INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI
LABORATORY INVESTIGATIONS
IN THE PSYCHOPHYSIOLOGICAL DETECTION OF DECEPTION

by

ISATO FURUMITSU

A thesis submitted in conformity with the requirements for
the Degree of Doctor of Philosophy, Graduate Department of
Psychology, in the University of Toronto.

© Isato Furumitsu 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.
LABORATORY INVESTIGATIONS IN THE PSYCHOPHYSIOLOGICAL DETECTION OF DECEPTION

Doctor of Philosophy, 1999

Isato Furumitsu

Department of Psychology, University of Toronto

Abstract
In the psychophysiological detection of deception, the Guilty Knowledge Test (GKT) is a far more objective way of distinguishing between guilty and innocent suspects, but is less in use in field practice than the controversial Control Question Test (CQT). In an effort to contribute towards full understanding and standardization of the GKT, two experiments were carried out to examine accuracy of the GKT in a laboratory situation where participants role-played a crime scenario. Experiment I was a conceptual replication of a study by Nakayama et al. (1988), who reported that detection was superior with skin conductance response (SCR) as dependent measure, when participants were required to delay their answers for 8 seconds rather than answer immediately as is customary. In the experiment, the GKT questions were presented in a visual mode on computer screen, a procedure that is more standardized than the usual procedure where a human examiner asks the questions. The results indicated that both SCR and reaction time (RT) discriminated between
relevant (larger SCR and shorter RT) and neutral items, but with no difference between immediate- and delayed-answer conditions. In Experiment II, auditory computer-synthesized voice presentation of the GKT was employed and the participant's emotional reaction to the relevant items was manipulated. To examine an orientation and habituation account of the GKT effect, a simple orienting response (OR) paradigm for obtaining participants' SCRs to repeated stimuli was carried out prior to the detection phase. The emotional-reaction manipulation did not affect detection, which, however, was again significant both for the SCR (larger) and RT (shorter). One of the OR effects, reinstated OR, showed a small but significant correlation with the GKT effect. Taken together, the present results suggest that GKT is effective regardless of response manner, and its effectiveness can be partly attributed to an attentional, orienting process in the participant, rather than to emotional factors. Additionally, computerized presentation may be useful for increasing objectivity in the detection of deception through the GKT. Non-physiological dependent variables like reaction time may also provide additional information for detection of deception.
Acknowledgments

First and foremost I want to thank my supervisor, John J. Furedy, for his valuable encouragement and advice at every stage of this work. I also want to thank the other members of my committee, Kenneth L. Dion and Ronald J. Heslegrave, for their advice and illuminating discussions at various stages of the project.

I thank Yo Miyata and Shoji Kakigi in Japan, who jointly recommended me to study abroad and originally introduced me to John J. Furedy. I extend my thanks to Magnús Kristjánsson and his wife, Guðrún Arnarsdóttir, for their co-operation and friendship throughout graduate school. I also express my appreciation to Makoto Nakayama and Shinji Hira, two scholarly active Japanese police polygraphers, for their helpful comments on an earlier version of this thesis.

However, my warmest thanks go to my wife, Shoko. I thank her for her patience, understanding, and unrelenting support.
TABLE OF CONTENTS

INTRODUCTION .................................................................1
  Psychophysiological Studies of Deception .........................1
  Brief History of Detecting Deception ..............................3
  General Description of the Control Question Test (CQT) ..........6
  Problems Inherent in the CQT ...........................................11
    Ambiguous meaning of "control" questions .......................11
    Lack of standardization of the method ..........................13
    Poor psychophysiological quantification ........................15
    Contamination problems in diagnoses .............................17
    Failure to separate detection of deception from confession inducement .............................................19
  General description of the Guilty Knowledge Test (GKT) ........20
  The GKT as A Scientific Detecting Method ........................22
    Scientific control questions ......................................22
    Ease of standardization ............................................23
    Less contamination in the procedure ................................23

SCOPE OF THE STUDY ..........................................................26

EXPERIMENT I .................................................................28
  Design of Experiment I ..................................................31
  Method .................................................................33
    Participants .........................................................33
    Apparatus ...........................................................33
    Procedure ...........................................................34
    Dependent Measures ................................................41
  Results .................................................................44
    Skin Conductance Responses during the Matching Trial ........44
    Skin Conductance Responses during the Detection Period .........46
    Classification Index .................................................50
Reaction Time to the Relevant and Neutral Question ........................................ 51
Discussion ........................................................................................................ 53
EXPERIMENT II .................................................................................................. 59
Design of Experiment II ..................................................................................... 65
Method ................................................................................................................ 66
Participants ......................................................................................................... 66
Apparatus ........................................................................................................... 68
Procedure ........................................................................................................... 68
Dependent Measures .......................................................................................... 73
Results ................................................................................................................ 75
Skin Conductance Responses in the Preliminary Novelty Paradigm (PNP) .... 75
Effects of the Alarm Manipulation on the Experimental Group ......................... 76
Skin Conductance Responses during the Detection Period ................................. 79
Reaction Time to the Relevant and the Neutral Questions ................................ 85
Relationship between the Novelty Effects and the GKT Effect ......................... 89
Classification Index ............................................................................................ 91
Discussion ........................................................................................................... 92
GENERAL SUMMARY AND DISCUSSION ....................................................... 100
REFERENCES ..................................................................................................... 113
APPENDIX A ....................................................................................................... 124
APPENDIX B ....................................................................................................... 126
INTRODUCTION

Psychophysiological Studies of Deception

Human society has always been interested in detecting deception, especially when the deception is practised to cover up guilt. A central concern of the present thesis is the psychophysiological detection of deception, also popularly known as the lie detection or polygraph test, in which, by definition, only unobtrusively measured physiological responses are examined for detecting deception. The name "lie detection" or "detection of deception" suggests that these techniques directly deal with deception or identify "specific lie responses" in humans, but, as we shall later see in detail, accurate detection of guilt could, in principle, be achieved without there being any specific lie responses.

Another problem in the psychophysiological approach to deception is to specifically differentiate between deception and truthfulness as psychological processes, which would be reflected by physiological indices. To demonstrate deception as a psychophysiological phenomenon, a comparison of physiological responses should be made between two conditions, experimental and control conditions, which differ only with respect to deception (Furedy, Davis, & Gurevich, 1988). Furedy and his colleagues (Furedy et al., 1988; Furedy, Gigliotti, & Ben-Shakhar, 1994; Furedy, Posner, & Vincent, 1991; Vincent & Furedy, 1992) have pursued this area
of research with their Differentiation-of-Deception Paradigm (DDP), in which research participants are presented with a series of questions, usually concerning their own autobiographies, and requested to produce deceptive answers to half of them, and true answers to the other half. Given that all the questions are equivalent in their significance to the participants and in their frequencies in the sequence, only the difference between the experimental and control conditions, it is argued, constitutes deception itself. However, this is not the case in any applied psychophysiological detection techniques. The DDP studies with electrodermal response as a dependent variable have yielded the expected increase in responding to deceptive question relative to honest questions (Furedy et al., 1988), and have indicated that this effect is probably not due to cognitive factors such as differential retrieval difficulty, and that it is also relatively unaffected by motivational factors (Vincent & Furedy, 1992). However, as argued by Furedy et al. (1988), the DDP does not have any direct applied significance. A related point is that it is important to distinguish between the applied aim of detecting deception and the scientific aim of differentiating deception from other related but different psychological processes (Ben-Shakhar & Furedy, 1990). As stated above, the present thesis is concerned with the former, detection-oriented type of research. My central purpose is to contribute towards understanding and standardization of a psychophysiological
detection of deception that has a genuine scientific rationale.

**Brief History of Detecting Deception**

From ancient days to the Middle Ages, commonly used methods of detecting deception included trial by combat, trial by ordeal, and physical torture (Kleinmuntz & Szucko, 1984a; Trovillo, 1939). These methods may have been the principal means of solving disputes for many administrators in those days, but they are far from determining, on a scientific basis, whether deception has in fact occurred. Torture, in particular, was designed to detect guilt by eliciting a confession. However, there is no guarantee that such a confession is true; it might be given only to stop the interrogation process and the torture.

A more subtle and plausible way to detect deception is to observe overt behavioral indices of deception, and this may be commonly practiced by ordinary people in their daily life, although such practice is not totally based on scientific knowledge. Several scientific attempts have also been made by psychologists to examine the behavioral correlates of lying. Indices of this sort that have been investigated include measures of reaction time (e.g., Goldstein, 1923; Jung, 1906/1973, 1910; Marston, 1920), interference with motor behavior (Luria, 1932; Runkel, 1936), changes in voice quality (Alpert, Kurtzberg, & Friedhoff, 1963; Horvath, 1978), and changes in frequency of other
nonverbal behaviors, such as postural changes, eye contact, eye movements, gestures, facial expressions, and speech hesitation (e.g., Ekman, 1985; Ekman & Friesen, 1969, 1974). However, such behavioral indices are under voluntary control to a significant degree, which severely limits their potential usefulness for detecting deception (Furedy, 1986). Indeed, De Paulo and Pfeifer (1986) had their participants try to detect deception from verbal and non-verbal cues and found that very few participants had ever achieved an accuracy level greater than 60 percent (in tasks for which 50 percent accuracy would represent a chance level), and some even performed worse than chance.

In contrast, small changes in psychophysiological functions, such as skin conductance, blood pressure, respiration, and heart rate, are usually unnoticeable to human participants but are recordable through surface electrodes. This property minimizes the potential problems associated with the participant's awareness of both his or her intention to deceive and of learned voluntary behaviors taken as signs of honesty, such as looking straight into the questioner's eyes. Hence, by elimination, the psychophysiological method appears to offer the only feasible way to detect deception in human communication to the extent that such detection is possible (Furedy, 1986).

