subject in the present study received instructions which outlined the crime and some subjects had their memories primed for salient details after they committed the crime (see Method section).

One final methodological issue worth noting is a situational distinction made by Ben Shakhar, Lieblich, and Kugelmass (1970), that of "certain" and "uncertain" situations. In the "certain" situation it is known that the subject did in fact do something, eg. pick a card from a deck of five. The examiner's task is to determine which of the cards was picked. In this case the examiner knows one response is deceptive and must identify it. In an "uncertain" situation the examiner is not sure that the subject is guilty so he is no longer concerned with maximizing correct identifications but he must also attempt to control misses and false identifications. Clearly, "uncertain" situations are a much more rigorous and realistic test for polygraphic detection of deception. By design the present study confronted the examiner with an "uncertain" situation since there was an innocent control group who provided no deceptive responses during
testing.

**Countermeasures**

As pointed out earlier, the commonly used field polygraph procedures tend to yield high false positive rates (9-55%). This issue is important because our legal system is designed, and should serve, to protect the innocent. Another issue, important in terms of polygraph validity, is that of false negatives, or the number of guilty suspects who "beat" the test. There are a variety of countermeasures that can potentially be used by the guilty to appear innocent, including: mental countermeasures, practice in making false responses, hypnosis, physical countermeasures such as unobtrusive flexing of a muscle, pain, adrenal exhaustion, controlling respiration, and chemical countermeasures (Barland & Raskin, 1973). Of these, one that has attracted recent research interest is that of chemical countermeasures. Drugs are often available, even to individuals in prison, and their use before or during polygraph testing could represent a serious threat to validity.

Waid, Orne, Cook, and Orne (1981) found that 73% of guilty subjects were able to avoid detection while under the influence of the tranquilizer meprobamate. Their study was conducted in a laboratory setting and the "guilty" condition consisted of having subjects overlearn a list of six common words. The guilty subjects were randomly assigned to one of three groups: a group which received a 400 mg dose of meprobamate, a group which received a placebo, and a control group which received no drugs. Thirty minutes after ingesting the drug, both innocent and guilty
subjects were given a polygraph test (GKT) by an experimenter who was blind to the subject's assignment. After the test the examiner rated the subject's drug condition, indicating whether or not it appeared that a tranquilizer had been taken. Results indicated that 73% of guilty subjects who were under the influence of meprobamate escaped detection, while the placebo group escaped detection only 23% of the time, and 100% of the innocent subjects were correctly identified. Furthermore, the drug affected behavior so subtly that the examiner could not distinguish between tranquilized and nontranquilized subjects.

In contrast to the Waid et al. (1981) study, Iacono, Boisvenu, and Fleming (1984) found that 100% of guilty subjects who had taken a tranquilizer, in this case diazepam, were detected. Similarly 100% of guilty subjects who took a stimulant (methylphenidate) were identified. Determination of guilt or innocence was made on the basis of blindly scored charts. The examiner in this case could identify those subjects who had taken a tranquilizer. Apart from the difference in tranquilizing drug used, these two studies also differed methodologically. Subjects in the Iacono et al. (1984) study watched a videotape depicting a burglary rather than memorize a list of words, and subjects were tested to see what details of the videotape each had remembered. A significant relationship between the ability of subjects to recall crime-related details and detectability was found; those who remembered more details were more likely to be detected. In light of the procedural and drug differences in these studies, the matter of tranquilizer effects remains unresolved.
Evidence for another sort of drug countermeasure has recently emerged. Bradley and Ainsworth (1984) conducted a study in which subjects committed a mock crime while either alcohol-intoxicated or sober. The following day the group of guilty subjects was further subdivided so that half underwent polygraph testing while intoxicated and half while sober. Each subject was tested with both the Guilty Knowledge Test and the Control Question Test. The principal conclusion of this study was that alcohol intoxication during commission of the crime reduced subsequent detectability on both the GKT and CQT. Intoxication while taking the polygraph appeared not to be an important variable, which argues against its effectiveness as a countermeasure during testing. In addition, the examiner, although blind to the alcohol state, could accurately judge which subjects were inebriated at the time of testing. Although the alcohol effects appeared to be substantial with both the CQT and GKT, the authors could not explain why the alcohol had acted to reduce detectability. They speculated that the alcohol affected emotional and memory processes that are important to polygraphic detection of deception. This stems from the assumption that the CQT relies on a person's emotional reaction to a question whereas the GKT depends more on memory for crime details.

The issue of emotionality at the time of the crime and its subsequent effects on polygraph outcome is not well understood. Davis (1961) outlines three theories that attempt to explain why physiological responses to polygraph questions occur: conditioned response, conflict, and punishment. Conditioned response refers to a process whereby the relevant polygraph questions are stimuli
conditioned to evoke emotional responses because of an association with the crime. The conflict theory assumes people are socialized to be truthful, and deceit causes conflict with this truthful tendency leading to the physiological response. The third theory, punishment, suggests that fear of the consequences of detection is what causes an autonomic response. None of these theories adequately explains all the situations in which successful detection of deception has been observed (see Raskin, 1979). In an attempt to account for previous theoretical inconsistencies, Raskin (1979) describes a conceptual framework based on the ideas of arousal and attention. He suggests that questions on the CQT would create greater emotional arousal in the subject than questions on the GKT. This is a result of the CQT relevant and control questions being accusatory, threatening, and personal, and the fact that these questions involve a direct denial of guilt by the subject. Raskin further postulates that the GKT and CQT may be utilizing characteristically different physiological responses. Where the GKT may be evoking an orienting response (OR) to relevant items, the CQT may be evoking a defensive response (DR). It follows that events at the time of the crime that in some way disrupt one or the other of these responses would affect polygraph outcome, and would affect the CQT and the GKT differentially.

The hypothesis that memory is an important variable in response to the GKT questions is consistent with the results of Iacono et al. (1984) which suggested that the ability to recall crime details is related to detectability. However, a recognition
memory test given by Bradley and Ainsworth (1984) showed high rates of recall across all groups. The authors postulated that their memory test may have been insensitive to memory disturbances for two reasons. The GKT was serial and auditory, and thus the subjects had to make a decision about each word as they heard it, which may have caused some uncertainty about the critical item. The final memory test, in contrast, presented all items simultaneously and asked the subject to recognize the critical item, which may have been an easier task. The second possible reason for memory test insensitivity is that it came after the GKT, and therefore the subjects were given exposure to the critical information before the memory test which may have aided them in recall. The authors conclude that memory effects should not be discounted on the basis of their results.

Although Bradley and Ainsworth (1984) did not demonstrate any memory effects as a result of alcohol intoxication, there is evidence that such effects do occur. The effects of various drugs, including those in common use such as marijuana and alcohol, have been studied and reported in the state-dependent learning literature. In reference to retrieval of information and drug-induced state, Eich (1977) states, "A person asked to remember a simple event such as the appearance of a familiar word in an otherwise unfamiliar list or collection of other words, typically shows impaired retrieval for the word-event when his state is changed between the study and test sessions of the experiment, in comparison with conditions where his state remains the same on both occasions" (p. 141). In a review of drugs that have been shown to produce memory effects Eich (1977) reported
that the effective dose of absolute ethanol was between 0.92 to 1.58 mls/kg of body weight. However, the relationship between memory disturbances and drug dosage has not yet been thoroughly researched.

Another perspective comes from research on the psychophysiology of attention and memory, which has shown that cognitive processing of stimulus information during the polygraph testing is an important determinant in the detection of deception. Corteen (1969) showed that when a list of words is presented to a subject, those that are remembered later are words which produced significantly larger electrodermal responses during presentation. Those that were forgotten tend to be words that produced the smaller electrodermal responses. Similar results have been obtained in studies by Sampson (1969) and McLean (1969). The assumption of these investigations is that the more intensely a subject focuses on the stimulus, the greater the electrodermal response will be and the more likely the subject will be to remember the word. These and other studies of arousal and memory suggest that subjects who do not fully process the semantic information of each question show reduced electrodermal responses and may be detected less frequently on a polygraph test. Factors that could possibly affect the processing of information are: cognitive countermeasures such as counting backwards by threes, drugs which would disturb the focus of attention during testing, and the salience of the test items. With regard to this factor of salience, it seems reasonable to suggest that more attention will be paid to items that have
greater meaning for the subject. This effect would be especially important in the Guilty Knowledge Test which depends on item recognition. Any condition that affects a person's memory for crime details, or the processing of these details during the crime, could affect the utility of those items in a polygraph test and reduce detectability.

Memory for crime details was looked at in several ways in the present study. After the polygraph test was given to each subject he was asked to complete three memory tests: one a free recall task often used in research on eyewitness memory of crime (see Yuille & Cutshall, 1986) which simply asks the subject to remember and note all the details he can, one a recall test which asks for specific information, and the third a recognition test in which the subject must choose the relevant item from a list of irrelevant items (see Appendix VII). The recognition test contained crime-relevant items not included in the Guilty Knowledge Test as well as those that were. In this way a correlation can be calculated between polygraph scores and memory for general crime details as well as between polygraph scores and memory for items specific to the test.

Study Rationale

One area of study that is important to a determination of polygraph validity is the issue concerning effects of drugs on polygraph outcome. Specifically, evidence has emerged suggesting that alcohol intoxication at the time of the crime has an effect on subsequent polygraph performance. Since the use of alcohol is widespread in this society (Ledain Commission Report, 1972) and
many crimes are committed under its influence (Glaser, 1978) any effect it might have that serves to decrease polygraph validity would be important.

Bradley and Ainsworth (1984) reported that alcohol intoxication during a mock crime can significantly decrease detectability by a polygraph test (both CQT and GKT) administered the following day. Although they found an alcohol effect, they could not explain why this had occurred. Since the mechanisms of physiological responding, which are the basis of the polygraph method, are poorly understood, the authors could only speculate that alcohol acted through memory or emotional processes. It seems reasonable to assume that either one or both of these processes are important in the detection of deception. And if Raskin (1979) is correct in his proposal that the CQT and GKT rely on characteristically different physiological responses to detect deception, memory for crime details and emotional responses during the crime could affect each test differently. The present study looked at the effects that emotion (threat) and memory for crime details have on the outcome of both polygraph methods. The present study was also an attempt to replicate the findings of Bradley and Ainsworth (1984) and to investigate how ethanol intoxication might affect emotional and memory processes as they relate to polygraph outcome.

Subjects recruited from undergraduate classes were administered both a Control Question Test and a Guilty Knowledge Test two days after they had committed a simulated crime. Procedures were similar to those of Bradley and Ainsworth (1984)
whenever possible. Conditions of threat, memory for crime details, and alcohol intoxication were manipulated in a fully factorial design. Guilty subjects committed the mock crime under conditions of high or low threat, then were primed or not for crime details, and half of the subjects in each of these groups were intoxicated while half were not. Some subjects came to the laboratory but did not commit the mock crime and were told none of the relevant details. They served as innocent control subjects during the polygraph testing, which all subjects underwent two days after committing the mock crime. The polygrapher was blind to group assignment of subjects at the time of testing, and after the study was completed the charts were scored blindly. All charts were numbered and all identifying information was removed. Besides the manipulation of threat and memory, physiological and subjective responses of the subjects were monitored at several important points during the study. As well, extensive memory testing was done following the polygraph session to determine what effects the various experimental manipulations had on recall and recognition of crime details. Outcome of polygraph testing was used to determine which factors (emotion during the crime, memory for crime details, and alcohol intoxication) were important to the physiological detection of deception.

The primary hypothesis being tested in this thesis concerns the effects of alcohol intoxication at the time of the crime: if a subject is intoxicated while committing a mock crime then he will be more difficult to detect as guilty on physiological tests of deception. A secondary hypothesis concerns the mechanism through which alcohol might act: if crime-intoxication does
reduce subsequent polygraph detectability then emotional and memory processes at the time of the crime will have been disrupted in the intoxicated subjects. Expectations based on current literature are that crime-intoxication will reduce detectability of deception by both the CQT and GKT procedures. The literature does not provide evidence to allow such a straightforward prediction concerning the mechanism of alcohol's action, but it is apparent that memory is an important variable in the GKT procedure and the CQT procedure appears to rely on a subject's emotional reaction to relevant questions. The expectation, therefore, is that memory disruptions due to intoxication would be important to the GKT procedure and emotional disruptions would be important to the CQT.
Method

Subjects

Eighty male university students were recruited from undergraduate courses at the University of British Columbia to participate in a study involving the detection of hidden information. The mean age of the subjects was 20.53 yrs with a range from 19 yrs (the legal drinking age) to 51 yrs. Most of the subjects were Caucasian, but the group also included some who were of Oriental descent. Before they volunteered, subjects were given a consent form to read and sign (see Appendix I).

The consent form told the subjects that this project was investigating the effects of alcohol on the detection of concealed information, and they would be asked to volunteer about 3 hours of their time over 2 days. On the first day, most subjects would be required to perform a task which involved taking money from a room, and they might also be asked to drink alcohol. All subjects were to return two days after the initial lab visit to undergo a polygraph examination. Since some subjects were required to drink alcohol, they had to agree to remain in the laboratory with a research assistant for one hour following the first session to allow some of the alcohol's effect to wear off. Subjects who had consumed alcohol were chauffeured home by a research assistant.

Subjects who were assigned to the alcohol group were also administered a medical questionnaire that asked for information about general health and consumption of drugs or alcohol (see Appendix II). Any subject who reported a major health problem
would be required to obtain a physician's written certification that he could participate in the study. There was a physician participating in the study who assisted in constructing the health questionnaire and who was on call in case any medical problems arose during the study. Also, subjects who were to be in the alcohol group were asked not to eat for 2 hours and not to consume alcohol for 24 hours before coming to the lab. In addition, they were asked not to drink coffee or smoke cigarettes 1 hour before coming for the polygraph test.

**Apparatus**

A Beckman R612 polygraph was used to record three physiological channels: skin conductance response (SCR), respiration, and finger pulse amplitude. SCR was recorded using Ag-AgCl electrodes attached to the medial phalanges on the first and second fingers of the left hand. The electrodes were filled with Unibase NaCl electrode paste. A Beckman Skin Conductance Coupler, type 9844, was used to process the signal. Respiration was measured with a single chest bellows positioned above the diaphragm. The bellows were attached through a strain gage to a Beckman Integrating Strain Gage Coupler, type 9825. Finger pulse amplitude was recorded with a photoelectric transducer attached to the middle finger of the right hand and covered with a black, foam finger sheath. The signal went to a Beckman Photocell Coupler, type 9874, and was recorded with the time constant set at .03 second. The 50% bandwidth for high frequency filtering was set at 30 Hertz for all three channels. Chart speed for all
subjects was 2.5 mm/sec. Blood pressure and heart rate was monitored throughout the study with a low-weight sphygmomanometer with pressure cuff, a manual pump, and a microprocessor with digital display (model UA 102, Thought Technology Inc., Montreal).