Indeed, the psychophysiological detection of deception has been a major research area in psychophysiology since the early 1960s (e.g., Davis 1961; Gustafson & Orne, 1963, 1964;
Kugelmass, Lieblich, & Bergman, 1967; Lykken, 1959, 1960; Orne, Thackray, & Paskewitz, 1972). It has been of major applied significance, because testing of this sort would bring a potentially serious consequence into the criminal justice system. However, a consensus on the scientific validity of the psychophysiological detection has not yet been established among psychophysicologists.

According to Trovillo (1939), basic research into modern psychophysiological detection of deception can be traced to the late nineteenth century, when Lombroso, an Italian criminologist, several times assisted the police in identifying criminal suspects through the use of blood pressure, which had become possible with the development of equipment that would measure psychophysiological functions relatively unobtrusively (Furedy, 1986). Later in the United States, Hugo Münsterberg (1908), a well-known psychologist at Harvard, suggested the application of psychophysiological measurement techniques such as blood pressure recording to forensic problems. One of his American students, Marston (1917), conducted a psychophysiological detection study and reported a 96% accuracy rate using blood pressure as the single measure. Lykken (1981, 1998) credited Marston as being the first to introduce the notion of a "specific lie response," which could be detected by a unique psychophysiological pattern of responding, a notion now discredited by many "detecting-deception" opponents (e.g., Ben-Shakhar & Furedy, 1990; Ben-Shakhar, Lieblich, & Bar-

Subsequent to Marston, several interrogation methods for psychophysiological detection of deception have been developed and are currently used in field practice (Saxe, Dougherty, & Cross, 1985).

According to Lykken (1998), thousands of Americans are subjected to polygraph or "lie detector" tests each year; and countries such as Canada, Israel, Japan, South Korea, Mexico, Pakistan, the Philippines, and Thailand use the polygraphic examination in criminal investigation and security applications. The United States, however, is the largest consumer of psychophysiological detection of deception, and the place where the most traditional and controversial interrogation method called Control Question Test (CQT) has been employed.

**General Description of the Control Question Test (CQT)**

The CQT was developed by John Reid and his associates in 1947 for use in criminal investigations (Reid & Inbau, 1977). Several interrogation methods exist and are currently used in various areas. A central feature of these methods is mostly exemplified in the CQT, and its nature has been the topic of considerable and often vitriolic debate in the scientific
community (e.g., Lykken, 1978 versus Raskin, 1978; Lykken, 1979 versus Raskin & Podlesny, 1979; Furedy, 1993, 1996 versus Honts, Kircher, & Raskin, 1995; Furedy & Ben-Shakhar, 1993 versus Honts, 1993). The basic assumption underlying the CQT is that an examiner can detect deception in an individual by comparing his or her physiological responses to questions relating to the crime or other critical events of interest and to questions not directly relating to the crime but still causing high alertness in most examinees. On the basis of this assumption, three types of questions are administered to an examinee, while his or her psychophysiological responses (e.g., electrodermal response, plethysmographic response, and respiration) are concurrently recorded. These three types of questions are referred to as relevant, irrelevant (or neutral), and "control" questions. The relevant questions refer directly to the critical incident under investigation (e.g., "Did you steal the money from Mr. Smith's drawer last Friday night?"), whereas the irrelevant questions are unrelated to the incident (e.g., "Were you born in Japan?"). The control questions are designed to be unrelated to the incident under investigation, but are assumed to elicit at least as much emotion as the relevant questions do for an innocent suspect (e.g., "Apart from the present incident, did you ever take something that did not belong to you?"). If the psychophysiological responses are generally greater to the relevant questions, then the examinee is considered deceptive on the relevant
questions, and hence guilty of the crime; if the responses are greater to the control questions, the examinee is considered truthful on the relevant questions, and hence innocent of the crime. If there is no consistent difference in either direction, an inconclusive decision is given (Raskin, 1989). Before noting possible problems with the inference involved in this rationale, the whole CQT procedure should be briefly described because there are several other scientific problems inherent in the CQT procedure other than its inference rules.

A typical CQT procedure, as a whole, can be divided into four phases, namely: a pretest interview, administration of CQT polygraph "tests," a scoring phase, and a posttest interview (Ben-Shakhar & Furedy, 1990). The first phase is a pretest interview which lasts 30-60 minutes. During the interview, both the relevant and control questions are formulated through a discussion between the examiner and examinee. The relevant questions are reformulated until the examinee indicates that he or she finds them unambiguous, and can clearly answer "no" to them (i.e., indicating his or her innocence). The control questions are reformulated until a version is arrived at for which, in the examiner's view, the examinee's answer "no" is either totally deceptive or at least not confidently truthful. In other words, it leads the examinee to believe that admissions will cause the examiner to form the opinion that he or she is dishonest and is therefore inclined to guilty. Raskin (1989) summarized the
relationship between the relevant and control questions in the CQT as follows:

Control questions are designed to give an innocent suspect an opportunity to become more concerned about questions other than the relevant questions, thereby causing the innocent suspect to react more strongly to the control than to the relevant questions. If the subject does react with greater strength to the control questions, the result is interpreted as truthful. On the other hand, stronger reactions to the relevant questions are interpreted as indicating that the subject was deceptive to the relevant questions. The problem of no "specific lie response" is circumvented by the procedure of drawing inferences about truth or deception by comparing the relative strength of this particular subject's reactions to relevant and control questions. (p. 253)

The second phase is a "test" phase which lasts for about half an hour and is begun by connecting the physiological recording instrument to the examinee. The test phase proper is often preceded by the card or "stim" test which is designed to convince the examinee of the validity and infallibility of the polygraph examination as a lie detector. In this card test the examinee is asked to choose a card from a deck. The examiner then calls out the names of several cards to which the examinee is asked to give negative answer. While doing so, the examiner pretends to evaluate the polygraph recording. After the last call is made, the examiner correctly informs the examinee which card he or she chose and asserts that the polygraph revealed the choice.
However, the deck of cards is prearranged so that the examiner knows exactly which card has been picked by the examinee.

Usually, a sequence of questions consists of 10 questions, three pairs of the relevant and control questions, with the remainder being the irrelevant questions. The irrelevant question serves as an initial buffer designed to habituate the reaction that normally occurs to whatever the question is. These questions are presented about 30 sec apart and each repetition through the list is called a "test." The test is repeated at least three times, after that the examiner can decide whether or not to give one or two more.

When the examiner has decided to stop giving the "test" phase, the third phase is initiated. During this phase, the examinee is left alone for some 20 minutes, while the examiner leaves to score the psychophysiological records. Clearly the third phase has a role in the confession-inducing function of the polygraphic interrogation, because during this phase the examinee has little else to do except worry about whether he or she will be judged "deceptive" by the examiner.

If the examiner decides that the examinee has been deceptive, the examiner tries, upon returning, to induce a confession of guilt during the "posttest interview." This fourth phase can last from 10 minutes to several hours. That is, this last phase is terminated either by a confession or
by the examiner's decision that he or she will not be able to get a confession.

Problems Inherent in the CQT

There are several elements of the CQT that are of particular concern to its opponents (e.g., Ben-Shakhar, 1991; Ben-Shakhar & Furedy, 1990; Furedy, 1985, 1987, 1989, 1991, 1993; Furedy & Heslegrave, 1988, 1991; Kleinmuntz & Szucko, 1984a; Lykken, 1979, 1981; Saxe, 1991; Saxe, Dougherty, & Cross, 1985). They are summarized as follows: (a) the ambiguous meaning of "control" questions, (b) lack of standardization of the method, (c) poor psychophysiological quantification, (d) contamination problems of diagnosis by factors other than psychophysiological responses, and (e) failure to separate detection of deception from confession inducement. These claims will be briefly reviewed below.

Ambiguous meaning of "control" questions. The first and foremost problem inherent in the CQT is that it employs questionable "control" questions which do not truly constitute control in the scientific sense of the term (Ben-Shakhar & Furedy, 1990; Lykken, 1981). The normal sense of the term "control" in an experimental/control comparison is that the control condition is identical in every respect to the "experimental" condition, except for the critical difference being studied. If the purpose of the CQT administration is to detect deception and this is accomplished by comparing responses to the relevant
(experimental) and "control" questions, then the only difference between these two types of questions should be presence and absence, respectively, of deception. However, this is not the case in the CQT or in any other psychophysiological detection technique. As mentioned earlier, the differentiation of deception from other psychological processes can be accomplished only by the Differentiation-of-Deception Paradigm (Furedy et al., 1988).

Even if the target for detection is not the process of deception, but the guilt of the examinee, the relevant/control comparison does not constitute control in the normal scientific sense of the term, because the two types of questions differ on a number of dimensions besides that of whether or not the examinee is guilty or innocent. One such confound is a different arousal level supposedly introduced in the examinee by the relevant and control questions. In the CQT, it is relatively easy to identify which questions are the relevant ones because during the pretest interview, the examinee has thoroughly learned about the event under investigation, and then only the relevant questions seem to place the examinee in serious jeopardy, whether the examinee is innocent or guilty. In such a threatening situation, especially for a naive innocent person, one might suppose that everyone would respond more strongly to the relevant questions, meaning that all examinees, honest or deceptive, should fail the CQT. Indeed, reviews of the literature have revealed false-positive rates
(i.e., the rate of innocent suspects classified as guilty) for the CQT as high as 50%, with a mean in the mid-30% range (see Ben-Shakhar & Furedy, 1990; Kleinmuntz & Szucko, 1984b; Lykken, 1974, 1979; Saxe et al., 1985). This not only raises the issue of test validity but also the issue of the ethics of the examiners.