Study Design

The design of this study involved a total of 80 subjects: 64 had committed a simulated crime, and 16 had committed no crime serving as innocent controls (see Figure 1). The guilty subjects were subdivided so that half (32) were intoxicated with alcohol at the time of the crime, and half were sober. The alcohol and no alcohol groups were also divided, into a group of 16 subjects who committed the crime under conditions of high threat and a group of 16 who were under low threat conditions. Each of the threat groups had 8 subjects who read a list of details relevant to the crime, and 8 subjects who did not. These subjects served as the primed memory group and unprimed memory group respectively. All subjects returned two days after committing the crime (or just having their baseline measurements taken, in the case of innocent subjects) to be administered the polygraph tests. The two polygraph tests used were the traditional Control Question Test (CQT) and the Guilty Knowledge Test (GKT). The order of test administration was counterbalanced in each group so that half of the subjects received the CQT first, and half received the GKT first.

In summary, the design for the guilty subjects involved 2 levels of alcohol intoxication at the time of the crime (sober,
Figure 1
Study Design

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primed</td>
<td>n = 8</td>
<td>n = 8</td>
<td></td>
</tr>
<tr>
<td>High Threat</td>
<td></td>
<td></td>
<td>n = 16</td>
</tr>
<tr>
<td>UnPrimed</td>
<td>n = 8</td>
<td>n = 8</td>
<td></td>
</tr>
<tr>
<td>Primed</td>
<td>n = 8</td>
<td>n = 8</td>
<td></td>
</tr>
<tr>
<td>Low Threat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprimed</td>
<td>n = 8</td>
<td>n = 8</td>
<td></td>
</tr>
</tbody>
</table>

Note: Order of polygraph test administration was counterbalanced in each factorial cell
legal intoxication), 2 levels of emotional salience of the crime (high threat, low threat), and 2 levels of ability to recall crime details (primed, unprimed). As well, there were 2 orders of polygraph test administration (CQT first, GKT first).

Procedure

Before volunteering for the study, each subject was informed that he would be required to participate in two experimental sessions, two days apart. On the first day each subject was met and escorted into the lab by a research assistant. About five minutes was spent explaining the subject's part in the study, and answering any questions he might have. The research assistants were instructed to answer questions in a broad way without revealing the specific purpose of the experiment or any details that might compromise the outcome. The subject was then given the consent form to read and sign. Once the subject understood what was required of him, and had signed the consent form, he was asked to complete the first of three State Anxiety Inventories (Spielberger, Gorsuch, & Lushene, 1970). It was explained that the subject should be honest in answering the questionnaire, and that none of the information would be available to the polygrapher. Next, the subject's blood pressure and heart rate were recorded while he was seated. The above procedure was the same for all subjects, and only after the blood pressure measurements were made did the treatment of subjects begin to differ, depending on the experimental group to which they were randomly assigned.

Those subjects who were assigned to the alcohol intoxication
group were asked their body weight. The research assistant then determined the amount of ethanol needed to administer a dose of 1.32 mls per kg of body weight by referring to the alcohol-weight chart presented in Appendix III. The alcohol was served in three drinks made with a mixture of orange and pineapple juice. The subject had 15 minutes to finish each drink, so the total time taken to consume the alcohol dose was 45 minutes. Subjects were then asked to remain seated for 10 minutes to allow the alcohol to be absorbed into the bloodstream.

After finishing the drinks the subject was read either the High Threat or the Low Threat crime instructions (see Appendix IV). The instructions outlined a crime scenario that the subject would carry out. In the High Threat condition, the subject was told he had just gone to buy a psychology lab manual in his professor's office (an office set up specifically for this study). He has seen some money in a cashbox and now decides to steal it. The professor has left the office for a short time and the subject must go in, get the key to the cashbox from a coat pocket, remove the cashbox from the desk, steal the money, and return with it to the lab. The office number was given and the important steps necessary to accomplish the crime were outlined. The High Threat instructions stressed that the subject should be unobtrusive and very careful because he has no idea when the professor will return to the office. The subject was also told that people in the area of the office had not been informed of this study, and due to recent thefts from offices in the building, he had to be careful not to arouse suspicions. The
subject was given directions to the office and was required to go there by himself. Although the office is on the same floor as the lab, it is located on the opposite side of the building, so the subject had to walk down three different corridors to get to it. When the subject had understood the instructions, his blood pressure and heart rate were recorded for a second time, then he proceeded on his own to the crime office.

Once inside the office the High Threat subject encountered several crime features designed to increase his arousal. To begin, there was a cigarette burning in an ashtray on the desk. The intention was to give an appearance that someone had just left the office and was likely to return shortly. The subject passed the desk and went to get the cashbox key from a coat hanging on a coat rack. When he reached into the coat pocket a concealed trip wire caused some empty pop cans to fall off a ledge behind the coat. There were many cans stacked along the ledge, and the set up gave the appearance that the subject had been clumsy in handling the coat and had knocked the cans over himself. The idea was that the noise of the falling cans would startle the subject and perhaps increase his anxiety that someone would hear him. Finally, when the subject opened the desk drawer to get the cashbox he also found a bag of simulated marijuana. This was a final attempt to increase arousal by having the subject unexpectedly encounter something with high stimulus value.

When a subject in the Low Threat group had finished his last drink he was also read instructions, but these differed substantially from the High Threat instructions (see Appendix
IV). The Low Threat subject was told that we were interested in how the act of removing some money from a cashbox affects blood pressure and heart rate. The same outline for finding the money was given: the subject had to get a key from the coat pocket, retrieve the cashbox from the desk, unlock it, and take the money. There was no threat implied in the instructions, and the research assistants attempted to be as casual as possible in answering any questions. To keep the scenario low key, the subject was accompanied by a research assistant from the lab to the door of the crime room. While the research assistant waited outside, the subject went in and got the money. Unlike the High Threat condition, there was no burning cigarette, no trip wire to the pop cans, and no marijuana in the desk. Apart from the threatening instructions and arousal manipulations, the Low Threat subjects saw the same things and performed the same actions as the High Threat subjects did inside the crime room. The goal was to have the High and Low Threat crime situations differ only in the amount of arousal the subject experienced while taking the money.

The same High Threat and Low Threat scenarios were carried out with subjects in the No Alcohol group, so that there were 16 subjects in each of four groups: Alcohol/High Threat, Alcohol/Low Threat, No Alcohol/High Threat, No Alcohol/Low Threat.

After taking the money from the cashbox, all subjects, regardless of group assignment, were met outside the crime room and immediately asked to sit in a chair. Their blood pressure and heart rate were recorded for a third time and they completed the
State Anxiety Inventory for a second time. Each subject was then taken back to the lab. Those who were in a primed memory group were taken into a small room and read a paragraph that described 11 details of the crime room, 5 of which would later be used on the Guilty Knowledge Test (see Appendix VI). The subject was then quizzed on each item to make sure he had remembered all the details.

Subjects who were to serve as innocent controls came to the lab, read and signed the consent form, then had their blood pressure and heart rate recorded. Instead of having High or Low Threat instructions read to them, a brief explanation of the experimental procedure was given and they were told they would be functioning as control subjects, being innocent of the mock crime. No details of the crime scenario were given to them.

Before leaving the lab all subjects were given a reminder slip with an appointment time for the polygraph examination and directions to meet the polygrapher. Also, the purpose of the polygraph session was explained: the polygrapher was going to try and find out, by measuring physiological responses to questions, if the subject had removed money from the cashbox. The subject was to act as if he did not know anything about the missing money but to otherwise cooperate with the polygrapher. It was also explained that the polygrapher would not know if the subject was guilty because he was testing both innocent and guilty subjects and was blind to their group assignment. Finally, all subjects were asked not to discuss the details of the study with anyone else.

Two days after the mock crime, subjects returned to the
psychology building and met the polygrapher near the main entrance. The subject was taken to the polygraph room, located two floors up from the lab and crime room, and was seated in a straight-backed chair. The polygrapher then had the subject fill out the third State Anxiety Inventory, and recorded his blood pressure and heart rate for the fourth time. Following this, the polygrapher explained that the subject was suspected of taking some money from a professor's office, and the polygraph would be used to find out if the subject was guilty or innocent. The polygrapher reiterated that he did not know if the subject was guilty because both innocent and guilty subjects came for polygraph testing. While the subject was having the polygraph electrodes and sensors attached to him, a general explanation of polygraph principles was given. The subject was told that the polygraph was very good at detecting guilty subjects, and only people who were intelligent and emotionally stable would be able to lie and go undetected. This ploy was an attempt to increase the subject's arousal to polygraph questions, and is discussed more fully below.

Once all attachments were made, and the amplifier sensitivity in each channel was adjusted, a demonstration of the machine's operation was conducted. The subject picked a playing card from a deck of five and placed it face up on his knee. The polygrapher, standing behind the subject, asked if he had picked a given card in the deck and the subject, as instructed, answered "no". The polygrapher asked the question about each of the five cards in the deck, with the subject answering "no" each time.
This procedure forced the subject to lie about the card he had actually picked. Since the subject had the card face up on his knee the polygrapher knew its identity, and at the appropriate moment he adjusted the gain on the GSR channel to make the subject's response to the target card the largest of the five. The subject was subsequently shown the chart paper and the polygrapher pointed out that the subject's response to the target card was the largest, indicating he was being deceptive when he answered "no". The ostensible purpose of the card test was to demonstrate, for the subject, how the polygraph works, but the real intention was to make the testing situation more threatening than it might otherwise have been and to thereby enhance the subject's physiological responses. Gustafson and Orne (1965) and Bradley and Janisse (1981) demonstrated that if a subject believes he has responded emotionally (as reflected on the polygraph chart) to a relevant question, he is more likely to be aroused by subsequent questions. This arousal, coupled with the idea that detection of his guilt would indicate he was unintelligent or emotionally unstable, should have the effect of enhancing the subject's physiological responses to the polygraph questions.

Immediately following the card test, the first of the polygraph tests was administered, order of presentation being balanced in each experimental group. The Control Question Test consisted of 9 questions: 3 questions irrelevant to the crime, 3 control questions, and 3 relevant questions (see Appendix V). Prior to turning the polygraph on for the CQT, a 5-10 minute interview was conducted with the subject. He was told that the
test consisted of a set of 9 questions and the set would be asked three times. The questions were then reviewed. The main purpose of the interview was to discuss anything in the subject's past that might prevent him from answering "no" to the control questions. It is important that the subject answer in the negative to control questions so that a comparison can be made to the relevant questions, as discussed in the Introduction. If the subject did raise some ambiguity (e.g. to the control question: Have you ever stolen anything of value? he might say: I once stole a record from a friend.) the control question would be prefaced by the statement, "Except for what you have told me...." When the subject was satisfied he could answer "no" to all the control questions, the polygrapher turned on the polygraph and asked each of the questions in order. After the set of 9 questions was asked the polygraph was shut off and the subject was allowed to stretch and make himself comfortable before the next set of questions was asked. In each question set the order of presentation of the control and relevant questions was changed (see Appendix V).

The Guilty Knowledge Test questions were not reviewed prior to presentation. The subject was informed that if he had taken the money there were certain details about the crime that he would know. He would be asked specific questions about the room where the money was taken, and he was to answer "no" to each one. There were 5 GKT questions asked, each with 5 alternate responses (see Appendix V). Each of the 5 critical items was embedded in a set of similar items that were determined, in pilot testing, to
be equally plausible responses. The key item in each set was randomly assigned to one of the last four positions. The first item in each set served as a buffer question and was not considered in scoring the test.

Between the first and second polygraph test, each subject was disconnected from the machine and taken for a five minute walk through the building. Since the entire testing session took approximately 1 hour, during which the subject was required to sit very still, the walk was designed to give the subject a break and to decrease habituation to the testing situation.

After the final polygraph test the subject moved to a separate room and was asked to complete three memory tests (see Appendix VII). The first was a recall test that required the subject to list all the details he could remember about the crime room, including: descriptions and locations of furniture, decorations, and other objects in the room, as well as physical features of the room. When he had finished, the subject put his answer sheet in an envelope so the polygrapher would not see what he had written. Subjects were instructed to wait at least five minutes, even if they were innocent, before returning the envelope to the examiner. The subject then completed the cued recall test, in which answers to specific questions were required, such as: "What was the amount of money taken?" When the cued recall test was returned, the subject completed his final task, a recognition test in which he simply had to pick the correct answer from a choice of five alternatives. For each of these tests it was stressed that the subject should answer honestly and try to do his best, and that the polygrapher would
not see the memory test results until after the polygraph charts were scored. Subjects were also asked to make a guess on the recognition and cued recall tests if they were not sure of the answer or if they were innocent.

Once all the memory tests were completed, the polygrapher showed the subject his polygraph charts and explained whether he appeared innocent or guilty, indicating that the subject should not, even at this point, reveal what group he was in. Any questions the subject had were answered, and he was debriefed with regard to the card test. He was told that detection of guilt did not depend on the subject's intelligence or emotional stability, and that this was said at the beginning of testing merely to enhance his responses to the polygraph questions. Lastly the subject was thanked for his participation and asked once more not to discuss details of the study with anyone else.

**Polygraph Chart Analysis**

Three dependent measures were derived from the polygraph recordings: skin conductance response (SCR), respiration cycle response, and finger pulse amplitude. Scoring of these responses was done according to the Utah system and was taught to the polygrapher by Dr. David Raskin. The SCR was measured as the maximum increase in conductance (indicated as an upward pen deflection on the chart) occurring in the interval from the beginning of the question up to 5 seconds after the response. A respiration response was considered to be any shortening of the respiration cycle during the same interval described for the SCR.
A shortening of the respiration cycle could be manifested on the chart as an increase in cycle baseline, a decrease in amplitude, or a cessation of breathing (apnea). A cardiovascular response, in this case finger pulse amplitude, was considered to be any decrease in pulse amplitude during the time period used for SCR and respiration response. The decrease was calculated by taking the difference in mm between the largest pulse amplitude after the onset of the question, and the smallest pulse amplitude within 5 seconds following the subject's response. The expectation in using these three physiological measures is that a deceptive response by a subject will be indicated by an increase in skin conductance, a shortening of respiration (shallow breathing), and a decrease in finger pulse amplitude. A subject may respond in all these channels, one or two channels, or he may not respond in any.