In summary, then, no control in the normal, scientific sense of that term is involved in the CQT, hence the use of the term in the CQT is misleading.

Lack of standardization of the method. Because the technique is referred to as the Control Question "Test," its standardization ought to be a minimum requirement. Psychological tests are usually standardized in two important ways (Blinkhorn, 1988). First, the test is the same for everyone who takes it. Second, an examinee's score is compared with population norms, in order to render the results of the test interpretable. However, neither criterion can be accomplished in the CQT, because the relevant questions are event- or crime-specific and, hence, they are different from case to case. Even if the CQT is used with the same type of crime with the same type of relevant questions for different suspects, then the "control" questions are still different from suspect to suspect, because of the manner in which the control questions are formulated. Thus no standard norm can ever be established.

Formulating appropriately sensitive control questions seems to be extremely difficult, and creates an obstacle to
standardization of the test. The whole logic of the method depends on the ability of the examiner to make the "control" questions look more threatening for innocent suspects and to make the relevant questions look more threatening for the guilty, and this is carried out through a dynamic examiner-examinee interaction during the pretest interview. This means that the validity of the procedure will be difficult, if not impossible, to assess even in principle (Ben-Shakhar & Furedy, 1990).

Advocates of the CQT have recently recommended the use of a purportedly more standardized form of the CQT, called the Directed Lie Test (DLT; e.g., Honts et al., 1995; Raskin, 1989). In the DLT, directed-lie questions like "Have you ever told a lie?" or "Have you ever made a mistake?" are introduced instead of so-called "control" questions. To each directed lie question, the examinee is requested to give an answer "no," and it is made clear that, for example, anyone who denies having ever lied in the past would be lying, which is reflected on the tracings of the polygraph. The DLT assumes that an innocent examinee will be more disturbed when instructed to answer falsely about some past misdeed than when truthfully denying an accusation of which he or she now is suspected of having perpetrated. The DLT is scored in the same way as the CQT.

Advocates believe that the DLT is an improvement because there is greater certainty that the examinee's answers to directed-lie questions are false. Moreover, they claim that
the DLT is a more standardized test than the CQT because the same directed-lie questions can be used in every case. However, the assumption of the DLT, that is, the innocent examinee is more disturbed when answering falsely to the directed-lie questions, seems very moot because one can never know whether the examinee is really more disturbed by the questions. Indeed, it seems likely that naive and innocent examinee will show strong responses to the relevant questions, not to the directed-lie questions, because the examinee is well aware that the relevant questions are more important for the outcome of the test. Moreover, the psychological impact of the directed-lie questions (e.g., "Have you ever lied?") seems less than the impact of the "control" questions in the CQT (e.g., "Apart from the present incident, did you ever take something that did not belong to you?"), and even much less than the impact of the relevant questions (e.g., "Did you take money?"). Thus the result of the increased use of the DLT, in the name of a more standardized lie detection method, might be an increased number of false-positive classifications.

Poor psychophysiological quantification. An important aspect of the polygraphic examination is the interpretation of the physiological responses. Currently, there are two types of scoring method used: qualitative and semi-quantitative. The qualitative method consists of simply inspecting the shape of the responses and deciding whether the person has been deceptive by generally responding more to
relevant questions (see Reid & Inbau, 1977). This impressionistic method would be particularly effective if there were "specific lie responses" that would be easily identified. However, this is not the case. Unfortunately, the method is still employed by most field polygraphers (Furedy & Heslegrave, 1991).

The semi-quantitative method, which has been recommended by the CQT advocates (see Barland & Raskin, 1975), assigns numbers (ranging from -3 to +3) to each pair of relevant/control questions for each physiological measure, depending on the relative magnitudes of the responses to each question. If the relevant question yields a greater response, then the sign of the number is negative; if the reverse is true, the sign is positive. The scores are summed, using question pairs (usually three), physiological measures (three or four), and at least three "tests." If the absolute value of the sum exceeds 5, then, depending on the sign of number, the examinee is classified as deceptive (negative sign) or truthful (positive sign). If the scores falls between -5 and +5, the examinee is classified as "inconclusive."

One basic problem of the method is that the score (ranging from -3 to +3) is assigned by subjective and qualitative means. Secondly, the setting of the cutoff point (-6 or +6) for inconclusive is arbitrary. In addition, there is no allowance for number of physiological measures and number of tests. Thus, the chances of scoring an examination
inconclusive could be decreased as a function of the sum of the number of channels used and tests administered. These arbitrariness features severely weaken the objectivity of the CQT.

To counter these claims, a computerized evaluation system for polygraph data was developed by Kircher and Raskin (1988). However, Furedy (1987) described it as "garbage-in, garbage-out." If the CQT process itself is unreasonable, the output of the computer is still unreasonable. Moreover, the system does not seem to subdue the inherent problems of the CQT, although it could increase its face validity value.

Contamination problems in diagnoses. Since the beginning of polygraph use, polygraph examiners had used "global evaluation" (e.g., Reid & Inbau, 1977). The same individual reads the case file, conducts the pretest interview, formulates the questions, administers the tests, observes the "behavior symptoms," evaluates the polygraph charts, and then reaches the final diagnosis of deceptive or truthful (Lykken, 1998). Hence, during the course of the procedure, a great deal of complex information is available to the examiner, and it is impossible to differentiate between the impressions formed by this prior information and those gained from the purely physiological data obtained during the so-called test phase of the polygraph interrogation procedures. This feature, which characterizes all types of polygraph-based interrogation procedures, but in particular the CQT, has been labeled "contamination" (Ben-
Shakhar, Bar-Hillel, & Lieblich, 1986), meaning that judgments and conclusions derived from the physiological information are contaminated with various kinds of non-physiological information.

The contamination problem is particularly severe when the interpretation of the physiological charts is impressionistic and subjective. When no clear a priori rules of chart interpretation exist, the interpretation depends upon the judgment of the examiner, who cannot ignore the prior information that was provided.

Another possible contamination problem is that the non-physiological information may influence the manner in which the polygraph interrogation is conducted. In the course of pretest interview and/or when reading the case file, it is likely that the examiner forms an opinion of whether the examinee is guilty or not, and then the examiner's opinion may affect the way the interrogation is conducted. In addition, and perhaps more significantly, a prior impression that the examinee is probably guilty can lead to the examiner formulating relatively ineffective control questions. Such an "experimenter expectancy effect" has been well documented (e.g., Rosenthal & Rubin, 1978) both in laboratories and real-life situations; and it is generally demonstrated that expectations formed by an experimenter may influence the behavior of subjects.

Because the CQT includes the pretest interview as an essential part of procedure, in which so-called control
questions are formulated through the interaction between the examiner and examinee, it would be impossible to overcome these contamination problems.

**Failure to separate detection of deception from confession inducement.** According to practicing polygraphers (e.g., Lee, 1953), an important function of the polygraph is not only its use as a detector of deception, but also its use to induce a confession. This raises two basic problems concerning the CQT; one is ethical and the other is scientific.

The ethical problem is that the confession-inducing function can be administered independent of the result of psychophysiological detection. Even if an examiner is not sure that an examinee is lying, that doubt is resolvable by pressing the examinee into a confession of guilt, asserting that the 'machine' indicates that the examinee has been lying (Furedy & Heslegrave, 1991). This is particularly likely with the unstandardized CQT, in which an examiner is induced to form his or her own opinion for an examinee through an extensive interaction during the pretest interview, the case record reading, and the behaviors of the examinee during the interview. Moreover, there is the logical possibility that certain confessions obtained by such "psychological-rubber-hose" (Furedy & Liss, 1986) or "fourth-degree" (Lykken, 1981) procedures may be false. In fact, many false confessions have been made by innocent persons in criminal cases without a "third degree" procedure (Gudjonsson, 1992).
The scientific problem is that one can never estimate the accuracy of the CQT as a pure psychophysiological detection technique, as long as the confession-inducing function is regarded as a part of the CQT procedures. In other words, it is impossible to distinguish whether the final judgment by the polygrapher on an examinee's guilt has been provided by confession or physiological recordings, or both.

Accordingly, the CQT is not a scientific-based technique to detect deception or to discriminate between honest and deceptive individuals. The only way to overcome the difficulties mentioned above is by completely decontaminating the procedures, using a standardized interrogation method and objective measurements (Ben-Shakhar, 1991; Ben-Shakhar & Furedy, 1990). However, it seems impossible to decontaminate the CQT because its theoretical basis is scientifically unsound and obscure.

**General Description of the Guilty Knowledge Test (GKT)**

An alternative and less controversial detection technique is the Guilty Knowledge Test (GKT), also known as Concealed Information Test (CIT), which is commonly used in laboratory experiments but less in field practice. The GKT was developed by Lykken (1959, 1960) based on entirely different assumptions from the CQT as follows: (a) The purpose of psychophysiological detection in criminal cases is detecting guilt rather than detecting deception, (b) a guilty
person possesses specific information concerning a crime which is not usually available to innocent persons (e.g., guilty knowledge), and (c) recognition of pieces of information unique to the crime will lead to an enhanced physiological response to them only in the guilty person.