Numerical scoring of the Control Question Test charts was done using the Utah method (see Raskin, 1986), and the Guilty Knowledge Test was scored following the procedure described by Lykken (1959). For each pair of control and relevant questions on the CQT the strength of each physiological response was compared, and a score was assigned for each comparison. There were 3 question pairs and 3 physiological measures, resulting in 9 scores for each chart. Three charts were recorded for each subject. The scores on each question pair can range from +3 if the physiological reaction to the control question is dramatically stronger than that of the relevant question, to -3 if the reaction to the relevant question is dramatically stronger. If the physiological reaction is substantial but not
dramatic, a score of +2 or -2 is assigned, but most scores are +1 or -1, indicating a small difference, or 0, indicating no difference between responses to the control and relevant questions. A total score for each subject is calculated by adding the scores for each question pair, in each physiological channel, across all three charts. Subjects whose scores are in the positive direction would be responding more strongly to the control questions, an indication of innocence. The converse is true for guilty subjects, they would score in the negative direction. In most scoring systems, total scores of +6 or higher are considered indicative of truthfulness, -6 or lower would indicate deception, and scores falling in between these two scores are considered inconclusive.

On the Guilty Knowledge Test, all 3 physiological channels in each of the 5 question sequences received a score of 2, 1, or 0, depending on the relative strength of the response to the critical item. If the response to the critical item was strongest, it received a score of 2, if the response was second strongest or equal in magnitude to the strongest response it received a score of 1, and any other response received a score of 0. There were 5 question sequences and 3 physiological measures, so the range of possible scores was 0 to 30. To make a determination of deceptive or truthful, the strategy outlined by Lykken (1960) was used. The cutoff score was set at 50% of the subject's possible SCR score. If a subject did not make a physiological response to any of the 5 alternative answers to a question, that question was excluded from the scoring. The test
was considered inconclusive if a subject did not respond at all to 3 or more of the 5 questions. Therefore, a subject who gave an SCR response to 4 questions would have a total possible score of 8 and would have to score 4 or higher to be considered deceptive. A score of less than 4 would indicate truthfulness. Any subject responding to 2 or less questions would be classified as inconclusive.
Results

In this section discussion of the results will begin with an analysis of the three independent variables to determine if all manipulations had the desired effects. This is followed by an evaluation of polygraph accuracy and a comparison is made to other studies that have used similar testing procedures. Next, the influences of alcohol intoxication, threat, and memory priming on polygraph outcome are investigated. Further investigation of the threat variable is conducted by looking at its two components; subjective and physiological arousal. Also, an evaluation of memory test scores is done and the relation of memory to polygraph scores is investigated.

To control for any liberal bias in the repeated measures ANOVAs following, the Greenhouse-Geisser (1959) correction was used. This correction procedure adjusts the degrees of freedom of the F-test based on the homogeneity of both the variances and covariances of the repeated measures. When there is reason to question the homogeneity assumptions a more conservative F-test is used. The degrees of freedom are adjusted by a factor ($\varepsilon$) that reflects the degree of heterogeneity of the repeated measures variances and covariances. As well, since most of the chi square analyses will contain some small expected frequencies (ie. $<5$) calculations were made using a formula that includes the Yates' correction for continuity. For small values of $N$ the curve of the sampling distribution of chi square may provide a poor fit to the data which could result in appreciable error when estimating probabilities. The Yates' correction reduces by $.5$ the obtained
frequencies that are greater than expectation and increases by .5 those frequencies that are less than expectation. This correction should always be used when 20% or more of the expected cell frequencies are less than 5 and some statisticians recommend a cutoff frequency of 10 (Ferguson, 1976).

In the course of evaluating these data, multiple ANOVAs and chi square analyses were often carried out for the sake of complete investigation of interesting variables. When this situation arose Type I error rate was controlled by reducing the alpha level, according to the Bonferroni approach (cf. Harris, 1975), in both the initial analyses and the post hoc tests. In order not to lose excessive power in the tests of significance the initial alpha level was set at a liberal 0.1 rather than the usual .05, and it was reduced by the appropriate steps to a level consistent with the number of analyses performed. Wherever this was done the Bonferroni alpha level is indicated with the F values.

One way to control the alpha level is to do it in an experiment-wise manner, or in other words, consider all analyses done in the study. This was not done because in an investigation of this size the estimations of significance would be too conservative. Instead, the Bonferroni inequality principle was applied in a family-wise manner. That is, the alpha level was controlled for theoretically related sets of analyses. An example of this would be the case where univariate analyses were done with each of the three GKT dependent measures and each of the three CQT dependent measures. Instead of adjusting the alpha level for six univariate analyses the CQT and GKT were treated as
theoretically distinct cases and the alpha level was adjusted for three analyses.

The model of MANOVA used to evaluate the dependent variables in this section was the same for all cases where it was applied. The model is a fully factorial, between subjects design with no covariates, and three dependent variables. This design provides for analysis of all the main effects as well as interactions of the independent variables. Whenever the results of a MANOVA are reported the levels of independent variables are indicated eg. 3-way MANOVA (2 levels of threat x 2 levels of alcohol x 2 levels of priming). Given that there were 3 dependent variables and that the main effects were calculated with an N=32, and the two-way interactions with an N=16, the power of the analyses is sufficient. However, the three-way interactions were necessarily calculated with a smaller number of subjects (N=8) and the power of these analyses is consequently weaker.

Independent Variable Manipulations

To ensure that the independent variables (alcohol intoxication, memory priming, and threat) were appropriate and adequate treatments, several manipulation checks were carried out. Those subjects in the alcohol intoxication group were administered a 1.32 ml/kg of body weight dose of ethanol, as described in the method section. The objective was to raise this group's blood alcohol concentration above the legally defined level of intoxication; 80 mg ethanol/100 mls of blood (commonly expressed as .08). In a similar study conducted in this lab a few
months earlier, officers from the local Royal Canadian Mounted Police detachment were asked to give a breathalizer test to twelve subjects drawn from the same subject pool used in this study: male undergraduates. Ten minutes after having consumed 3 drinks containing a total of 1.32 ml/kg ethanol, each subject was given a breathalizer test. The average blood-alcohol level was 97 mg/100 mls with a range of 61 to 115 mg/100 mls. Of the twelve subjects, only two were below, but close to, the legal limit. Their readings were 61 and 77 mg/100 mls. These blood-alcohol readings contrast with those reported by Bradley and Ainsworth (1984). They gave subjects less alcohol (1.0 ml/kg body weight) over a similar time period (1 hr) and they indicate that the blood alcohol level of their subjects approached 120 mg/100 mls, a somewhat higher level than that achieved in this study, but there is no indication that they conducted a breathalizer check, and no explanation of how they arrived at the .12 estimate.

Memory for crime details was manipulated in two groups: one group was primed following the crime and the other was not. Following the polygraph session three tests of memory were administered, two of which are relevant to the issue of memory for specific crime details: a cued recall test and a recognition test. A 3-way ANOVA (2 levels of alcohol x 2 levels of threat x 2 levels of priming) was performed with scores on the recognition test. There was found to be a significant effect of priming \[ F(1,56)=293.42, \ p<.001 \]. The primed group had a mean test score of 10.31 out of a possible 11, compared to a score of 6.28 for the unprimed group. There were no significant main effects of alcohol or threat on recognition test scores, nor were there any
significant interactions between memory priming, threat, or alcohol.

The same 3-way ANOVA was conducted with the cued recall test scores. The ANOVA showed a similar difference between the primed and unprimed groups, with the mean of the primed group (19.56 out of a possible 22) being significantly higher than the unprimed group (10.47) [F(1,56)=355.16, p<.001]. In contrast to the recognition test score analyses, alcohol intoxication had a significant effect on cued recall scores [F(1,56)=28.89, p<.01]. Although this alcohol main effect was statistically significant the mean difference between the two groups was not substantial, with the mean score of the alcohol intoxicated subjects (14.34) being only slightly less than that of the sober subjects (15.69). This alcohol effect will be discussed in more detail later in conjunction with polygraph test outcome.

To determine if manipulation of the threat variable had been successful, both physiological and subjective measures of arousal were taken at various points during the study. The physiological measures were cardiovascular responses (heart rate, systolic and diastolic blood pressure) and the subjective measure was the State portion of the State-Trait Anxiety Inventory (Spielberger, Gorsuch, and Luchene, 1970). Although the postcrime measures of arousal were made immediately after the subject left the crime room, there was a few minutes delay before all readings were completed. Due to this delay, and the fact that the subjects were young, healthy students whose physiological responses recover quickly, we expected that the best physiological indicator of arousal would be systolic blood pressure, since it is responsive
to changes in arousal and is sustained longer than heart rate changes. This turned out to be the case as a 3-way (2 levels of alcohol x 2 levels of threat x 4 time periods) repeated measures ANOVA indicated a significant group difference in systolic blood pressure across time \( [F(3,171)=4.54, p<.025] \). Simple effects analysis of mean scores showed the mean systolic pressure of the high threat group to be significantly higher than the mean of the low threat group immediately after the crime \( [F(1,59)=7.48, p<.025] \) but not significantly different at any other time (see Figure 2). There was also a significant interaction of alcohol and time \( [F(3,171)=3.56, p<.025] \). Simple effects analysis of the means showed that the alcohol group had lower mean systolic pressure shortly after drinking, likely due to the vasodilation properties of ethanol, but this difference was not significant when the Bonferroni adjustment of alpha was made \( [F(1,59)=6.75, p>.008] \). See Figure 3. There were no significant group differences across time in either diastolic blood pressure \( [F(3,171)=0.68, p>.05] \) or heart rate \( [F(3,171)=2.64, p>.05] \) nor were there any significant interactions with alcohol. Blood pressure measurements of the innocent subjects were taken only at baseline and before the polygraph session, but the mean pressures of 127.62 mm/Hg at baseline and 138.56 mm/Hg before testing are consistent with the mean pressures of the other subjects.

The subjective, state anxiety measures were taken at three points in the study: baseline, post crime, and the beginning of the polygraph session. A 3-way (2 levels of alcohol x 2 levels of threat x 3 time periods) repeated measures ANOVA, using STAI scores as the dependent measure, showed no main effect due to
FIGURE 2

MEAN SYSTOLIC PRESSURE FOR HIGH AND LOW THREAT GROUPS

Legend

- HIGH THREAT
- LOW THREAT

SYSTOLIC BLOOD PRESSURE IN MM HG

TIME OF MEASUREMENT

BASELINE POST INSTRUCTION POST CRIME POLYGRAPH SESSION
FIGURE 3

MEAN SYSTOLIC PRESSURE FOR ALCOHOL AND NO ALCOHOL GROUPS

Legend

▲ ALCOHOL

★ NO ALCOHOL
alcohol \( [F(1,59)=3.13, p>.05] \) but there was a significant main effect of threat \( [F(1,59)=7.34, p<.05] \), the high threat group having a higher mean STAI score (42.0) than the low threat group (37.8). There was also a significant interaction between threat and time \( [F=(2.118)=21.88, p<.025] \). Simple effects analysis showed a significant group difference between means only at the post crime measurement \( [F(1,61)=6.10, p>.025] \). See Figure 4. The mean STAI scores of the innocent group at baseline (36.26) and at the polygraph session (35.00) were of similar magnitude as those of the low and high threat groups.

In summary, manipulations of the three independent variables appear to have resulted in significant and substantial differences between the groups. Subjects in the high threat group scored higher on both a physiological and a subjective measure of arousal taken immediately following the crime. Tests of memory for specific crime details showed that subjects in the primed memory group remembered substantially more details than those in the unprimed group, in both recognition and cued recall of items. Finally, breathalyzer tests done prior to this study suggest that the mean blood-alcohol level of our subjects would be in the range of 97 mg/100 mls, well above the legally defined intoxication point of 80 mg/100 mls.

**Accuracy of Polygraph Tests**

The purpose of this study was to investigate factors at the time of a crime that might affect subsequent polygraph test outcome. It was not designed to answer broader questions about the validity of polygraph tests, and that sort of discussion
FIGURE 4

MEAN STAI SCORES FOR HIGH AND LOW THREAT GROUPS

Legend

△ HIGH THREAT

× LOW THREAT
would be inappropriate. However, before proceeding with an analysis of the experimental manipulations some comparisons of polygraph accuracy in this study to that in other studies will be made to be sure that the procedures were carried out well enough to support the proposed hypothesis testing.

Using an inconclusive range of +5 to -5 as described in the Method section, an assessment of deceptive or truthful was made for each subject based on the Control Question Test results. Of those subjects classified as deceptive or truthful 73.5% were correctly identified as deceptive (true positives) and 26.5% were wrongly identified as truthful (false negatives). However, 46.9% of the subjects in the crime condition fell in the inconclusive range and could not be classified as either deceptive or truthful. Of the innocent control subjects, 83% of those classified were correctly identified as truthful (true negatives) and 17% were classified as deceptive (false positives) with 62.5% falling in the inconclusive range. Taking into account both guilty and innocent subjects, the overall rate of correct classification was 78.2%. This compares with a survey of CQT field studies compiled by Iacono and Patrick (1987) in which they compared the results of 8 studies done between 1969 and 1984. The overall mean accuracy rate in these studies was 81.6% correct decisions, with a range of 64% to 93%. Raskin (1986) reported 5 mock crime studies of the control question test that used non-student subjects. These studies were conducted mainly by Raskin and his associates with field polygraph techniques and trained polygraph examiners. The mean accuracy of decisions in these studies was 95% correct classification, with a range of 91% to
96%. The inconclusive rate was 8%.

A post hoc manipulation of the inconclusive region, using the total score from all three charts, was done for cutoff points ranging from ±1 to ±5. It is evident from Table 1 that the hit rate, that is the percent of accurate decisions, does not change substantially even when the cutoff scores are decreased to ±1. At the same time the number of subjects being classified as inconclusive is greatly reduced. Consistent with the studies of Bradley and Ainsworth (1984) and Raskin and Hare (1978), cutoff points of +2 and -2 were found to yield good accuracy of decisions while keeping the number of inconclusives low.

In addition to using the method of cutoff scores to determine guilt or innocence, a multivariate, discriminant function analysis was also conducted. A discriminant function optimally weights all dependent variables so that maximal discrimination between the two target groups is achieved. The analysis provides a test of the null hypothesis that the two groups are identical in terms of the dependent measures. In this case the interest was in seeing how well the innocent subjects could be distinguished from the guilty on the basis of numerical scores without recourse to standard cutoff scores which weight the three physiological measures equally. Since a discriminant analysis can be biased by unequal sample sizes, it was important that the ratio of innocent to guilty subjects be equal, or in other words, since there are only 16 innocent subjects in the study only 16 guilty subjects could be included in the analysis. The guilty group used was the high threat/no alcohol group since the treatment of this group most closely corresponds to that of
### Table 1

*Classification Accuracy for Various Cutoff Limits*

<table>
<thead>
<tr>
<th>Cutoff Limits</th>
<th>±5</th>
<th>±4</th>
<th>±3</th>
<th>±2</th>
<th>±1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceptive</td>
<td>25 (73%)</td>
<td>28 (76%)</td>
<td>32 (74%)</td>
<td>37 (71%)</td>
<td>38 (68%)</td>
</tr>
<tr>
<td>Guilty Subjects</td>
<td>9 (27%)</td>
<td>9 (24%)</td>
<td>11 (26%)</td>
<td>15 (29%)</td>
<td>18 (32%)</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>30 (47%)</td>
<td>27 (42%)</td>
<td>21 (33%)</td>
<td>12 (19%)</td>
<td>8 (12%)</td>
</tr>
<tr>
<td>Truthful</td>
<td>5 (83%)</td>
<td>7 (78%)</td>
<td>8 (80%)</td>
<td>8 (67%)</td>
<td>10 (71%)</td>
</tr>
<tr>
<td>Innocent Subjects</td>
<td>10 (63%)</td>
<td>7 (44%)</td>
<td>6 (37%)</td>
<td>4 (25%)</td>
<td>2 (12%)</td>
</tr>
</tbody>
</table>
other laboratory studies and the field situation.

A discriminant function analysis using the innocent group and the high threat/no alcohol group of guilty subjects, with the dependent variables being numerical scores derived from the three physiological measures of the Control Question Test, showed that guilty subjects were correctly classified in 75.0% of the cases, and the innocent subjects were correctly classified 68.8% of the time, making the false positive rate 31.2%. A chi square analysis of the discriminant functions showed group separation to be significant \( \chi^2(3) = 9.30, p < .05 \). The results of this analysis, which maximally discriminates between the two groups, compare with the true positive rate of 71.1%, a true negative rate of 66.6%, and a false positive rate of 33.4% resulting from the categorical scoring system with cutoff points of +2 and -2. This suggests that a cutoff region of ±2 yields near optimal accuracy of classification while substantially decreasing the number of subjects in the inconclusive category.

Unlike the CQT, the Guilty Knowledge Test has not been used in field studies and there is no evidence that the GKT is used to any extent by field polygraphers (Iacono and Patrick, 1987). For the purposes of comparison then, estimates of GKT accuracy derived from laboratory studies must be used. In a review of studies that used mock crime investigations, carried out between 1959 and 1984, Iacono and Patrick (1987) reported that GKT accuracy rates for guilty subjects ranged from 59% to 92%. The mean classification rate in these seven studies was 80% for guilty subjects and 97% for innocent subjects. Since the GKT is not widely used there is no well established procedure of
determining cutoff scores for guilt or innocence, but as discussed in the Method section, the strategy used in this study is the one outlined by Lykken (1960).

With the cutoff score being 50% of a subject's total possible score, the accuracy rate for the three groups was as follows: 26 subjects in the primed memory group were classified as deceptive (86.6%), 21 of the unprimed group were classified as deceptive (72.4%), and 15 of the innocent group were classified as truthful (97.7%) making the overall hit rate 82.6% with a total of 5 subjects (6.2%) falling in the inconclusive range. The one innocent subject who was classified as deceptive had a SCR score of 7 out of a possible 10 and appeared to have knowledge of the crime details since he responded strongly and clearly only to the relevant items. It is interesting to note also that his score of -8 on the CQT also put him clearly in the deceptive category. I attempted to contact him after the results had been tabulated but failed to reach him so his case remains a mystery.

The primed memory group subjects are probably most representative of the subjects used in laboratory studies of the GKT, since many studies have the subject memorize relevant information or use well learned material (ie. mother's name) as the target for polygraph detection. An aggregate of the primed and unprimed groups, which includes people who remember all relevant crime details and those who remember only some, may be more representative of a field situation. In this aggregate group of guilty subjects, 47 of 59 or 79.7% were correctly classified as deceptive on the GKT.

The same discriminant function analysis conducted with the
CQT data was performed using the three physiological measures derived from the Guilty Knowledge Test. In this analysis, guilty subjects were correctly classified as guilty in 75.0% of the cases and innocent subjects were correctly classified in 93.8% of the cases, meaning there was a false positive rate of 6.2%. A chi square analysis of the discriminant functions showed that the separation of the groups was significant \( \chi^2(3) = 9.8, p < .05 \).

Since some researchers use only the SCR measure when conducting a GKT interrogation (cf. Lykken, 1960; Iacono, Boisvenu, & Fleming, 1984) it is also of interest to see how well the innocent and guilty groups can be separated on the basis of this measure alone. A discriminant function analysis using just SCR channel of the GKT correctly classified 75.0% of the guilty subjects and 68.8% of the innocent subjects, giving a false positive rate of 31.3%. The hit rate with guilty subjects is the same in both discriminant functions, but using only the SCR measure increases the rate of false positives from 6.2% to 31.3%. However, neither of these analyses classifies subjects as well as Lykken's (1960) categorical system which, as outlined earlier, correctly identified 79.7% of the guilty subjects and 97.7% of the innocent subjects, making the false positive rate only 3.3%. The difference in hit rates of the discriminant function and Lykken's categorical system may lie in the fact that the categorical system, in assigning a subject, uses only the GKT questions to which the subject responds, whereas the discriminant function uses the total score, obscuring the fact that some subjects do not respond at all to certain questions.

When a discriminant function analysis was performed using
both the GKT and CQT data (i.e. 6 dependent variables) the accuracy of classification was increased markedly. In this combined score analysis, guilty subjects were correctly classified in 81.3% of the cases, and innocent subjects in 93.7% of the cases, with a false positive rate of only 6.3%.

In summary, it appears that the accuracy of the CQT and GKT procedures made in this study is comparable to accuracy achieved in studies reviewed by Iacono and Patrick (1987), but the CQT classification rate was somewhat lower than that achieved in several laboratory studies done by Raskin and his associates (see Raskin, 1986). One difficulty in using the CQT in this study was the limitation of having to run only 3 charts per subject. In a field situation the polygrapher would run more charts if, after 3 charts, the suspect scored in the inconclusive range. In this study, however, the design called for two different test procedures (CQT and GKT) which were counter-balanced for order of presentation, and in the interest of consistency the two tests had to be kept approximately the same length. Using 3 charts in the CQT procedure meant the subject was asked 27 questions, making it about the same length as the GKT which had 25 questions. Running more charts for those subjects in the inconclusive range would have meant increased and unbalanced testing time. As well, the polygrapher was blind to the subject's experimental condition and did not score the charts until all subjects had been run.

Although the inconclusive rate was high when the traditional CQT cutoff limits of +5 to -5 were used, a discriminant function analysis suggested that near optimal separation of the innocent
and guilty groups could be achieved when an inconclusive range of +2 to -2 was used. The ±2 cutoff is optimal in terms of maintaining a high hit rate among guilty subjects while minimizing the number of those falling in the inconclusive category. This is also the inconclusive range used by Bradley and Ainsworth (1984) in the initial study of alcohol effects on polygraph outcome. Considering the above data, the accuracy of both the Control Question Test and the Guilty Knowledge Test appears to be well within the range of accuracy achieved by other researchers using similar procedures.

Reliability of the CQT and GKT scoring procedures was examined and shown to be high. Fifteen of the polygraph charts were randomly chosen and blindly rescored by the polygrapher 6 months after the initial scoring. Pearson correlations of the two scores were calculated for each channel of the CQT and GKT charts. The SCR scores of the CQT correlated highest (.99) followed by the finger pulse amplitude scores (.98) and respiration (.78). The correlation between total CQT score and rescore was .99. The pattern of correlations was the same for the GKT procedure with SCR scores correlating highest (.99) followed by finger pulse amplitude (.86) and respiration (.83). The correlation of the total GKT scores was .94. On the basis of the above correlations it appears that this numerical scoring system is reliable.

Effect of Independent Variables on Polygraph Outcome

As described in the Method section, the design of this study was fully factorial with three independent variables; threat,
alcohol intoxication, and memory priming, and all groups contained an equal number of subjects. The analysis of the effects of these three variables was done in two ways; multivariate analysis of the numerical scores and nonparametric analysis of the classification rates. Outcome of both the GKT and CQT procedures will be investigated in this manner.

To determine if any of the three independent variables had a significant effect on the Control Question Test results, a fully factorial, 3-way (2 levels of threat x 2 levels of memory priming x 2 levels of alcohol intoxication) MANOVA was performed using the numerical scores derived from the three dependent measures (SCR, respiration, and finger pulse amplitude). Results showed that the mean CQT scores of subjects were unaffected by the manipulations; neither threat \[F(3,54)=1.89, \ p>.05\] nor memory priming \[F(3,54)=0.68, \ p>.05\] nor alcohol \[F(3,54)=0.18, \ p>.05\] produced significant group differences in polygraph scores. Similarly, none of the 2-way interactions or the 3-way interaction were significant. Table 2 summarizes the CQT total scores for each of the factorial cells.

Although multivariate analysis showed no mean numerical score differences between groups, there was also accuracy of classifying subjects into deceptive or truthful categories to be considered. Following the post facto procedure of manipulating cutoff boundaries that was discussed above and outlined by Raskin and Hare (1978), and using the optimum inconclusive range of +2 to -2, outcome of the polygraph test was assessed based on total scores of all three charts. Table 3 shows the results of
<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Threat Primed</td>
<td>-5.38</td>
<td>-6.75</td>
<td></td>
</tr>
<tr>
<td>High Threat UnPrimed</td>
<td>0.00</td>
<td>-3.63</td>
<td></td>
</tr>
<tr>
<td>Low Threat Primed</td>
<td>-4.88</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Low Threat UnPrimed</td>
<td>-0.38</td>
<td>-6.25</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Alcohol</td>
<td>No Alcohol</td>
<td>Innocent</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>High Threat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Truthful</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>UnPrimed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Truthful</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Low Threat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Truthful</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>UnPrimed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Truthful</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
classifying all the guilty subjects on the basis of CQT scores. Chi square tests of independence were conducted to determine if alcohol, threat, or memory priming had affected the distribution of classifications. Again, there were no significant effects found; neither alcohol \( \chi^2(1) = 0.94, p > .05 \) nor threat \( \chi^2(1) = 0.03, p > .05 \) nor memory priming \( \chi^2(1) = 1.09, p > .05 \) affected subject classification.

The results of the Guilty Knowledge Test were subjected to the same analysis conducted above with the CQT results. A 3-way (2 levels of threat x 2 levels of memory priming x 2 levels of alcohol intoxication) MANOVA, using the numerical data derived from each of the GKT physiological channels, indicated that there was no significant difference in the mean scores of the two threat groups \( F(3,54) = 0.55, p > .05 \) or in the mean alcohol group scores \( F(3,54) = 0.016, p > .05 \). However, in contrast to the CQT analysis, there was a significant difference between the memory priming groups. The mean score for the primed memory group (11.75) was significantly higher than the mean of the unprimed group (8.66), \( F(3,54) = 3.05, p < .05 \). None of the 2-way interactions or the 3-way interaction were significant. It is interesting to note that the mean GKT score of the unprimed group (8.66), although smaller than that of the primed group, is still substantially higher than the mean score of the innocent group (5.00). Univariate tests of the individual physiological measures revealed that the mean SCR score of the primed group (6.87) was significantly higher than that of the unprimed group (4.75), \( F(1,56) = 8.20, p < .01 \). The finger pulse amplitude \( F(1,56) = 0.073, p > .05 \) and respiration \( F(1,56) = 1.52, p > .05 \) scores did not
differ significantly between the two groups. Table 4 summarizes the GKT total scores for each of the factorial cells.

As with the CQT method, subjects can be classified as truthful, deceptive, or inconclusive, based on the GKT score. As described in the method section, to be classified as guilty the subject had to have an SCR score that was at least 50% of the total possible score. If his score was less than 50% of the total possible, he was classified innocent, and if he did not give a response in at least 3 of the 5 GKT questions his chart was considered inconclusive. The hit rate based on this classification system was analyzed to determine if alcohol, threat, or memory priming variables would have a significant effect. Chi square tests of independence showed that none of the independent variables, neither alcohol ($\chi^2(1)=0.07, p>.05$), or threat ($\chi^2(1)=0.82, p>.05$) or memory priming ($\chi^2(1)=1.07, p>.05$) had a significant effect on the outcome of classification. It appears that although memory priming produced a significant group difference in mean numerical scores, it did not affect the final judgment of truthful or deceptive. Table 5 shows the results of classifying subjects on the basis of the SCR data only.

To further assess the relationship between memory for crime details and GKT score, Pearson product moment correlations between SCR score and the recognition and cued recall memory scores were calculated. Recognition test score correlated .53 while cued recall test score correlated .54 with the SCR score of the GKT.

Post hoc Analysis of Effects due to Arousal

In most experimental situations there are subjects who
Table 4

**Cell Means - GKT Composite Score**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primed</td>
<td>11.25</td>
<td>12.25</td>
<td></td>
</tr>
<tr>
<td>High Threat</td>
<td></td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td>UnPrimed</td>
<td>9.50</td>
<td>6.88</td>
<td></td>
</tr>
<tr>
<td>Primed</td>
<td>12.00</td>
<td>11.50</td>
<td></td>
</tr>
<tr>
<td>Low Threat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UnPrimed</td>
<td>9.63</td>
<td>8.63</td>
<td></td>
</tr>
</tbody>
</table>
Table 5

GKT Hit Rates - Based on SCR scores only

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deceptive</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Primed</td>
<td>Truthful</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Inconclusive</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>High Threat</td>
<td>Deceptive</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Truthful</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Inconclusive</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Deceptive</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Primed</td>
<td>Truthful</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Inconclusive</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Low Threat</td>
<td>Deceptive</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Truthful</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Inconclusive</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
respond better to the independent variable manipulations than others. In an attempt to compensate for any subjects who may not have been aroused by the high threat manipulation or, conversely, may have been aroused in the low threat condition, it was decided to regroup subjects on the basis of their physiological and subjective arousal at the time of the crime. Rachman (1978) outlined a three-systems analysis of fear comprising three components: subjective experience of apprehension, psychophysiological changes, and attempts to avoid or escape certain situations. In this study the subjects were directed to commit the crime, so avoidance as a measure of threat was not feasible; therefore we have to rely on physiological and subjective arousal to estimate the subject's experience of threat. As well, Rachman (1978) hypothesized that in certain situations there would be a discordance between the three systems, a person could display high physiological arousal and low subjective arousal, or vice versa. Discordance was in fact evident in this study as the postcrime STAI scores and systolic blood pressure were not highly correlated (Pearson r = .26).