To satisfy these presumptions, the GKT utilizes multiple-choice questions, each having one relevant alternative (e.g., a true feature of a crime) and several neutral, irrelevant alternatives. The alternatives are chosen so that an innocent person with no guilty knowledge cannot discriminate among them (Lykken, 1981), for example, "Was the weapon used in the particular case a: A) gun, B) knife, C) lead pipe, D) club, or E) baseball bat?" Since it is assumed that the true weapon is known only to a person involved in the crime, this weapon has a special meaning only for this person. If the person's physiological responses to the relevant item are consistently larger than to the neutral ones, knowledge about the event in question (e.g., crime) is inferred. In contrast, an innocent suspect with no guilty knowledge should not respond differently to the alternatives.

Naturally, a single presentation of one question with a few alternatives would not be sufficient because an innocent suspect might then show greater response to the correct alternative just by chance (i.e., a false-positive error). To prevent such an occurrence, the GKT uses several different questions that focus on different features of the crime (e.g., the kind of clothes the victim wore, the amount of
money stolen) and by frequently repeating each series of questions. With only 10 such guilty knowledge items, each with five good alternatives, there would be only about 1 chance in 10 million that the suspect without guilty knowledge would provide his or her largest response to the correct alternative in all 10 items (Lykken, 1974).

The GKT as A Scientific Detecting Method

Unlike the CQT, the GKT rests on sound psychophysiological and psychological principles as a standardizable test or detector of guilt, although it does not detect deception directly. Advantages of the GKT as a scientific method over the CQT are summarized as follows: (a) use of sound scientific "control" questions, (b) ease of standardization of the method, and (c) a less contaminated procedure. These advantages will be briefly described below. Scientific control questions. The control questions in the GKT are formulated exactly like the relevant ones. The only difference between the relevant and control questions is that only the relevant questions refer to true features of the event under investigation. Therefore, only people possessing such knowledge (i.e., guilty knowledge) are capable of distinguishing relevant from control questions. The recognition then leads to differential psychophysiological responsivity to the relevant questions only by the guilty person. This aspect is not accomplished in the CQT, in which all questions, both the relevant and control questions, are
serious to an examinee but their degrees of difference are never known.

**Ease of standardization.** In the GKT, the questions are determined by the feature of the event under investigation, and they do not depend on the examiner-examinee interaction like in the CQT. It is possible that two different examiners would construct different sets of questions from the features of an event, but these two GKT procedures may be viewed as two equivalent forms of a vocabulary test, each containing different items sampled from the set of words (Ben-Shakhar & Furedy, 1990). In addition, the set of questions in the GKT can be prepared in advance on the basis of a thorough examination of the event and even the same set of questions can be applied to different examinees. Such a procedure is more standard and objective than is the case in the CQT.

**Less contamination in the procedure.** As mentioned previously, the whole CQT procedure requires an extensive examiner-examinee interaction, especially in the pretest interview, and this leads to contamination in the examiner's final judgment whether the examinee is guilty or not. However, the GKT does not require the pretest interview to construct questions and thus the examiner can be freer from any bias against the examinee. Moreover, the GKT can be completely decontaminated from tester biases by using blind interrogation or even prerecorded questions (Ben-Shakhar & Furedy, 1990; Lykken, 1981).
In summary, the GKT enables scientific exploration of psychophysiological detection of deception, although it purports to detect guilty knowledge rather than deception itself. Unfortunately, however, it has not been used widely in the field practice of criminal investigation in North America. Only Israeli and Japanese police polygraphers are frequent users of the GKT, and in the latter case, around half of some 5,000 annual polygraph examinations have been carried out with the GKT (Ben-Shakhar & Furedy, 1990). The lack of field use of the GKT is partly due to the reasonably strict requirements of the test to keep important details of the crime from being revealed to the public in general and to the various suspects in particular. A requirement of the GKT is that an examiner be available who is well informed about the crime so that he or she can find the "keys," that is to say, the items of information that only a guilty suspect would recognize and that could be used as GKT questions (Lykken, 1991).

In laboratory studies, the GKT has proven to be quite reliable in detecting guilt (e.g., Ben-Shakhar, Lieblich, & Kugelmass, 1970; Davidson, 1968; Dawson, 1980; Furedy & Ben-Shakhar, 1991; Gustafson & Orne, 1963, 1965; Lieblich, Ben-Shakhar, & Kugelmass, 1976; Lykken, 1959, 1960). Ben-Shakhar and Furedy (1990) provided the average accuracy rates after reviewing 10 GKT laboratory studies, which had been originally designed to examine the validity of the method. The accuracy rate for guilty participants was 84%, and 94%
among those simulating the innocent persons. They also provided the average accuracy rates of the CQT in both laboratory mock crime studies and field studies. The results of the 9 mock crime studies revealed 80% correct classification of the guilty participants and 63% of the innocent persons, while the review of 9 field studies provided the accuracy rates of 84% for the guilty suspects and 72% for the innocent suspects. It should be pointed out that both the GKT and CQT correctly identified guilty subjects around 80% of the time, but the CQT's accuracy rates for innocent persons were substantially lower compared with the GKT. This means relatively high false-positive errors occurred with the CQT. Overall, Ben-Shakhar and Furedy (1990) concluded that the GKT is the method which can best protect the innocent and recommended abandoning the CQT.
SCOPE OF THE STUDY

In the psychophysiological detection of deception, the GKT is apparently a far more objective way of distinguishing between guilty and innocent suspects, but less frequently used in field practice than the more controversial CQT. To the extent that applied psychology professes itself to be based on scientific knowledge, standardization of the method and explanation of its underlying mechanism are essential requirements. In an effort to contribute towards a fuller understanding and standardization of the GKT, two different experiments were carried out to examine accuracy of the GKT in a laboratory situation where participants role-played a crime scenario.

Experiment I was a conceptual replication of a GKT laboratory study by Nakayama, Mizutani, and Kizaki (1988), who reported that detection was superior when participants were required to delay their answers for 8 seconds rather than answer immediately as is customary. In their experiment, the GKT questions were presented in a visual mode on computer screen—a more standardized procedure than the usual one in which a human examiner asks the questions. If the result by Nakayama et al.'s (1988) study is again confirmed in another experiment, a simple procedural change requiring delayed-answer would produce improved detection.

In Experiment II, an auditory, computer-synthesized voice presentation of the GKT was employed, and the participant's emotional reaction to the relevant items was
manipulated in hope of changing their "signal values" (e.g., Lykken, 1974). In addition, to examine an orientation and habituation account of the GKT effect (e.g., Ben-Shakhar, 1977; Lieblich, Kugelmass & Ben-Shakhar, 1970; Lykken, 1974; Raskin, 1979), a simple orienting response (OR) paradigm for obtaining participants' skin conductance response (SCR) characteristics to repeated stimuli was carried out prior to the detection phase, and several components of OR were compared to the GKT effects.

In both experiments, SCR were used as a dependent variable. The SCR has been proven to be the most efficient psychophysiological measure, both in laboratory studies and field practice of detection of deception (e.g., Barland & Raskin, 1973; Cutrow, Parks, Lucas, & Thomas, 1972; Elaad & Ben-Shakhar, 1989; Furedy & Heslegrave, 1988; Podlesny & Raskin, 1977; Thackray & Orne, 1968; Waid, Orne, Cook, Orne, 1981).
EXPERIMENT I

The primary purpose of the Experiment I was to provide a conceptual replication of a GKT laboratory study by Nakayama et al. (1988). They reported, with SCR as dependent measure, not only that the delayed-answer condition was superior to detect guilty, but also that the immediate-answer condition failed to yield a reliable GKT effect.

The delayed-answer condition was originally introduced by Dawson (1980) in his laboratory version of the CQT to separate two hypothetical processes involved in the deception phenomenon, namely the intention to deceive and the act of deception. Thus, psychophysiological responses were measured twice in the delayed-answer condition: one immediately after question onset and the other immediately after the participant's answer. It was conceptualized that the former measure was an index of the intention to deceive, and the latter was an index of the act of deception. Dawson (1980) found a greater difference between relevant and control questions when electrodermal responses to the question onset were used as a detection measure, and hence concluded that it was the intention to deceive rather than act of deception that was critical for the detection of deception in the CQT. However, there was no difference between the immediate-answer and delayed-answer conditions when the electrodermal responses to the questions were compared.

Later, Furedy and Ben-Shakhar (1991) adopted the delayed-answer condition in addition to the conventional
immediate-answer condition to study the role of deception in the GKT. They found equally significant detection rates when participants' SCRs were obtained immediately following question presentation, whether the participants had answered immediately or had delayed their answers. In contrast, the differential SCR to the delayed answer was markedly attenuated. Based on their results, Furedy and Ben-Shakhar (1991) also suggested that the crucial factor determining differential responsivity to the relevant and neutral items is the intention to deceive rather than act of deception.

Contrary to other laboratory GKT studies, Nakayama et al. (1988) failed to obtain a reliable GKT effect in the immediate-answer condition, but obtained a significant GKT effect in the delayed-answer condition. Hence they recommended the use of the delayed-answer condition in the field practice of the GKT to produce improved detection. Their presentation method of GKT questions, however, varied from the normal practice in several ways. First, the GKT questions were presented on a computer screen in a visual, rather than auditory mode. Thus it seems difficult to generalize the result of the Nakayama et al. study into field practice in which a human examiner presents questions orally. The second deviation in their method was that the question remained on a computer screen for 8 s, no matter whether the participants had answered immediately or had delayed their answers for 8 s. In the immediate-answer condition, therefore, the GKT question would have remained on screen
long after the participant had answered it, and this raises the possibility that, at least for some participants, the immediate-answer condition may have seemed to lack sense, and could have accounted for these investigators' failure to find any reliable GKT effect under this conventional condition. In other words, some participants in the immediate-answer condition may not have treated their task as seriously as the participants in the delayed-answer condition, because the questions were remaining on without any specific sensible purpose. Moreover, the participants could have answered without paying any attention to the very onset of question. To prevent this problem, Nakayama et al. requested their participants to repeat the name of item that was presented on the computer screen as an interrogatory. Aside from speculative nature of their attentional hypothesis, it is not clear that the word repetition procedure eliminates the attentional difficulty confound. More importantly, however, it does not deal with the original problem of the immediate-answer version of the task.

To make the computerized GKT procedure more like the normal oral arrangement, the present experiment introduced an additional immediate-answer condition, where the questions disappeared from a computer screen as soon as participants gave their oral answers (Immediate-Answer-Contingent condition). This was carried out by requesting participants to hit a computer key simultaneously when they gave their oral answers, which means the duration of questions on
computer screen was depending on participants' oral answers. The answer-dependent procedure was also applied to the delayed-answer condition where participants were required to delay their answers for 8 s and then hit a computer key to erase questions (Delayed-Answer-Contingent condition). In addition to these modified conditions, the original arrangement by Nakayama et al. of the question being on screen for 8 s was carried out both in the immediate- and the delayed-answer conditions (Immediately-Answer-Non-contingent and Delayed-Answer-Non-contingent conditions, respectively).

In the light of the Nakayama et al. (1988) study, the present experiment explored these independent variables to shed more light on their role in the GKT effect.

**Design of Experiment I**

A $2 \times 2 \times 2 \times 2$ (Question Type x Interval x Contingency x Stimulus Type) mixed-design with repeated measures on Question Type and Stimulus Type was used. The Question Type variable refers to the relevant and neutral questions: the variable the most critical in the GKT study, which was a 'within' factor. The Interval variable refers to the immediate- versus delayed-answer conditions, and the Contingency variable refers to the duration of the question on computer screen, which was controlled by the participant's key response. These two factors were examined as 'between' factors and hence generated a total of four different groups of participants described above. Finally, the Stimulus Type
variable refers to the two types of stimulus which were provided for each question item; one was the question itself and the other was a tone stimulus presented 8 s after the question onset. The latter originally served as a prompt to give oral answers by participants in the delayed-answer groups but was also presented to the immediate-answer groups just to equalize experimental conditions. Thus this variable was also a 'within' factor. To examine these independent variables, SCR was used as a dependent measure.
Method

Participants

The participants were 24 male and 24 female volunteers, aged 19-39 years, from the Introductory Psychology class at the University of Toronto. Prior to obtaining their consent, it was explained to the participants that they were being asked to play the role of a burglar in the experiment. All of them were willing to play this role. None of the participants had previously taken part in a detection-of-deception study. They were randomly assigned to one of four groups of 12 participants each with a restriction that each group had 6 female and 6 male participants.

Apparatus

The participants were tested in a sound attenuated one-way vision room, which was described as "interrogation room" and was located adjacent to the experimenter's room; it was lit with a 100 w light situated overhead and to the back of the participant.

The participants were seated on a one-armed chair in front of a lap-top computer (Toshiba J3100-GT) with 9 inch monitor and a keyboard and two outside speakers. Instructions about the procedure as well as all questions during the detection period were provided in visual mode on the computer screen. The computer also produced a beeping tone during the matching trial, where participants' orienting responses to simple tone stimuli were evaluated. The tone
was 80 dB at speaker level, 2500 Hz, and 800 ms in duration. It was presented five times with a fixed inter-stimulus-interval of 30 s. The computer program controlling the experimental procedure was written in Microsoft Quick Basic 4.5.

Continuous skin conductance was recorded with a Coulbourn Instruments preamplifier (S71-22) on a Narco Bio-Systems MK-III-S (Physiograph®) polygraph running at a chart speed of 0.5 cm/s. Two Ag/AgCl electrodes (8-mm in diameter), in conjunction with commercial electrolyte from Teca Corp, were attached by double-sided adhesive collars to the volar surfaces of the participant's distal phalanges of the first and third fingers of the left hand, which had been cleaned with Isopropyl rubbing alcohol.

Procedure

When participants arrived for the experiment, they were greeted by the experimenter and then given general information concerning the nature of the experiment. They were told that the purpose of the experiment was to examine the reliability of a computer-assisted detection of deception (lie detection). Their task was to play, in an imagined crime situation, the role of burglar who faced a polygraph test he or she was trying to beat in order to be released by the police. The participants were also told that two electrodes would be attached to two fingers of their left hands before starting the experiment. It was further
explained that these were only recording electrodes for detection of deception, and the participants would not experience any discomfort from the electrodes. Finally, participants were told that the experiment itself would last for approximately 30 minutes.

Following general information on the running of the experiment, each participant signed an informed consent which stated that the participant was aware of the nature of the experiment and that he or she could terminate the experiment at any time with impunity.

The participants were then led into a room described as the "interrogation room," adjoining the one used for the physiological recording by the experimenter. The participants were seated at a table facing the computer and then electrodes were attached for SCR measurement. Once the electrodes were in place, the participants were asked to rest the left hand on a soft pad and to keep it still and avoid unnecessary movement.

After the completion of electrode attachment, the participants were told that detailed instructions for the experiment and the description of the crime, for which the participants had to take a polygraph interrogation, would be given on the computer screen in written form. The participants were then asked to initiate the computer program by hitting any key on the computer once they were ready. While the participants were reading the instructions, the experimenter went to the experimenter's room to calibrate the
On the first screen of the instructions, the participants were reminded again to play the role of a burglar and were given a background why they had been brought into the "interrogation room" as suspects. Below are the instructions that were presented on the first screen.

As you already know, this experiment is about lie detection. The room you are now in is an interrogation room. Here, I would like you to play the role of a burglar who actually committed a burglary and has been under arrest for that reason.

But you are a tough criminal and you have not confessed to your crime during normal interrogation procedures. Although the police are convinced that you are the burglar, there is not enough material evidence that you are guilty.

So, the irritated police department decided to make you take a lie detection test, and you have agreed. That's why you are here.

If you can beat this lie detection test, you will be released.

Do you understand your role? If not, please ask the experimenter now.

Next, I'll explain the crime you committed. You don't have to memorize the details. Just try to grasp the situation in general.
On the bottom line of the screen, a note prompted the participants to advance to the next screen by hitting any key so that they could read instructions at their own pace. The second screen was a crime scenario that read as follows:

The crime concerned here is, of course, the burglary.

The other day, at midnight, someone broke into a house in a Toronto suburban area. After the police investigations, it was revealed that the burglar, maybe just one burglar, broke into the house and stole some amount of money and jewels. Fortunately, no one in the family was hurt because they all slept deeply. From the skillful way in which the burglary had been carried out, the police identified and arrested you as the burglar. You had three records of burglary before. But you were smart enough not to leave any material evidence this time.

As I said before, the police are convinced that you are the burglar, but it is also true that they have not enough evidence. You will be released if you can beat the following lie detection test. If you fail, you will still be kept in custody. Moreover, the result of the test will be used in your trial as an evidence. Thus, try to beat the following test!!!

Next, I'll explain the experimental procedures.

On the next six screens, the general interrogation procedures were explained to the participants and then different instructions about a manner of responding to the interrogatory questions were given, depending on which group each participant had been allocated. All participants were
instructed that they would be requested to choose one out of five spots thorough which they had intruded the house as a burglar. They were also told that the aim of the following polygraph test was to identify the spot that each participant would chose by means of examination of psychophysiological recordings. The participants then practiced responding to dummy questions in a way that had been explained differently for each group until they felt comfortable to do so.

After the completion of practice session, the participants were requested on the next screen to select one out of five spots which they had intruded a house as a burglar by referring to a layout sheet of the alleged house (Appendix A), which had been put face down besides the computer. On the sheet, design of a house and its layout with five alternative burglary spots marked was presented. These five alternatives were "1) Entry," "2) Family room," "3) Garage," "4) Kitchen," and "5) Living room." The participants were requested to type the number key on the computer keyboard which corresponded to the selected spot after the experimenter left the interrogation room. To confirm the selected spot, the participants were asked to type the same number key again on the next screen. No one failed to type the corresponding key twice in succession.

Following the choice of burglary spot, the participants were requested to initiate the five minute baseline period by hitting any key once they were ready. During the baseline period, the participants were asked to sit quietly at ease
for five minutes during which the computer screen went blank. When the participant hit a key to initiate the baseline period, the experimenter turned on the recording apparatus. In the middle of the baseline period, the tone was presented five times by the computer, separated by fixed intervals of 30 s. This procedure, the matching trial, would later permit assessment of the group differences in skin conductance responses to simple stimuli.

After the five minute baseline period passed, the computer let the participants know with several beeping tones that the baseline recording had just finished and asked them to initiate the detection period once they were ready. On the same screen, the participants were reminded of the break-in spot they had previously chosen.

In the detection period, each question was presented with a font size of approximate 14 points at the center of the computer screen, asking "Did you break in through the entry?" There were five different questions in which the name of burglarized spot was replaced among five alternatives mentioned above. All the participants were asked to deny every question by saying orally "No, not the entry," in which the name of the spot had to be changed correspondingly to the question. Because the experimenter was in the adjoining room to monitor their physiological responses during the detection period, the response method, namely, denying the spot orally rather than simply say "no" as usual, gave the experimenter feedback about the spot in the question and enabled him to
mark it on the recording chart when the detection process was under way. This procedure also did not let the participants neglect the question because they had to name the spot in their answer to each question.