Given that the physiological and subjective components could reflect a different aspect of threat, it was decided to assess each independently. The approach was to do a median split within the alcohol and no alcohol groups so that subjects who scored highest on the post-crime STAI, regardless of their threat group assignment, were put into a new group: high subjective arousal. Those subjects below the median were assigned to the low subjective arousal group. The same procedure was done with the post-crime systolic blood pressure measure, resulting in a group
of subjects who showed high physiological arousal immediately following the crime and a group with low physiological arousal. The STAI score and systolic blood pressure were used in the median splits because they proved to be the measures that best discriminated between the high threat and low threat groups as discussed earlier. When the alcohol group was split on the basis of STAI scores, 4 subjects previously in the high threat condition moved to the low subjective arousal condition, being replaced by 4 subjects previously in the low threat condition, and 5 subjects changed groups in the no alcohol group. The split on the basis of systolic blood pressure resulted in 6 subjects in the alcohol group moving from the high threat condition to the low physiological arousal condition, and 4 subjects in the no alcohol group changing conditions.

Statistical analyses similar to those performed on the original threat groups were conducted on the new groups created by the median splits. Since the splits were done without regard to memory priming, the multivariate analysis was a 2-way (2 levels of arousal x 2 levels of alcohol intoxication) MANOVA. Using the data derived from the three physiological channels of the GKT, the MANOVA revealed no significant mean differences between the high subjective arousal and low subjective arousal groups \(F(3,58)=1.44, p>.05\) and the alcohol/subjective arousal interaction was also not significant \(F(3,58)=1.01, p>.05\). Similar results were obtained when GKT data of the high and low physiological arousal groups were examined. The 2-way MANOVA suggested that there was no significant difference between GKT scores of these groups \(F(3,55)=1.40, p>.05\) nor was there a
significant interaction of alcohol with physiological arousal \(F(3,55)=0.28, p>.05\). Since the GKT classification procedure used in this study depended on only the electrodermal measure, a separate analysis of the SCR scores was also conducted. A one-way ANOVA showed the SCR measure to be significantly different between groups \(F(1,60)=4.33, p<.05\), the high subjective arousal group had a higher mean score (6.59) than the low arousal group (5.03). This difference is important to note because unlike the CQT, the GKT categorical scores were based solely on numerical scores derived from the SCR channel and only changes in this measure can affect classification of subjects. Table 6 summarizes the total GKT scores as well as the SCR scores for each of the multivariate factorial cells.

In addition to multivariate analysis, nonparametric analyses of GKT classification rates were done. When the accuracy of classifying subjects as deceptive or truthful in the high subjective arousal group was compared to the accuracy of classifying subjects in the low subjective arousal group, a significant difference in the 2 x 2 contingency table was found \(X^2(1)=6.77, p<.016\) with subjects in the low subjective arousal group being harder to detect as guilty (see Table 7). A check of the distribution of subjects in the memory priming condition showed that the median split resulted in a high subjective arousal group containing 15 subjects whose memory had been primed for crime details and a low subjective arousal group containing 17 primed subjects, therefore the significance of the chi square is not due to an inadvertent unbalancing of groups with respect to the priming variable.
Table 6

**GKT Numerical Scores - Subjective Arousal Groups**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>11.37</td>
<td>11.25</td>
</tr>
<tr>
<td>High Subjective Arousal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR</td>
<td>6.75</td>
<td>6.43</td>
</tr>
<tr>
<td>Low Subjective Arousal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composite</td>
<td>9.81</td>
<td>8.37</td>
</tr>
<tr>
<td>SCR</td>
<td>5.31</td>
<td>4.75</td>
</tr>
</tbody>
</table>
### Table 7

**GKT Categorical Scores - Subjective Arousal Groups (SCR only)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Subjective Arousal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Truthful</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Low Subjective Arousal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Truthful</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
The analyses done above with the high and low subjective arousal groups were repeated for the high and low physiological groups. A 2-way MANOVA using numerical scores from all three physiological channels of the GKT showed no significant mean difference between the high and low physiological arousal groups $[F(3,55)=1.40, p>.05]$ nor a significant alcohol/physiological arousal interaction $[F(3,55)=0.28, p>.05]$. A chi square test of classification accuracy similarly showed that there was no significant difference in detection rates between the two groups $[X^2(1)=0.04, p>.05]$.  

The Control Question Test data of the two median split groups was evaluated in the same way as the GKT data. A 2-way MANOVA (2 levels of subjective arousal x 2 levels of alcohol intoxication) showed that there was no significant difference in mean CQT scores between the two groups formed by a median split of the post-crime STAI scores $[F(3,58)=0.85, p>.05]$. Table 8 shows the numerical scores for each of the factorial cells. Although there were no significant main effects, the alcohol by subjective arousal interaction was significant $[F(3,58)=4.59, p<.01]$. Univariate analysis of the individual physiological measures revealed that both respiration $[F(1,60)=6.20, p<.016]$ and skin conductance response $[F(1,60)=8.96, p<.016]$ scores were significantly different in the alcohol interaction. Figures 5 and 6 show that alcohol interacts with the condition of high subjective arousal to produce positive SCR and respiration scores, meaning that subjects who are both subjectively aroused and intoxicated during the crime tend to score more in the
Table 8

CQT Cell Means - Subjective Arousal Groups

<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Subjective Arousal</td>
<td>1.00</td>
<td>-5.50</td>
</tr>
<tr>
<td>Low Subjective Arousal</td>
<td>-6.31</td>
<td>-2.31</td>
</tr>
</tbody>
</table>
FIGURE 5

MEAN SCR SCORES FOR SUBJECTIVE AROUSAL × ALCOHOL INTERACTION

Legend

- HIGH AROUSAL
- LOW AROUSAL
FIGURE 6

MEAN RESPIRATION SCORES FOR SUBJECTIVE AROUSAL x ALCOHOL INTERACTION

Legend
• HIGH AROUSAL
• LOW AROUSAL
truthful direction.

Analysis of classification rates showed that the numerical differences between groups, which suggest that high arousal subjects will be less easily detected when intoxicated, were large enough to significantly affect subject classification with the CQT procedure. Chi square analysis of hit rates (2 levels of classification x 2 levels of arousal) showed that there was a significant difference between high arousal and low arousal subjects when they were intoxicated \[X^2(1) = 4.49, p<.05\]. More subjects in the high arousal group were misclassified as truthful than in the low arousal group (see Table 9). The same chi square analysis was conducted with the no alcohol group but there was no significant classification difference \[X^2(1) = 0.20, p>.05\] between high arousal and low arousal subjects (see Table 9). The final nonparametric analysis conducted with these groups was between crime-sober subjects and crime-intoxicated subjects in the high arousal group. There was no significant difference found between these groups \[X^2(1) = 2.46, p>.05\].

When CQT data of the two groups formed by a median split of the post crime systolic blood pressure measure were examined, there were no differences found. The 2-way MANOVA showed no significant difference between the mean numerical scores of the high physiological arousal group and the low arousal group \[F(3,55) = 0.13, p>.05\] and there was no significant alcohol/physiological arousal interaction \[F(3,55) = 1.31, p>.05\]. Similarly, chi square analysis of the hit rate distribution revealed no significant difference between the classification of subjects in the high physiological arousal and low arousal groups.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Alcohol</th>
<th>No Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Subjective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Truthful</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Low Subjective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceptive</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Truthful</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
\[ \chi^2(1) = .002, p > .05 \].

To summarize the above analyses, it appears that all three independent variables (threat, memory priming, and alcohol intoxication) contributed to polygraph test outcome in some way. Unlike the results reported by Bradley and Ainsworth (1984), this study found that alcohol intoxication did not produce any significant group differences in numerical or categorical scores on either the CQT or GKT. However, when subjects were regrouped on the basis of their subjective arousal during the crime, there was a significant alcohol effect found. Those subjects in the high subjective arousal group scored higher (i.e. in the innocent direction) on the CQT when intoxicated, and this difference in numerical scores was substantial enough to be reflected in the categorical scores where the crime-intoxicated, high subjective arousal subjects were misclassified as truthful significantly more often than their low subjective arousal counterparts. There was no evidence of this alcohol/subjective arousal interaction in the GKT numerical or categorical scores. When subjects were grouped according to postcrime physiological arousal, no alcohol/arousal interaction effects were found in either the GKT or CQT scores.

Analysis of the threat variable showed that it had no significant effect on the numerical or categorical scores of either the CQT or GKT procedures. But when threat was examined in terms of its separate components, physiological and subjective arousal, a significant effect was evident in the GKT scores. Subjects in the low subjective arousal group were correctly classified as deceptive less often than subjects in the high
arousal group, that is, low arousal subjects escaped detection more often than high arousal subjects. This effect was not seen in analyses of the CQT scores, and the effect was not found in either the GKT or CQT scores when the subjects were split on the basis of postcrime physiological arousal.

Memory for specific crime details was found to have no effect on CQT numerical or categorical scores but there was a significant effect of the memory manipulation on the GKT outcome. Those subjects in the primed memory group had significantly higher numerical scores on the GKT than subjects whose memory was not primed. Although the mean group scores were different there was no significant difference between the primed and unprimed groups when the subjects were classified as deceptive or truthful. These results indicate that memory for crime details may be an important factor in determining a subject's GKT score, so that variable is examined more closely in the analyses of memory test scores.

Memory Test Outcome

As outlined in the Method section, three tests of memory for the mock crime details were administered to each subject immediately following the polygraph session. The tests were designed to evaluate the subject's ability to freely recall information, recall information when given cues, and to recognize crime details in a list of 5 alternatives.

The scoring system used to evaluate the free recall test produced scores in three categories: total number of correct details recalled, number of details incorrectly recalled, and
unscorable responses. The unscorable responses were usually subjective judgments such as "a pretty color" or "a typical chair." Since much of the scoring of this test was subjective, a set of guidelines was formulated and used for scoring, then 15 of the 64 tests were randomly chosen and blindly rescored by a research assistant using the same guidelines. The reliability of scoring varied for each of the three categories. Total details recalled was the most reliable category, having an interrater correlation of .90 followed by details incorrect with a correlation of .85, and the poorest reliability, .68, was with unscorable responses.

The results of the free recall test were analyzed in a 3-way (2 levels of alcohol intoxication x 2 levels of threat x 2 levels of memory priming) MANOVA with the dependent variables being each of the 3 scoring categories. Of the main effects, alcohol intoxication \([F(3,54)=2.95, p<.05]\) and memory priming \([F(3,54)=5.28, p<.01]\) were significant, but threat was not \([F(3,54)=1.75, p>.05]\). Univariate analysis of the alcohol effects showed that subjects who were intoxicated at the time of the crime made more errors in recalling crime details (3.47) than subjects who were sober (1.91) \([F(1,56)=6.86, p<.016]\) but there were no significant group differences in total details recalled or unscorable responses. Inspection of the univariate analyses of memory priming effects showed, not surprisingly, that the subjects in the primed memory group remembered more details (28.91) than those in the unprimed group (21.66) \([F(1,56)=12.29, p<.016]\). There were no group differences in the number of incorrect details recalled or in unscorable responses. The MANOVA
also showed no significant interactions.

The cued recall test data were subjected to a 3-way (2 levels of alcohol intoxication x 2 levels of threat x 2 levels of memory priming) ANOVA. Two of the main effects were significant: alcohol intoxication \( F(1,56)=7.25, p<.01 \) and memory priming \( F(1,56)=355.16, p<.001 \) but the effect due to threat was not \( F(1,56)=2.62, p>.05 \). The mean score of the primed memory group (19.56 out of 22) was substantially larger than the unprimed group (10.47). In contrast, the mean score of the alcohol intoxicated group (14.34) was significantly but not, in practical terms, substantially smaller than the sober group (15.69). None of the ANOVA interactions was significant.

When the results of the recognition test were analyzed in the same 3-way ANOVA design as above, there was only one main effect found to be significant. The mean score of the primed memory group (10.31 out of 11) was significantly larger \( F(1,56)=293.42, p<.001 \) than that of the unprimed group (6.28). The effects on test scores of threat \( F(1,56)=0.86, p>.05 \) and alcohol intoxication \( F(1,56)=0.02, p>.05 \) were not significant. Again, there were no significant interactions.

Along with memory for general crime details there is also the issue of how well subjects remember details that are central to the polygraph test. Both the recognition and the cued recall test contained the 5 central questions that made up the GKT as well as 6 questions about peripheral crime details. The 5 central questions all concerned items the subject had come in direct physical contact with. Further analysis of the recognition test results was done to determine if there was any difference between
the primed and unprimed groups in terms of memory for central or peripheral details. When a one-way ANOVA was conducted with each of the 5 central questions the mean scores of the two memory groups were significantly different on only 2 of the 5 questions. This contrasts with the score on the peripheral questions where the primed group scored significantly higher on all 6 questions. This means the difference in recognition memory test score between the two memory primed groups is due largely to differences on the peripheral questions.

It appears, from the above analyses, that the threat manipulation had no significant effect on memory for crime details since there were no group differences in any of the memory scores. Alcohol intoxication did have an effect that resulted in the intoxicated group making more errors on the free recall test and on the cued recall test, but alcohol intoxication did not significantly impair the performance on the recognition test. The most consistent effect on test scores was due to the memory priming manipulation. The primed memory group remembered more details on the free recall test and scored higher on both the cued recall test and the recognition test.

Since subjects in the high subjective arousal group scored higher on the GKT, they may have had a better recollection of the central details. A 2-way (2 levels of alcohol intoxication x 2 levels of subjective arousal) ANOVA was performed with the total score on the 5 central questions of the recognition test being the dependent variable. There was no significant group difference due to subjective arousal \( F(1,60) = 1.67, p > .05 \) and no significant interaction of alcohol with arousal. Similarly, there
were no significant effects of arousal on the recall of peripheral details when subjected to the same analysis. The same ANOVA was conducted with data from the cued recall test. Again there were no group differences in total score on the 5 central questions due to subjective arousal \([F(1,60)=2.24, p>.05]\) nor was the alcohol by arousal interaction significant. Results were similarly nonsignificant when the analysis was conducted with the total score on cued recall of peripheral details. The above analyses suggest that subjective arousal had no effect on recognition memory for either peripheral or central crime details.

Additional Analyses

To reiterate the major results with respect to polygraph test outcome, the data have suggested that subjective arousal at the time of the crime has an effect on subsequent GKT results. Subjects in the low subjective arousal group were less detectable with the GKT procedure than subjects in the high arousal group, and this group difference appears to be manifested in the SCR channel only. Subjective arousal was also shown to interact with alcohol intoxication, resulting in subjects in the high subjective arousal group receiving higher numerical scores on the CQT and thus being less easily detected as guilty if they were intoxicated at the time of the crime.

Given the above results, the question that suggests itself is: do the low and high subjective arousal groups differ in ways other than STAI score at the time of the crime? It is of interest to know if the groups differed on physiological and subjective
measures at other times during the study. A 2-way (2 levels of subjective arousal x 2 time periods) repeated measures ANOVA showed that the groups did have significantly different STAI scores throughout the study. The high subjective arousal group scored higher on the STAI at both baseline and at the beginning of the polygraph session, than the low subjective arousal group (see Figure 7). The mean STAI score for the high arousal group (37.68) was significantly higher than the low arousal mean (34.22) [F(1,61)=5.03, p<.05]. In contrast to the differences in STAI scores found above, a similar 2-way (2 levels of subjective arousal x 3 time periods) ANOVA showed no differences between groups in systolic blood pressure [F(1,59)=.03, p>.05] at either baseline, post instructions, or during the polygraph session. The interaction of time with systolic BP measures was also not significant [F(1,61)=.07, p>.05]. See Figure 8.

Since subjective arousal at the time of the crime has been shown to be important in GKT outcome and to interact with alcohol in determining CQT outcome, it is of interest to know how well the crime STAI score correlates with polygraph scores. Pearson correlations were calculated but none was significant. The crime STAI score correlation with CQT total score was .06 (p>.05), and with the CQT electrodermal score it was .08 (p>.05). The correlation of STAI score and GKT total score was .11 (p>.05) and with the GKT electrodermal score it was .17 (p>.05). This suggests that there is not a significant relationship between subjective arousal itself, as measured by the STAI, and any of the polygraph test scores.

Results of the discriminant function analysis, discussed
FIGURE 7

MEAN STAI SCORES FOR SUBJECTIVE AROUSAL GROUPS

Legend
- HIGH AROUSAL
- LOW AROUSAL

TIME OF MEASUREMENT

MEAN STAI SCORES
FIGURE 8

MEAN BLOOD PRESSURE SCORES FOR HIGH AND LOW AROUSAL GROUPS

Legend
- HIGH AROUSAL
- LOW AROUSAL

TIME OF MEASUREMENT

MEAN SYSTOLIC BP in MM Hg

BASELINE  POST INSTRUCTIONS  POLYGRAPH SESSION
earlier in the results section, suggested that the use of the CQT and GKT together would give better separation of the innocent and guilty subjects than either test achieved on its own. This raises the question of order effects. Does the use of one procedure before the other affect the test results? To investigate the effects of presentation order a 3-way (2 levels of alcohol x 2 levels of threat x 2 levels of order) MANOVA was performed on the GKT and CQT numerical scores. Order of presentation was counterbalanced in this study so each procedure was administered first the same number of times that it was administered as the second test. As with the other MANOVAs performed in this study the dependent measures were derived from the three physiological channels; SCR, finger pulse amplitude, and respiration. The MANOVA of GKT results showed no effect of order \[ F(3,54)=2.05, \ p>.05 \] nor were there any significant interactions with the alcohol \[ F(3,54)=.94, \ p>.05 \] or threat \[ F(3,54)=.98, \ p>.05 \] variables. Similarly, the MANOVA of CQT results showed no order effects \[ F(3,54)=1.42, \ p>.05 \] and no interaction with alcohol \[ F(3,54)=.65, \ p>.05 \] or threat \[ F(3,54)=.65, \ p>.05 \]. It appears then, that using the two polygraph procedures with the same subject will not significantly affect the outcome of either, regardless of the order of presentation.
Discussion

The results of this study demonstrate that alcohol intoxication during the enactment of a mock crime, subjective arousal during the crime, and memory for crime details, are all factors that can have an effect on the subsequent physiological detection of deception. The effects, however, are not always substantial and are manifested differently in the Control Question Test and the Guilty Knowledge Test. Manipulation checks confirmed that the independent variables; alcohol intoxication, threat, and memory priming, were successfully implemented and the following discussion is organized around their individual and combined effects on polygraph outcome.

The primary purpose of this study was to investigate the effects of these independent variables and the power of the analyses of main effects and interactions is felt to be adequate. Consequently, the discussion of these results is on firm ground. However, the post-hoc and nonparametric analyses that were done (such as the analysis of the two threat components) may be less powerful and therefore these results could be less reliable. With this in mind the discussion of the results of these latter analyses should be regarded as speculative although suggestive of directions for further research.

Effects of Alcohol

In the present study, subjects who consumed alcohol prior to committing the mock crime did not, as a group, show significantly different scores on the CQT than subjects who were crime-sober.
Unlike the results reported by Bradley and Ainsworth (1984), results in the present investigation reveal no generalized effect of alcohol on either numerical or categorical polygraph scores. Bradley and Ainsworth (1984) reported that a multivariate test of their polygraph scores showed that subjects who were intoxicated at the time of the crime had significantly higher CQT numerical scores (i.e. in the direction of truthfulness) than crime-sober subjects. Univariate analysis indicated that the electrodermal scores of the crime-intoxicated subjects were significantly higher than those of the sober group, but respiratory and cardiovascular measures did not significantly differ. As well as the significant difference in electrodermal scores, Bradley and Ainsworth (1984) also reported that classification of subjects as deceptive, truthful, or inconclusive, was less accurate with the crime intoxicated subjects. Using a 3 x 2 chi square analysis they found significant differences between sober and intoxicated groups when subjects were classified on the basis of the composite score (i.e. the total of 3 physiological measures). There appear, however, to be some problems with the way the chi square analysis was conducted. Although the expected frequency was less than 5 in 4 of their 6 cells (67%), they did not use the Yates' correction for continuity, and this correction factor should be used where 20% of the expected frequencies are less than 5, with some statisticians suggesting 10 (see Ferguson, 1976, p. 201). Without the correction factor the analysis of classification rates yields $\chi^2(2) = 6.3$ significant at the .05 level. Using Yates' correction factor the analysis yields $\chi^2(2) = 3.5$ which is only significant at the .20 level. In addition
to omitting the Yates' correction factor from the calculations, the authors included an inconclusive category in the analysis. In such a laboratory study where motivation to beat the test may be low and where subjects are only administered 3 charts instead of the 5 charts commonly administered in the field, including the inconclusive category in the analysis could be misleading. If the chi square analysis is repeated with only the deceptive and truthful categories, the result is even less significant; $X^2(1) = .75, p > .30$. It could therefore be argued that although Bradley and Ainsworth (1984) reported significant numerical differences in CQT scores of the crime-intoxicated and crime-sober groups, these differences did not significantly affect the classification of subjects. So in terms of hit rates, the present study has replicated the results but not the conclusions of Bradley and Ainsworth (1984) because neither study demonstrated significant differences between the crime-intoxicated and crime-sober groups when subjects were classified on the basis of CQT scores (see Table 3).

The GKT data were subjected to the same statistical analyses as the CQT data, and the results were similar to those of Bradley and Ainsworth (1984). The multivariate analysis showed no significant mean differences in numerical GKT scores between the crime-intoxicated and crime-sober groups, and chi square analysis showed no group difference in classification rate. Bradley and Ainsworth (1984) also found no significant group differences in their multivariate analysis of GKT numerical scores, but they reported a significant univariate difference between intoxicated
and sober groups on the electrodermal measure, indicating that the intoxicated group had a lower (less deceptive) mean score than the sober group. This may be misleading because by using the univariate analysis from the MANOVA they have neglected to control the alpha level. If the Bonferroni inequality principle is applied to adjust the alpha level of the three univariate analyses \( p < .016 \) the electrodermal group difference in the Bradley and Ainsworth (1984) study \( F(1,24) = 5.45, p < .05 \) would no longer be significant. The authors did not conduct a chi square analysis of the GKT classification rate because of low expected cell frequencies. It appears, then, that in terms of GKT results the present study has replicated the results but again differed in the conclusions of Bradley and Ainsworth (1984), since neither study could demonstrate significant differences between the intoxicated and sober groups on either the numerical scores or classification scores.

When statistical irregularities of the Bradley and Ainsworth (1984) study are addressed, the only difference between their results and those of the present study is their finding of an alcohol effect on the electrodermal numerical scores of the CQT. The present study did not find this effect on CQT scores. A reason why this difference in results between the two studies may have occurred came to light when effects of the threat variable were examined, and these results are discussed below.

**Effects of Threat**

In light of their results, Bradley and Ainsworth (1984) speculated that alcohol intoxication at the time of the crime
might operate through emotional or memory processes to affect polygraph results. They suggested that if detection of deception by the CQT depended on "emotional arousal" of guilty subjects to crime-relevant questions, these questions could be less arousing if the crime were committed while intoxicated. That is, alcohol may reduce the emotional impact of committing the crime so subsequent questions about the crime would be less arousing. The results of the present investigation suggest that their hypothesis is not tenable for two reasons; alcohol did not significantly affect the emotional impact of the crime, and emotion during the crime did not significantly affect the subsequent polygraph results.

For purposes of the present study "emotional arousal" was defined in terms of threat to a subject during the crime and was operationalized as measures of physiological and subjective arousal. Results of the manipulation check revealed that the high threat group had significantly greater scores on the post-crime STAI measure and higher post-crime systolic blood pressure than the low threat group, but the 3-way ANOVA indicated no effect of alcohol on either of these measures. The mean post-crime STAI score for the alcohol group was 46.58 compared to 48.94 for the no alcohol group, and systolic blood pressure was 147.16 mm/Hg and 149.78 mm/Hg respectively. This suggests that alcohol intoxication, at the level used in this study, does not significantly blunt the emotional experience of the crime. Further, analysis of the threat variable effects on polygraph results revealed no significant difference between the high and
low threat groups on either CQT or GKT outcome. This is true for
the numerical scores and categorical scores of both procedures.
On the basis of the above results it seems clear that crime-
intoxication did not significantly reduce the subject's
experience of threat in this study. Alcohol affected neither the
subjective nor the physiological measures of arousal. In
addition, threat had no significant effect on the polygraph
outcome, so even if the emotional impact of the crime was to be
reduced in some way, it would not have a significant effect on
detection of deception.

Although the threat variable did not significantly affect
polygraph outcome, some interesting results were obtained when
the two components, physiological and subjective arousal, were
examined separately. Since the post-crime STAI score and systolic
blood pressure measures were not highly correlated (r = .26) and
each could possibly reflect a different emotional response,
analyses of their individual effects on polygraph results were
carried out. The approach used was to create two groups by
conducting a median split of subjects on the basis of their
scores on each measure. A split on the basis of systolic blood
pressure gave two groups, one composed of subjects with high
systolic blood pressure following the crime and the other with
low systolic blood pressure. Further analysis revealed that these
two groups were not significantly different, numerically or
categorically, on either CQT or GKT scores. However, when
subjects were formed into high and low subjective arousal groups
on the basis of post-crime STAI scores, significant polygraph
differences were found. While multivariate analysis of numerical
scores showed no difference in GKT scores between the high and low subjective arousal groups, the classification system used in this study utilizes only the electrodermal score so a separate analysis using only the SCR scores was done. Univariate analysis using the SCR scores showed a significant difference between groups. The low subjective arousal group had smaller SCR scores, making them more difficult to detect than high arousal subjects. This numerical score difference was substantial enough that it significantly affected the classification of subjects; low arousal subjects were classified as deceptive less often than high arousal subjects (see Table 7). Thus the subjective arousal variable had not only statistically significant effects on numerical scores but also clinically significant effects as reflected in the hit rates. There was no significant interaction of alcohol and subjective arousal with respect to GKT scores as was found with the CQT.

Subjective arousal had a substantial and straightforward effect on the GKT results but the same can not be said for its effect on the CQT results. When a multivariate analysis of CQT scores was performed, no significant differences were found between the high and low subjective arousal groups; however, a significant interaction between alcohol and arousal was revealed. Subjects in the high subjective arousal group who were intoxicated at the time of the crime had significantly higher CQT scores, meaning that they were harder to detect as deceptive. Univariate analysis showed that both respiration and SCR measures were significantly affected by the interaction. It is apparent
from Figures 5 and 6 that crime-intoxication increased both the SCR and respiration scores of subjects in the high arousal group. Crime-intoxication does not have the same effect on subjects in the low arousal group, and in fact there is a trend for this group to have lower scores if intoxicated during the crime. These numerical differences are not only statistically significant but are substantial enough to result in significant differences in classification rates. Analysis of hit rates showed that 9 of 16 intoxicated subjects who reported high subjective arousal during the crime were misclassified as truthful and this was significantly greater than the 2 of 15 intoxicated subjects reporting low crime arousal who were misclassified. This significant difference in hit rates was not found in crime-sober subjects where 2 of 10 subjects reporting high subjective arousal during the crime were misclassified as truthful and 2 of 9 subjects reporting low arousal were misclassified (see Table 9).

Given the above results one might expect that the crime-intoxicated subjects in the high arousal group would be more difficult to detect than crime-sober, high arousal subjects. Although there was a trend in this direction, crime-intoxicated high arousal subjects were not misclassified more often than their crime-sober counterparts. The alternate explanation of why no significant difference was found between the two high subjective arousal groups had to do with small sample size. Since more subjects in the no alcohol group were classified as inconclusive there were only 10 subjects from this group, compared to 16 subjects in the alcohol group, included in the chi square analysis. This smaller number of high arousal, no alcohol
subjects may have decreased the power of the analysis since the value of chi square is related to sample size. If a trend in the data represents an actual group difference, this difference will increase with sample size and the associated probability will decrease (see Ferguson, 1976, p. 204). Thus, if our sample size was larger the nonparametric analyses would have been more powerful.

To reiterate the above discussion, it is apparent that alcohol did not have a general effect on polygraph results in this study, but alcohol was found to interact with subjective arousal to make intoxicated subjects who experienced high arousal during the crime more difficult to detect with the CQT procedure. Although subjective arousal during the crime did not itself significantly affect the CQT outcome, it did affect GKT results. Subjects reporting low subjective arousal during the crime were more difficult to detect as deceptive with the GKT procedure than high arousal subjects. Unlike the CQT, the GKT results were not affected by the combination of high arousal and intoxication.