For the participants in the Immediate Answer-Non contingent (IANC) condition, each question remained for 8 s on the computer screen after its onset and disappeared automatically with a computer beep (70 dB SPL, 2500 Hz and 800 ms in duration). The participants were requested to respond immediately following each question although the question remained on the screen for awhile. In the Delayed Answer-Non contingent (DANC) condition, again each question remained for 8 s on the screen after its onset and disappeared automatically with a computer beep. However, the participants in this condition were requested to respond only immediately after the computer beep; in other words, the participants had to delay their response for 8 s. These two conditions were exact replications of conditions used in the Nakayama et al. (1988) study.

In the Immediate Answer-Contingent (IAC) condition, the participants were instructed to answer immediately after question onset and hit a space bar at the same time to get the next question. After 8 s from question onset, a computer beep was provided in the same way as in previous two conditions. This procedure corresponds more closely to the non-computerized, oral presentation arrangement because a question would disappear rather than stay long after the
participants had given their answer. For the participants of the Delayed Answer-Contingent (DAC) condition, each question remained for 8 s after its onset as in the DANC condition. After 8 s passed, the participants were prompted to give their oral response by a computer beep and requested to hit a space bar to get the next question at the same time.

A series of questions was constructed from the 5 spots described above and repeated six times in each participant. The questions were presented at a fixed interval of 30 s and there was a 60 s pause between each series. The questions were presented at random in a series with a restriction that the first question in each series would not be the relevant spot. Thus, the first question in each series was treated as a buffer and would be excluded from the data analysis.

Finally, as an invasion spot, 4 participants chose Entry, 10 participants chose Family room, 22 participants chose Garage, 8 participants chose Kitchen and remaining 4 participants chose Living room during the spot-selection period.

**Dependent Measures**

SCR was defined as any response that was initiated (showing an inflection point) within 1 to 5 s following stimulus onset and producing at least a 0.5 mm change on the chart. Magnitude (in mm) was expressed as the difference between response onset (inflection point) and the highest point following response onset until the end of the 1 to 5 s
latency window. SCRs with a change of less than 0.5 mm on the chart were scored as zero responses. In cases where a composite of curves occurred, the measure taken was from the first curve. In the matching trial, SCRs were measured following the simple tone onset. Because the simple tone was delivered five times in the matching trial, each participant produced five SCR scores. In the detection period, SCRs were measured twice for each question both in the Immediate and Delayed Answer condition—first, after question onset, and second, after tone onset, which was originally introduced to prompt the participants in the Delayed Answer condition to give their answers. Before statistical treatment, the SCR scores were range transformed (cf. Davey & Singh, 1988; Lykken, 1972) by dividing every score obtained by a participant with the maximum score obtained that participant during the experiment.

In addition, an attempt was made to classify participants as "guilty" or "innocent" using a procedure proposed by Lykken (1959), although the participants in this experiment were all guilty of the imagined burglary. In a series of five questions, the first one served as a buffer. Thus, in remaining four questions, if the largest response occurred after the relevant question, a "2" was assigned on that question. If the second largest response occurred after the relevant question, a "1" was assigned on that question. In other cases, "0" was assigned. Thus, with six repetitions of the series of questions, a sum of 12 would strongly
indicate guilt, while 0 would strongly indicate innocence. For analytical purposes, those scoring 7 or more were classified as guilty, while those scoring less than 7 were classified innocent. In the Delayed-Answer condition, this scoring procedure was applied separately to the question-elicited responses and answer-elicited responses.
Results

Skin Conductance Responses during the Matching Trial

To assess differences in electrodermal responsivity among the four different groups, the participants' range-corrected SCRs to simple tone in the initial matching trial were compared. The tone-elicited SCRs (ORs) were analyzed by means of a $2 \times 2 \times 2 \times 5$ (Interval $\times$ Contingency $\times$ Sex $\times$ Trials) mixed-design analysis of variance (ANOVA) with Interval (immediate versus delayed), Contingency (contingent versus non-contingent), and Sex (female versus male) as between-subject factors, and Trials (5 trials) as a within-subject factor. An alpha level of .05 was set for statistical significance. Univariate tests for repeated measures were adjusted for violation of sphericity by using the Greenhouse-Geisser epsilon correction procedure (Fowles, Christie, Edelberg, Grings, Lykkken, & Venables, 1981).

Figure 1 shows habituation curves to five tone stimuli in the four different groups. The only significant effect was a Trials effect, $F(4, 176) = 42.56, p < .0001$ (epsilon = .742), and the Sex main effect was nearly significant, $F(1, 44) = 3.48, p = .069$, $M_s = .43$ and .33 for male and female, respectively. These results indicate that there were no differences in participants' electrodermal responsivity between groups, at least in response to simple tone stimuli. This preexperimental equivalence among the conditions permits further comparison of SCRs in the detection period among the four groups without considering initial group differences in
Figure 1. Mean change in skin conductance response (SCR) to simple tone during the matching trial in the Immediate-Answer-Contingent (IAC), Immediate-Answer-Non-contingent (IANC), Delayed-Answer-Independent (DAC), and Delayed-Answer-Non-contingent (DANC) groups.

Although it was not an essential purpose of the matching trial, SCR changes were also compared among the five trials by means of the Helmert contrasts in order to see whether complete habituation had occurred. The Helmert contrasts compared each trial mean of SCRs to the average of all subsequent trial means with respect to trial order. When the
mean SCR amplitude of the third trial was contrasted with the averaged SCR amplitude of the fourth and fifth trial, the significant main effect of Trials disappeared ($F < 1$), which suggest that complete habituation had occurred within the five trials.

**Skin Conductance Responses during the Detection Period**

During the detection period, each participant had 6 SCRs to relevant questions and 18 SCRs to neutral questions. These responses were averaged to provide one mean SCR score for relevant questions and one mean SCR score for neutral questions. In addition, there were also 6 SCRs to the tones that followed relevant questions and 18 SCRs to the tones that followed neutral questions. These SCRs were also averaged to provide one mean SCR score for each tone stimulus.

Figure 2 shows the range-corrected mean SCR changes to question onset and following tone onset in each group. These data were analyzed by means of a $2 \times 2 \times 2 \times 2 \times 2$ (Question Type x Interval x Contingency x Stimulus Type x Sex) mixed-design ANOVA with Question Type (relevant versus neutral) and Stimulus Type (question versus tone onset) as the within-subject factors, and Interval (immediate versus delayed), Contingency (contingent versus non-contingent), and Sex (female versus male) as the between-subject factors. The main effect of Question Type was significant, $F (1, 44) = 32.89$, $p < .0001$ (epsilon = 1), with larger SCRs to the
Figure 2. Mean change in skin conductance response (SCR) to relevant and neutral questions during the detection period in the Immediate-Answer-Contingent (IAC), Immediate-Answer-Non-contingent (IANC), Delayed-Answer-Contingent (DAC), and Delayed-Answer-Dependent (DANC) groups.

relevant questions than the neutral questions, which means that a reliable GKT effect was obtained in the present experiment, $M_s = .39$ and $.32$ for the relevant and neutral questions, respectively. The main effect of Interval was also significant, $F (1, 44) = 8.17, p < .01$, indicating that the two delayed-answer groups showed larger SCRs than the immediate-answer groups, $M_s = .31$ and $.40$ for the immediate- and delayed-answer group, respectively. The Stimulus Type
main effect of was also significant, £ (1, 44) = 30.72, p < .0001 (epsilon = 1), indicating that questions elicited larger SCRs than tone stimuli, Ms = .42 and .28 for questions and tone stimuli, respectively. The tone stimuli were originally presented to the delayed-answer groups to prompt their oral answers, but they were only dummy stimuli for the immediate-answer groups. Finally, the main effect of Contingency and Sex not only failed to reach significance (£s (1, 44) = .05 and 1.52 for Contingency and Sex, respectively), but it also did not interact significantly with the other main factors, and, hence was excluded from the following statistical analyses.

A three-way interaction among Questions Type x Interval x Stimulus Type was significant, £ (1, 44) = 8.16, p < .01. From the point of the present experiment, it is important to examine the difference of the GKT effect between the immediate- and delayed-answer conditions, hence a 2 x 2 (Question Type x Interval) mixed design ANOVA was conducted at question onset and at tone stimulus onset separately. Figure 3 shows the range-corrected mean SCR changes to the question onset (left panel) and the tone onset (right panel).

At question onset, the main effect of Question Type was significant, £ (1, 46) = 61.23, p < .0001 (epsilon = 1), indicating that a reliable GKT effect was obtained, Ms = .48 and .36 for the relevant and neutral questions, respectively. The main effect of Interval was also significant, £ (1, 46) = 12.25, p < .001, indicating that the immediate-answer
conditions elicited larger SCRs than the delayed-answer conditions, $M_{s} = .50$ and $.34$ for the immediate- and delayed-answer condition, respectively. Moreover, a Question Type x Interval interaction was significant, $F(1, 46) = 6.50, p < .02$, hence a comparison was made between the mean SCRs to the relevant and neutral questions for the immediate- and delayed-answer conditions separately using a paired $t$-test. A significant GKT effect—namely, the relevant questions elicited larger SCRs than the neutral questions—was obtained in both the immediate- and delayed-answer conditions, $t(23) = 6.92, p < .0001$; $t(23) = 3.99, p <
However, inspection of Figure 2 suggests that the GKT effect itself is larger in the immediate-answer condition, and this difference might contribute to the significant interaction. Thus the GKT effect was compared between the immediate- and delayed-answer conditions with a 2 sample t test. The result showed that the GKT effect in the immediate-answer condition was significantly larger than that of the delayed-answer condition, t(46) = 2.55, p < .02.