Subjective arousal seems to have a somewhat contrary effect on the CQT and GKT procedures. While high subjective arousal combined with intoxication decreases detectability with the CQT procedure, it is low subjective arousal that reduces detectability with the GKT procedure. Although this may seem paradoxical it is consistent with Raskin's (1979) explanation of detection of deception in terms of orienting and defensive responses. He outlined a theoretical framework for physiological detection of deception that is based on the concepts of arousal
and attention. He suggested: 

"(1) that detection of deception and the successful differentiation of guilty and deceptive subjects are based on the differential arousal value of various stimuli for deceptive and truthful subjects, and (2) that arousal value of the stimuli is determined by the basic ingredients of the deceptive context and the psychological set established by the examination procedure that induces differential amounts of signal value to various stimuli and sometimes produces information-processing in response to certain stimuli within the stimulus sequence (p. 588)."

He further postulated that a subject's responses to the stimulus questions of the CQT and GKT may be mediated by different physiological processes. The relevant CQT questions are accusatory, threatening, and personal, and require a direct denial of guilt, therefore the subject responds in a defensive manner. In comparison, the GKT questions are not accusatory but are an attempt to determine if the subject has knowledge of crime details that only a guilty person would know. The critical item has greater signal value than noncritical items for the guilty subject but not for the innocent subject. The signal value of the critical item for a guilty subject would produce a large orienting response to that item whereas the innocent subject's response to the critical item would be no larger than his responses to the noncritical items.

Orienting and defensive reactions serve different biological purposes. The orienting response prepares the body to deal with a new or important stimulus in the environment and involves a physical and mental orientation towards the stimulus. A highly salient stimulus, such as a child's cry, will continue to evoke
an orienting response regardless of familiarity. The defensive response is a more dramatic physiological response that prepares the body for a threatening stimulus. If the CQT produces reactions associated with defensive responses and the GKT reactions are predominantly orienting responses, the physiological characteristics of responses to the two tests should be different. Raskin (1979) presented evidence that physiological responses observed with the GKT differ from those obtained with the CQT. The results of the present study further support the two-response theory albeit from a different perspective. If the CQT and GKT relied on the same physiological response system one would expect subjective arousal to affect both tests in the same way. This is not the case because high subjective arousal when associated with alcohol diminishes the accuracy of the CQT and has no effect on the GKT. Conversely, low subjective arousal diminishes the accuracy of the GKT but has no effect on the CQT. This suggests that each procedure is mediated by different response systems.

Following Raskin's (1979) model, the signal value of the CQT relevant questions is established during the polygraph session with the introduction and review of questions, but the signal value of the critical GKT items is established at the time of the crime. It seems likely that if a subject is highly aroused at the time of the crime, details of the event would be more meaningful and thus acquire more signal value than they would for the low arousal subject, and conversely, anything which reduced the signal value would reduce the subject's response to the
critical item and consequently would reduce the discriminatory power of the GKT. Arousal then, would have a direct influence on how a subject responds on the GKT. Since response to the CQT questions depends heavily on what transpires in the pretest interview, arousal at the time of the crime would not be such an important factor with that procedure.

Results in the present study are consistent with the idea that level of arousal during the crime would influence the GKT but not the CQT. However, the relationship of subjective arousal to GKT score is not clear. Subjects reporting low subjective arousal during the crime score significantly lower on the GKT and therefore avoid detection more often than subjects in the high arousal group, yet scores on the post-crime STAI do not correlate significantly with GKT electrodermal score or with the GKT total score. This suggests that it may not be subjective arousal itself that affects GKT score; rather, STAI score may be an index of another dimension, possibly a personality dimension, that is important to detection of deception with the GKT procedure.

Iacono, Boisvenu, & Fleming (1984) also found that precrime score on the A-State portion of the STAI did not correlate significantly with GKT electrodermal score. In addition, they administered the Differential Personality Questionnaire (DPQ), an omnibus personality inventory, to subjects and found no correlation between personality scores and GKT scores. But since their investigation of personality was cursory, they may not have touched on dimensions important to GKT outcome. Such a trait would have to correlate highly with A-State measures of crime arousal and persons scoring high on measures of this trait would
be expected to score higher on the STAI even during baseline conditions, as the high arousal group did in this study.

In terms of this study, persons scoring high on the proposed trait would be more likely to be engaged by the experimental situation and therefore might attach more importance and meaning to details of the crime. The crime details would then in turn have more signal value when encountered on the GKT. Conversely, subjects scoring low on measures of this trait would not attach as much meaning to the crime details and would give smaller orienting responses when presented with these details on the GKT. It is tempting to explain the effects of subjective arousal in terms of attention to, and memory of, crime details, but memory tests revealed no difference in recall or recognition of relevant details between the high and low subjective arousal groups.

A personality dimension related to STAI score, as described above, would also have to account for the significant interaction of alcohol with the high subjective arousal subjects which resulted in this group being harder to detect with the CQT procedure. Raskin's (1979) detection of deception model suggests a possible explanation. According to this model, the pretest interview creates a situation where the signal value of the relevant CQT questions is strong for guilty subjects but is reduced for innocent subjects. The signal value of relevant questions is strong for guilty subjects because they are in a situation where they have to deny or be defensive about their guilt. The signal value of the relevant questions is less for the innocent subjects because they have been convinced that the
polygraph is accurate and they do not have to be as defensive in their responses. The results in the present study suggest that the signal value of the relevant questions has been reduced for the crime-intoxicated, high arousal subjects. This would be possible if the subject's sense of guilt or responsibility was diminished by the fact that he had been intoxicated during the crime. Subjects who scored high on a personality dimension related to the STAI measure might be more likely to see the responsibility for the theft less as a personal matter and more as an external or situational matter related to intoxication. These high arousal subjects would then be less easily detected than the high arousal subjects who did not have the excuse of intoxication because the relevant questions would have less signal value for the intoxicated subjects. It may be the case that the study of Bradley and Ainsworth (1984) had a large number of subjects who fell into the high subjective arousal category and therefore they found significantly different CQT numerical scores between the alcohol and no alcohol groups, whereas this difference was not found in the present study until a median split was conducted.

Since the investigation of crime-arousal was done in a post-hoc analysis, no definitive explanation can be made for its mode of action. It seems clear that subjects who experience low subjective arousal during the crime are more difficult to detect as deceptive with the GKT procedure, and subjects reporting high subjective arousal are more difficult to detect with the CQT procedure when they are intoxicated during the crime. Clarification of why this occurs will require examination of the
subjective arousal variable in an appropriate experimental design. Such a study would manipulate subjective arousal in the same manner as the threat variable was manipulated in the present study. This would be the first step in determining if subjective arousal itself is important to polygraph test outcome or whether it merely reflects a personality dimension that is not only a factor in a subject's response to polygraph questions but is also influenced by alcohol.

**Effects of Memory Priming**

Besides emotional processes, Bradley and Ainsworth (1984) also suggested memory as a process through which alcohol might work to reduce detectability with the polygraph, especially with the GKT procedure. In their study they found no difference between the intoxicated and sober groups on a recognition memory test composed of 9 crime relevant items, but the authors felt that a simple recognition test of memory may not have been powerful enough to discriminate differences between the groups. They suggested that group differences in memory for crime details might emerge with a nonprompted recall test.

In the present study three tests of memory were employed: a recognition test similar to that used by Bradley and Ainsworth (1984), a cued recall test, and a free recall test. Results of the recognition test showed no significant difference in scores between the intoxicated and sober groups, as was the case in the Bradley and Ainsworth (1984) study. As expected there was a large and significant effect on recognition test score due to memory
priming, an indication that the manipulation of this independent variable was successful. Threat had no significant effect on memory scores.

Results of the cued recall test were markedly different than those of the recognition test. As postulated by Bradley and Ainsworth (1984), effects of crime intoxication were seen in cued recall test scores. Intoxicated subjects scored significantly lower (14.34) than crime-sober subjects (15.69), but despite the statistical significance of the difference between the groups, the practical significance of the difference is minimal. A difference in score of 1.35 between the two groups, with respect to a total possible score of 22, does not seem substantial. The effect of memory priming on cued recall score was significant and much more substantial than the alcohol effect. The primed memory group had a mean score of 19.56 out of 22 and the unprimed group scored only 10.47. As with the recognition test scores, threat was not a factor in cued recall test scores.

Three separate scores were generated from the free recall test: number of errors, total details recalled, and unscorable responses. Multivariate analysis revealed that crime-intoxication had a significant affect on only one score, number of errors. Crime-intoxicated subjects, on average, made more errors (3.47) in recalling crime details than did crime-sober subjects (1.91), but there were no group differences in total details or unscorable responses. Memory priming also had a significant effect, as one would expect, resulting in the primed group recalling more details (28.91) than the unprimed group (21.66), but group differences in errors and unscorable responses were not
significant. As with the other two memory tests, threat had no significant effect.

The effects of alcohol on memory reported above are consistent with the state-dependent learning literature which suggests that uncued recall of target events is reliably disrupted by alcohol intoxication at the time of learning, but recognition of the events is not. Cued recall of events is less reliably disrupted than free recall (see Eich, 1977). Given that alcohol has had the expected effect on memory processes in the present study, the larger issue of how these memory disturbances influence polygraph outcome can be addressed.

It is clear that subjects intoxicated during the crime recall less on a cued recall test of memory and make more errors on a free recall memory test, but while these differences are statistically significant they are not large. This distinction turns out to be important because the action of alcohol on memory processes was not substantial enough to affect polygraph outcome. The results of the GKT, a procedure potentially susceptible to memory disturbances, were not affected by alcohol. And the CQT results, although affected by an interaction of arousal and alcohol, were not affected by memory disturbances since both high and low subjective arousal groups had similar memory test scores.

Despite the fact that alcohol did not disrupt memory processes sufficiently to affect polygraph scores, memory for crime details was shown to be an important variable. Results of the memory priming manipulation showed that the primed memory group had significantly higher electrodermal scores on the GKT.
than the unprimed group. In contrast, there was no significant
difference in CQT score between the primed and unprimed group,
but this is a reasonable finding since memory is not thought to
be an important factor in the CQT procedure.

Although the primed memory group had higher numerical scores
on the GKT, when subjects were classified as deceptive or
truthful there were no significant differences in hit rates but
there was a trend for unprimed subjects to go undetected more
often than primed subjects. The reason for this incongruity in
significance of numerical and categorical scores is likely due to
the fact that subjects in both the primed and unprimed groups
scored well on the central recognition test questions, questions
about items which made up the GKT. There was a significant group
difference in score on only 2 of 5 recognition questions, and
likely it is these two questions which were responsible for the
numerical GKT score difference, but since the group differences
were confined to only 2 central questions the categorical score
was not significantly affected. This illustrates the importance
of choosing the GKT questions well. When the GKT question format
was constructed for the present study only objects the subject
came in physical contact with were used as target details. This
assures that the subject has attended to, at least briefly, each
of the target details. In the present study this approach was
successful because even the unprimed group remembered the central
details well. The same can not be said for the peripheral
details, of which all 6 were recognized significantly more often
by subjects in the primed memory group.

The above results, which indicate that memory for crime
details is important to GKT outcome, are consistent with the results of Waid, Orne, Cook, & Orne (1978) and Iacono, Boisvenu, & Fleming (1984). The latter study found a correlation of $r = .53$ ($p < .0001$) between a crime detail recognition test score and GKT electrodermal score. This correlation is identical to the one calculated in the present study between the full recognition test score and SCR score ($r = .53$, $p < .0001$) and close to the correlation of cued recall test score with SCR score ($r = .51$, $p < .0001$).

Although the correlations are highly significant they indicate that memory for crime details only accounts for approximately 28% of the variance seen in GKT score and this suggests that other variables such as salience or signal value of the crime details also play an important role in GKT detection of deception. This is congruent with the finding of the present study that subjective arousal at the time of the crime can influence GKT score in a manner that may involve changes in signal value of crime relevant details. Further support for the idea that item salience is important was provided by Bradley and Warfield (1984). In their study they found that subjects who had actually committed a mock crime scored higher on the GKT than subjects who only witnessed the crime or subjects who were merely told the crime-relevant details.

Together the memory data and the analyses of the threat variable support Raskin's (1979) hypothesis that the occurrence of autonomic responses to the GKT and CQT questions are mediated by different physiological response systems, the orienting and defensive responses. In fact none of the independent variables
affected the GKT or CQT in the same way. While low subjective arousal was found to decrease GKT scores, it was high subjective arousal that was important to the CQT. Memory variables affected GKT results but not CQT results, and the opposite was true of alcohol which selectively affected CQT results but not GKT results. Raskin's model provides a framework for understanding why some of the seemingly paradoxical effects of the independent variables occur.

A side issue that the present study was able to address is that of order effects. Discriminant function analysis of the CQT and GKT results showed that both tests used together discriminated the deceptive and truthful subjects better than either test could do on its own. Although the GKT is not used much in field situations, results here suggest that it might be a useful adjunct to the CQT. An obvious potential difficulty of using both tests is that one test might detract from the procedure it precedes. Such things as habituation, boredom, or fatigue during the testing session could diminish the effectiveness of the procedure administered last. Results in the present study indicate that order of test presentation did not significantly decrease the effectiveness of either the CQT or the GKT. Subjects were taken for a 5 minute walk through the building between administration of the first and second test procedure, and this may have helped to maintain alertness and interest. Under the above conditions it appears that both the CQT and GKT procedures can be used serially with a subject without compromising the effectiveness of either.
Summary

In conclusion, the primary hypothesis being tested, that crime-intoxication will decrease a subject's detectability with the CQT and GKT procedures, has only been partially supported. While it appears that there is no generalized effect of alcohol, crime-intoxication did reduce detectability with the CQT procedure in some subjects. Subjects who reported high subjective arousal at the time of the crime, and who were intoxicated, were misclassified as truthful more often than other subjects. Since the investigation of subjective arousal was done post-hoc explanations of its relationship to polygraph scores can only be speculative, although it appears that subjective arousal is not directly related to polygraph outcome but may instead be an index of another important dimension, possibly a personality trait. Alcohol had none of the above effects on GKT scores, so while the primary hypothesis was partially supported with respect to CQT outcome, no support was found in terms of GKT outcome.