At tone stimulus onset, only the main effect of Interval was significant, F(1, 46) = 12.25, p < .001, indicating that the SCRs were larger in the delayed-answer condition, Ms = .07 and .46 for the immediate- and delayed-answer condition. Since neither the main effect of Question Type nor interaction of Question Type x Interval was significant, no GKT effect was obtained at the point of delayed answers.

**Classification Index**

The scoring method for field practice proposed by Lykken (1959), described in the Method section, yielded classification of each participant into either an "innocent" or a "guilty" person, although, clearly, the participants in this experiment were all belonged to the "guilty" category.

Classification based on the immediate-answer response in the two immediate-answer conditions was 66.7% (N = 16) correct; classification based on the response to the question onset in the two delayed-answer conditions was 41.7% (N = 10) correct; and classification based on the answer response in
the two delayed-answer conditions was only 29.2% (N = 7) correct. There was a significant difference between the accuracy rates of the immediate- and delayed-answer conditions when the latter was based on the answer response, \( X^2 (1, N = 48) = 7.35, p < .01 \). However, when the accuracy rate of the delayed-answer condition was based on the response at question onset, there was no significant difference between the immediate- and delayed-answer condition, \( X^2 (1, N = 48) = 3.02, p < .09 \).

**Reaction Time to the Relevant and Neutral Question**

The participants in the IAC and the DAC groups were asked to hit a computer key at the same time they gave a denial to each question to get the next one on the computer screen. Thus, for these participants, it was possible to measure response latency to the response cues, namely, question onset for the IAC group and tone onset for the DAC group. However, this measurement was not essential for the experiment and a part of computer program to measure the reaction time was incorporated after several participants had been already tested. Thus, only 21 participants' data were available (10 participants in the IAC group and 11 in the DAC group). For each participant, mean reaction time was obtained for relevant and neutral questions, respectively, by averaging 6 responses to the relevant and 18 responses to the neutral questions. These data were incorporated into a 2 x 2 (Group x Question Type) ANOVA. The only significant effect
Figure 4. Mean reaction time to the relevant and neutral questions in the Immediate-Answer-Contingent and Delayed-Answer-Contingent groups.

was Group, $E(1,19) = 6.66, p < .05$, indicating that reaction time was shorter when participants were tested in the immediate-answer condition, $M_s = 2.48$ and $3.12$ s for the immediate- and delayed-answer condition, respectively (see Figure 4). The Question Type main effect was nearly significant, $E (1, 19) = 4.12, p = .057$ (epsilon = 1), with shorter reaction times to the relevant questions, $M_s = 2.75$ and $2.88$ s for the relevant and neutral questions, respectively.
Discussion

The major finding of the first experiment is that, with a visual mode of presentation of the questions, a reliable GKT effect in terms of SCR amplitude was obtained in both the immediate- and delayed-answer condition in the interval immediately following question onset. Moreover, the immediate-answer condition produced a larger GKT effect than the delayed-answer condition at question onset, whereas this was not clearly so when the classification index was applied. In the latter case, the accuracy rate of classifying participants as guilty did not differ significantly between the immediate- and delayed-answer conditions. These results thus, contrast with those of Nakayama et al. (1988), who reported not only that the delayed-answer condition was superior in obtaining the GKT effect, but also that the immediate-answer condition failed to yield any reliable GKT effect.

As previously noted, the GKT questions in the immediate-answer condition in the Nakayama et al. (1988), each remained on the computer screen for a considerable time after the participants had given their answer to it. It is possible that, as a result of this procedure, some of the participants in the immediate-answer condition did not treat their task as seriously as the participants in the delayed-answer condition. The fact of the questions remaining on the screen may look contrived or lack sense to the participants. To investigate this possible objection or confound in Nakayama
et al.'s study, the present experiment introduced an additional immediate-answer condition, where the questions disappeared from the screen, as soon as the participants gave their answers. In other words, the disappearance of the questions in this condition was contingent upon the participants' answers, and, hence, more similar to a normal oral presentation of questions than Nakayama et al.'s procedure. However, neither the main effect of Contingency nor its interaction with other main factors was significant. The nonsignificant Contingency main effect indicates that the original purpose of the present experiment did not work as expected, but, in turn, it also indicates that the reliable GKT effect was obtained in the unmodified version of the immediate-answer condition which was originally examined in Nakayama et al.'s study. Therefore, this discrepancy between the results of the two experiments is particularly difficult to explain.

A conspicuous difference between the two is the participants' motivational levels to avoid detection. In Nakayama et al., the participants were instructed that an electric shock would be delivered if they failed the polygraph test, and to back up this claim a real spark was fired from a stun gun as a demonstration, although the instruction served only to heighten the motivation of the participants to avoid detection. In the present experiment, however, local ethical restrictions made it impossible to duplicate this part of Nakayama et al.'s study and hence the
participants in the present study were only asked, in a crime scenario, to beat the polygraph test just in order to be released from the police investigation. Therefore, it could be assumed that the participants in the study of Nakayama et al. had higher levels of motivation than participants in the present experiment.

Although several GKT studies have manipulated levels of motivation to avoid detection, the results have been discrepant. Elaad and Ben-Shakhar (1989) and Gustafson and Orne (1963) found a better rate of detection under motivational instructions to avoid detection, while Davidson (1968), Furedy and Ben-Shakhar (1991), Horvath (1979) and Lieblich, Naftali, Shmueli, and Kugelmass (1974) did not. Moreover, when the effect of heightened motivation was obtained (e.g., Elaad & Ben-Shakhar, 1989; Gustafson & Orne, 1963), it enhanced the detection rate, not the reverse. The participants in Nakayama et al. may have used some kind of countermeasures to avoid detection. In fact, Nakayama et al. (1988) argued that the SCRs obtained under the immediate-answer condition might have been contaminated by participants' motor responses (e.g., oral answers), and hence this produced inferior detection in the immediate-answer condition. However, as neither Nakayama et al. nor the present experimenter performed any check on the use of countermeasures by the participants, this remains a matter of pure speculation.
Several studies have compared detection efficiency between the immediate- and delayed-answer conditions. However, the detection did not differ between these two conditions when responses were compared immediately following question onset in the CQT (Dawson, 1980), in the GKT (Furedy & Ben-Shakhar, 1991), and in their DDP (Furedy, Davis, & Gurevich, 1988). Only Furedy, Posner, and Vincent (1991) found relative superiority of the immediate-answer condition to differentiate deception in the DDP experiment. The delayed-answer condition in the present study also yielded a significant discrimination between the relevant and neutral questions when SCRs were compared following question onset. Moreover, the correct classification rates in the two conditions were not significantly different when the SCRs were compared following question onset. Accordingly, at least some of the discrepancy concerning the superiority of the immediate-answer condition over the delayed-answer condition may reflect nothing more than a sampling error. In other words, guilty knowledge can be detected equally successfully in the immediate- and delayed-answer conditions when the responses are compared immediately following question onset. Since no discrimination effect was obtained for the participants' answers in the delayed-answer condition, the crucial factor for the GKT effect may be the intention to deceive rather than the act of deception.

Although it was not an essential objective of the study, sex differences in the psychophysiological detection were
also examined in the present experiment. In the initial matching trial, a marginally significant sex difference was obtained with female participants showing smaller SCR-ORs to the repetitive tone stimuli. In the detection period, however, neither the main effect of Sex nor its interaction with other factors were significant. Several studies of ORs have indicated that females are less reactive than males when phasic electrodermal responses to simple stimuli are measured (e.g., Kimmel & Hill, 1961; Kimmel & Kimmel, 1965; Maltzman, Gould, Barnett, Raskin, & Wolff, 1979), which is consistent with the present result. Only a few studies have examined a possible sex differences in psychophysiological detection (e.g., Cutrow et al., 1972; Gudjonsson, 1982; Timm, 1982) but they all agree that sex is unrelated to measures of differential responsivity in the GKT. Sex differences in SCRs, if any, may be overridden by a task like the GKT.

An interesting feature emerged in the results of the reaction time measure although this was not planned as a measure in the present experiment. First, the Delayed Answer-Contingent (DAC) group showed a significantly larger reaction time than the Immediate Answer-Contingent (IAC) group to both the relevant and neutral questions. As the DAC group had 8 s to process a question and prepare for the key response, the reaction time in the DAC group ought to be shorter compared to the IAC group. Second, reaction time to the relevant questions was shorter than to the neutral
questions, although this effect was only marginally significant.

Davis (1961) suggested several hypotheses to account for augmented physiological responses in the psychophysiological detection of deception. One of his hypotheses was that the relevant question raises an emotional conflict in an examinee because he or she should answer it deceptively, and such conflict produces a larger psychophysiological responses in the examinee. If Davis's conflict hypothesis is true, reaction time to the relevant question should be slower. The present results, which showed acceleration of reaction time to the relevant question, apparently contradict the conflict hypothesis and hence need some considerations. However, the result would have to be replicated in another experiment for further discussion, as the difference in reaction time to the relevant and neutral question was only marginal.

In summary, the present experiment showed that guilty knowledge can be detected in the immediate- and delayed-answer conditions when the responses were compared immediately following question onset. Since the immediate-answer condition is easier to deploy compared with the delayed-answer condition, the former procedure seems to be better in field practice of the GKT. There were no sex differences in the GKT effect. Finally, a reaction time measure may provide additional information about the mechanism of differential responsivity to relevant and neutral questions in the GKT.
EXPERIMENT II

There are several theories to account for the GKT effect.