The analysis of subjective arousal is also relevant to the second hypothesis being addressed in this thesis: alcohol acts through memory and emotional processes. Analysis of the effects of memory priming showed that memory for crime details has a direct influence on GKT scores: those subjects who remember more details score higher on the GKT. Results also clearly showed that crime-intoxication did impair a subject's ability to remember crime details, but this impairment was not substantial enough to affect the GKT score. As expected, memory for crime details was not a factor important to CQT score. Analysis of the threat
variable suggested that the experience of threat during the crime was not a factor in determining the outcome of either the GKT or CQT, but when threat was examined in terms of its components, physiological and subjective arousal, an effect was found. If a subject reports low subjective arousal during the crime he is more difficult to detect as deceptive on the GKT. This contrasts with the CQT results which indicate that subjective arousal alone has no effect on polygraph score, but when high subjective arousal is combined with intoxication, subjects are less easily detected as deceptive. Thus the second hypothesis was also only partially supported. Memory processes are important for GKT outcome, but memory is not sufficiently affected by alcohol (at the dose used in this study) to significantly diminish GKT scores, and emotional processes during the crime may be important to polygraph outcome but they appear to be more important to the GKT rather than the CQT. Future research will be necessary to more fully understand the relationship of subjective arousal during the crime to physiological detection of deception.
References


and physiological measures in the detection of deception. 
*Psychophysiology*. 15, 344-359.

and the detection of deception. *Psychological Bulletin*. 
84, 782.

San Francisco: W.H. Freeman.

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

Raskin, D.C., (1979). Orienting and defensive reflexes in the 
detection of deception. In H.D. Kimmel, E.H. Van Olst, & J.F. 
Orlebeke (Eds.), *The orienting reflex in humans*. 587-605, 
New York: John Wiley & Sons.

professional and legal issues surrounding application and 
1, 29-74.

Reid, J.E., & Arther, R.O. (1953). Behavior symptoms of lie 
detector subjects. *Journal of Criminal Law, Criminology and 

polygraph ("lie-detector") technique.* 
Baltimore, Maryland: Williams & Wilkins.

polygraph ("lie detector") technique*. 2nd Ed. Baltimore: 
Williams & Wilkins.

factors in free recall of verbal and visual material. 
*Psychonomic Science*. 16, 221-222.

Schachter, J. (1957). Pain, fear, and anger in hypertensives and 

Skolnick, J.H. (1961). Scientific theory and scientific evidence: 
70, p. 714.

Slowick, S.M., & Buckley, J.P. (1975). Relative accuracy of 
polygraph examiner diagnosis of respiration, blood pressure, 
and GSR recordings. *Journal of Police Science and 
Administration*. 3, 305-309.

manual for the State-Trait Anxiety Inventory*. Palo Alto,


Appendix I

Consent Form

The purpose of this project is to study the effects of alcohol on the physiological detection of hidden or concealed information. The study involves two sessions. In the first, I will be asked to perform a certain task, which may involve taking some money from a room. I may also be asked to drink alcohol. At the conclusion of this session, I realize that I may still, to some extent, be under the influence of the alcohol I have received. I agree to remain in the laboratory with a research assistant for one hour following the session to allow some of the alcohol's effect to wear off. I also agree to avoid, for at least four hours after leaving the laboratory, any task that requires mental alertness such as operating machinery or driving a motor vehicle. Also, I agree to refrain during this period from taking any other drugs or alcohol.

I will be asked to return one week later for a second session in which my body's physiological responses will be recorded while I respond to some questions. Electrodes will be attached to my fingertips to measure skin conductance and on my arms to measure heart rate. A belt around my chest will monitor respiration. I understand that there is nothing dangerous or harmful about these procedures and that all information obtained in this project will be kept confidential and used only for the purposes of this study. Confidentiality will be maintained by assigning code numbers to collected data and by restricting access to the data to those working on this project. I understand that this study will require about 3 hours of my time over 2
days. I also realize that I may ask clarifying questions at any point during the session and that I am free to withdraw from the study at any time without prejudice. I have read the above statement, answered the medical survey, and agree to participate in this study. I also hereby acknowledge receiving a copy of this consent form.

Signature: __________________________

Print Name: __________________________

Date: ________________________________

Study I.D. #: _________________________

Witness: ______________________________
Appendix II

Medical Questionnaire

Caller ______________________  Date ________________
Subject Name ________________  Address ________________
Age _________________________  Weight ________________
Phone # ______________________

1. Have you been under the care of a physician for any medical problems over the past 2 years? If yes what are they? ________________

2. Have you ever been under a doctor's care for emotional or nervous problems (ie. depression, anxiety, etc.)? ________________

3. Have you ever had any unusual drug reactions, drug allergies or allergies of any kind? ________________

4. Have you ever had any of the following conditions:

   Seizures or loss of consciousness?  yes  no
   Persistent headaches?  yes  no
   Head injury with loss of consciousness?  yes  no
   Dizziness?  yes  no
   Problems with vision or glaucoma?  yes  no
   Persistent muscle weakness (myasthenia)?  yes  no
   Respiratory problems, including asthma?  yes  no
   Heart condition, hypertension, or other cardiovascular conditions?  yes  no

5. Do you use any of the following drugs?

   Major tranquilizers  how often? _________
   Minor tranquilizers  how often? _________
   Barbiturates  how often? _________
Amphetamines how often? 
Any other drugs how often? 

6. How many drinks of alcohol do you have each week on average?
   a) bottles of beer 
   b) glasses of wine 
   c) liquor 

7. Have you ever had an adverse reaction to alcohol or orange juice and pineapple juice (will be mixed with alcohol)?
### Appendix III

**Alcohol/Weight Chart - Total dose = 1.32 ml/kg**

<table>
<thead>
<tr>
<th>lbs</th>
<th>Kg</th>
<th>dose (mls)</th>
<th>1/3 dose</th>
<th>1/2 dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>54.43</td>
<td>72</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>125</td>
<td>56.70</td>
<td>75</td>
<td>25</td>
<td>37.5</td>
</tr>
<tr>
<td>130</td>
<td>58.97</td>
<td>78</td>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>135</td>
<td>61.23</td>
<td>81</td>
<td>28</td>
<td>40.5</td>
</tr>
<tr>
<td>140</td>
<td>63.50</td>
<td>84</td>
<td>28</td>
<td>42</td>
</tr>
<tr>
<td>145</td>
<td>65.77</td>
<td>87</td>
<td>29</td>
<td>43.5</td>
</tr>
<tr>
<td>150</td>
<td>68.04</td>
<td>90</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>155</td>
<td>70.31</td>
<td>93</td>
<td>31</td>
<td>46.5</td>
</tr>
<tr>
<td>160</td>
<td>72.57</td>
<td>96</td>
<td>32</td>
<td>48</td>
</tr>
<tr>
<td>165</td>
<td>74.84</td>
<td>99</td>
<td>33</td>
<td>49.5</td>
</tr>
<tr>
<td>170</td>
<td>77.11</td>
<td>102</td>
<td>34</td>
<td>51</td>
</tr>
<tr>
<td>175</td>
<td>79.38</td>
<td>105</td>
<td>35</td>
<td>52.5</td>
</tr>
<tr>
<td>180</td>
<td>81.65</td>
<td>108</td>
<td>36</td>
<td>54</td>
</tr>
<tr>
<td>185</td>
<td>83.91</td>
<td>111</td>
<td>37</td>
<td>55.5</td>
</tr>
<tr>
<td>190</td>
<td>86.18</td>
<td>114</td>
<td>38</td>
<td>57</td>
</tr>
<tr>
<td>195</td>
<td>88.45</td>
<td>117</td>
<td>39</td>
<td>58.5</td>
</tr>
<tr>
<td>200</td>
<td>90.72</td>
<td>120</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>205</td>
<td>92.99</td>
<td>123</td>
<td>41</td>
<td>61.5</td>
</tr>
<tr>
<td>210</td>
<td>95.25</td>
<td>126</td>
<td>42</td>
<td>63</td>
</tr>
<tr>
<td>215</td>
<td>97.52</td>
<td>129</td>
<td>43</td>
<td>64.5</td>
</tr>
<tr>
<td>220</td>
<td>99.79</td>
<td>132</td>
<td>44</td>
<td>66</td>
</tr>
<tr>
<td>225</td>
<td>102.06</td>
<td>135</td>
<td>45</td>
<td>67.5</td>
</tr>
</tbody>
</table>
Appendix IV

Crime Instructions - High Threat

You are about to commit a theft and this is the scenario. Listen carefully. It is very important that you do this correctly.

You have gone to see a psychology professor in his office to buy a lab manual. You see he has some money in a cashbox that is in his desk. He has left the office and you decide to go in and steal the money. You have already seen the prof and now you are going back to steal the money. The office number is 1234. You have no idea when the prof will return so you must accomplish the deed as quickly as possible. In order to increase the realism of this crime we have not told other people in the area of the office anything about this experiment, so you must be as unobtrusive as possible and don't arouse suspicions. There have been several thefts in the building recently so people are suspicious of strangers. I will show you the way to get to the office. What you must do when you get into room 1234 is:

- First, get the cashbox key from the coat pocket.
- Second, unlock the cashbox and take the money.
- Third, return here with the money.

Again, be very careful and do not make a mistake in carrying out this crime. Do you have any questions?
Crime Instructions - Low Threat

We are interested in seeing if the act of removing some money from a cashbox has any effect on blood pressure or heart rate. I will take you to a room, have you go in and remove some money, and then measure your blood pressure and heart rate. Afterwards, you will fill out a questionnaire which will give us additional information about this task. I will take you to the room which is numbered 1234. What you must do when you get into room 1234 is:

- First, get the cashbox key from the coat pocket.
- Second, get the cashbox from the desk, unlock it, and take the money.

I will be with you when you finish, waiting outside the door.
Appendix V

GKT Instructions

If you took the money you will be aware of certain crime details. I am going to ask you some questions about the crime and after each one I want you to answer no, as you did earlier in the card test.

Is that clear?

GKT Questions

1. The key to the cashbox was taken from a coat pocket. Was that coat hanging.
   a) in a closet
   b) on the back of a chair
   c) on a wooden stand
   d) on a wall hook
   e) on the back of the door

2. Regarding the amount of money taken from the cashbox. Was it
   a) $5
   b) $10
   c) $20
   d) $50
   e) $100

3. The cashbox was taken from a drawer of a desk. Was that drawer on:
   a) the lower left side
   b) the upper right side
   c) the upper left side
   d) the middle
4. The cashbox key was attached to something. Was that object:
   a) a rabbit's foot
   b) a bottle opener
   c) a whistle
   d) a knife
   e) a small wrench

5. There was something else besides money in the cashbox. Was that thing:
   a) a necklace
   b) a cheque book
   c) a diary
   d) a pocket watch
   e) a savings bond
Control Question Test Format: Chart 1

The test is about to begin:

1. Do you live in Vancouver? (irrelevant)

2. Regarding the theft, do you intend to answer all questions about it truthfully? (sacrifice relevant)

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

4. Before the age of 18, did you ever steal anything of value? (lie control)

5. Did you take the money from the professor's office? (relevant)

6. Have you ever lied about stealing something to stay out of trouble? (lie control)

7. Did you take the money from the cashbox? (relevant)

8. When you were younger did you ever steal any money? (lie control)

9. Did you take the money from the office? (relevant)
Control Question Test Format: Chart 2

The test is about to begin:

1. Were you born in the month of _______?
2. Regarding the theft, do you intend to tell the truth?
3. Are you convinced that I will only ask questions on this test that you have already okayed?
4. Have you ever lied about stealing something to stay out of trouble?
5. Did you take the money from the professor's office?
6. Before the age of 18 did you ever steal anything of value?
7. Did you take the money from the office?
8. When you were younger did you ever steal any money?
9. Did you take the money from the cashbox?
Control Question Test Format: Chart 3

The test is about to begin:

1. Is your first name ____________?
2. Regarding the theft, do you intend to tell the truth?
3. Are you convinced that I will only ask questions on this test that you have already okayed?
4. When you were younger did you ever steal any money?
5. Did you take the money from the cashbox?
6. Have you ever lied about stealing something to stay out of trouble?
7. Did you take the money from the professor's office?
8. Before the age of 18 did you ever steal anything of value?
9. Did you take the money from the office?
Appendix VI

Memory Priming Instructions

You have just taken some money and below are some details you should have noticed while in the office.

You entered the office and took $20 from a cashbox with a yellow top, found in the upper right-hand drawer. Also in the cashbox was a pocket watch. The key to the cashbox was in the pocket of a grey coat which was hanging on a wooden coat-rack near a brown recliner chair. Attached to the key was a small wrench. On the coat-rack next to the coat was a scarf. Immediately behind the desk was a painting hung on the wall. Just inside the office door, on the floor, was a stack of cardboard boxes.
Appendix VII

Recognition Test

1. The key to the cashbox was taken from a coat pocket. Was that coat hanging.
   a) in a closet
   b) on the back of a chair
   c) on a wooden stand
   d) on a wall hook
   e) on the back of the door

2. Regarding the amount of money taken from the cashbox. Was it
   a) $5
   b) $10
   c) $20
   d) $50
   e) $100

3. The cashbox was taken from a drawer of a desk. Was that drawer on:
   a) the lower left side
   b) the upper right side
   c) the upper left side
   d) the middle
   e) the lower right side

4. The cashbox key was attached to something. Was that object:
   a) a rabbit's foot
   b) a bottle opener
   c) a whistle
   d) a knife
   e) a small wrench
5. There was something else besides money in the cashbox. Was that thing:
   a) a necklace
   b) a cheque book
   c) a diary
   d) a pocket watch
   e) a savings bond
Recall Test

On the sheet provided I would like you to list features of the crime and the room where the theft occurred. These details of the crime can include your actions; descriptions and locations of furniture, decorations, and objects in the room; as well as physical features of the room itself. I would like you to list as many details as you can. The order in which you list things is not important.

When you are finished please place the answer sheet in the envelope provided and return it to the examiner.

If you did not commit the crime, it is not necessary to complete this task but I would like you to place the answer sheet in the envelope, and then wait 5 minutes before you return the envelope.

Thank you.

Name ________________________________

Date ________________________________
Cued Recall Test

1. The top of the cashbox was a distinctive color. What color was it?

2. The room was illuminated by a lamp. Describe the lamp and its location.

3. What was the amount of money taken?

4. The key to the cashbox was taken from a coat pocket. Where was the coat hanging?

5. Which drawer of the desk was the cashbox taken from?

6. Immediately behind the desk was something hanging on the wall. What was that thing?

7. The key to the cashbox was found in a coat pocket. What was the color of that coat?

8. What was the cashbox key attached to?

9. Name one object that was on top of the desk.

10. Near the coat was another article of clothing. What was that article?
11. Beside the door, just inside the office, there was something on the floor. What was that thing?

12. There was a stuffed recliner chair in the office. What was the color of that chair?