Lykken (1974), the originator of the GKT, conjectured that prior knowledge of the relevant items provides them with a "signal value" that constitutes the basis for differential responsivity to these items. This assumption rests on the theory of orienting responses (ORs). The OR describes a complex of physiological reactions evoked by a novel stimulus, change in stimulus modality, or change in stimulus significance (e.g., Berlyne, 1960; Lynn, 1966; Sokolov, 1963). Indeed, Lykken (1974) suggested, "...for the guilty only, the 'correct' alternative will have a special significance, and added 'signal value' which will tend to produce stronger orienting reflexes than that subject will show to other alternatives" (p. 728).

Ben-Shakhar (1977) elaborated Lykken's simple OR-based explanation by including habituation processes of ORs as an important factor of the GKT effect and proposed dichotomization theory. According to this theory, questions (or items) presented during the GKT are differentiated cognitively into two distinct categories: relevant and irrelevant stimuli. Furthermore, it is assumed that habituation takes place only within each category with little or no carryover across categories. Since there are fewer relevant questions compared to irrelevant or neutral questions in the GKT, less habituation occurs for relevant
questions. In other words, the GKT effect reflects differences in the amount of habituation occurring in the relevant and irrelevant question categories.

In contrast to the OR-based approaches, earlier accounts of psychophysiological detection focused mainly on emotional and motivational factors, such as conflict accompanying lying (Jung, 1906/1973; Luria, 1932), fearful consequences of detection (Davis, 1961), and/or motivation/intention to deceive (Gustafson & Orne, 1963, 1965a). Although it is difficult to create situations in laboratory experiments as intense as real-life polygraphic interrogations, there is experimental evidence that emotional and motivational factors affect detection. For example, several laboratory studies have demonstrated that participants who are motivated to beat the polygraph test are detected at a higher rate than nonmotivated participants (e.g., Gustafson & Orne, 1963, 1965a). As to the conflict hypothesis, several studies have reported findings that oral responses of "no" to the questions, which can be operationally defined as lying and expected to arise conflict in a participant, produce a larger GKT effect than "yes" or a silent response (e.g., Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Gustafson & Orne, 1965; Horneman & O'Gorman, 1985). Accordingly, it seems difficult to relate detection of information to a single underlying mechanism, such as an emotional-motivational mechanism or a cognitive mechanism (Ben-Shakhar & Furedy, 1990).
In line with this perspective, Elaad and Ben-Shakhar (1989) have proposed a unified conceptual framework to explain the GKT effect. They suggested that the response evoked by the relevant item or question reflects the degree to which this stimulus was attended to. Detection can be avoided if the participant manages to ignore the relevant items. However, doing this is not an easy task because various factors, such as the nature of the relevant stimulus and the motivational state of the participant, may operate to keep the focus of attention upon the relevant items. Elaad and Ben-Shakhar (1989) used the term "noteworthiness" of the stimulus to denote the degree of attention elicited by that stimulus.

According to the perspective, factors such as motivation and type of verbal response affect detection through their influence on the noteworthiness of the relevant items. For example, when the consequence of the detection test is important for the participant, it becomes more difficult to ignore the relevant item. Likewise, a deceptive answer to a relevant item may increase its noteworthiness because of the participant's conflict in that situation. Elaad and Ben-Shakhar (1989) varied the motivational factor (high and low motivation to beat the polygraph test) and the verbal response type ("no," "yes," silent, and repeating item) in their experiment. They predicted that a higher detection rate would be obtained in the heightened motivation group than in the low motivation group because motivation might
increase the noteworthiness of the relevant stimulus, and thus would make it more difficult to ignore. As to the verbal response type, the demand for a verbal response might be associated with more attention directed to the relevant stimulus than remaining silent, which allows a strategy of ignoring that stimulus; thus, a higher detection rate would be obtained in the verbal response groups than in the silent group. Furthermore, a variable verbal response (repeating items) to the different stimuli might require more attention than a standard response to all stimuli. Thus, the highest detection rate would be obtained in the heightened motivation with variable verbal response group.

The results indicated that highly motivated participants were detected at a higher rate than less motivated participants. The act of lying was associated with enhanced differential responsivity, but no differences were obtained between verbal response versus no response or between variable versus standard verbal response. Elaad and Ben-Shakhar (1989) concluded that the results only partially supported the noteworthiness hypothesis, because they failed to demonstrate predicted differences between verbal responding versus remaining silent and between variable versus standard verbal response. However, it is possible that the noteworthiness of relevant items in their study was so strong it inhibited the effect of a demanding verbal response, that is to say, a "ceiling effect."
On the basis of the noteworthiness explanation, it is predicted that if an experimental manipulation can produce clearly different signal values for each relevant item, the magnitude of psychophysiological response to each relevant item will vary with signal value. The present experiment was designed to evaluate this hypothesis.

First, two groups of participants were asked to choose two items relating to an imagined-crime situation (burglary) in a crime scenario, each from a different category (spot for invasion and item to steal). These two items comprised the two relevant items of the GKT which were given later. One group of participants were presented an alarm sound from the computer when they chose the item they imagined stealing. This procedure was designed to add some reality to a crime scenario and arouse participants' emotions. Therefore, the procedure enhanced the signal value of items accompanied by an alarm. It was expected that both of the relevant items would produce enhanced SCRs in the GKT, but that the item of heightened signal value would evoke a larger electrodermal response than the normal relevant item.

To assess the OR-based accounts of the GKT effect, the participants were also to be exposed to the Preliminary Novelty Paradigm (PNP) prior to the GKT procedure. This procedure is designed to measure individual differences with respect to some key characteristics of the OR (Furedy, 1993b). Briefly, the PNP consists of 9 repeated tone (or light) trials, followed by a cross-modal changed light (or
tone), and a re-presentation of the repeated tone (or light) stimulus. With the PNP, four well-known components of OR can be assessed: namely, habituation (trial 1 minus trial 9), dishabituation (trial 11 minus trial 9), OR reinstatement or ORR (trial 10 minus trial 9), and super ORR (trial 10 minus trial 1). All these four effects were expected to emerge reliably in the sense that the first term in the four equations exceeds the second term (see, e.g., Furedy, 1968; Ginsberg & Furedy, 1974). However, the critical assessment for the present experiment consisted in examining the correlations between these four OR-related effects and the GKT effect. If an OR mechanism is significantly involved in the GKT, all four OR components should be correlated with the GKT effect.

Another distinctive aspect of the present experiment was introducing a voice-synthesized interrogation by computer and assessing its efficacy. Since most polygraphers know the feature of the crime in question, this may lead to their differentially presenting the relevant and control questions. In addition, the polygrapher must interact with a suspect before and during the detection test, and this could create a prejudice about the guilt of the suspect. The voice-synthesized presentation by computer eliminates this source of possible examiner bias and, essentially, sets up a "blind" interrogation.
Design of Experiment II

A 2 x 2 x 2 (Question Type x Category x Signal Value) mixed design with repeated measures on Question Type and Category was used. The Question Type variable refers to the relevant and neutral questions: the variable the most critical in the GKT study. The Category variable refers to the spot for invasion and item to steal. A between-subjects factor was the levels of signal value of relevant items varied by the alarm treatment. The treatment was also varied as a within-subjects factor; providing the alarm treatment only to the stolen item in the experimental group. Reaction time was included as a dependent variable besides SCR, because it provided insightful information in Experiment I which contradicted the conflict explanation of the GKT effect.
Method

Participants

Sixty-four participants (37 females and 27 males, aged 19-51) were randomly divided into two groups. A majority of participants (N = 51) were recruited from the Introductory Psychology class at the University of Toronto for course credit and the remaining minority (N = 13) was recruited through posters on campus. The latter, predominantly university students, were paid $8.00 for their participation. None of the participants had previously taken part in any type of studies of detection of deception.

Apparatus

The participants were tested in the same room with the same SCR recording equipment as in Experiment I, except a computer (Primax-320SX, IBM PC compatible) with a 15 inch monitor and two outside speakers. The computer was equipped with a Sound Blaster Pro® card (Creative Technology, Model CT-1330), which enabled delivery of the interrogative questions in auditory mode by computer-synthesized voice during the detection phase. Its text-to-speech function was controlled by a software, Monologue® (First Byte), which was activated by the main program written in Microsoft Quick Basic 4.5®. The computer synthesized a male voice delivered at a speed of around 350 ms per word. All the interrogative questions that the computer provided orally and were used in the detection phase were checked and reviewed prior to the experiment by five persons whose first and predominant
language was English. All the questions were written and spelled correctly without difficulty by the reviewers, which was considered to demonstrate satisfactory phonic quality of the questions. To measure participants' reaction time to each question in 1 ms order, TIMERSET and TMRREAD subroutines (Graves & Bradley, 1987, 1988) were incorporated into the main program that controlled the sequence of the experiment. The computer was also used to give instructions about the experiment in a visual mode.

The Preliminary Novelty Paradigm (PNP) was carried out prior to the main experimental procedure. The PNP consisted of 9 repeated tone (or light) trials, followed by a cross-modal change light (or tone) trial, and a re-presentation of the repeated stimulus tone (or light). The light was provided by a 100 w bulb which was mounted on the wall of the room, above and to the right of the participant. The tone was 1 kHz, 80 dB at the source and about 70 dB at the participants' level, coming thorough a loud speaker placed adjacent to the light source. The duration of the tone and light was 800 ms, while the inter-stimulus interval was randomly varied between 25, 30 (mean), and 35 s. The presentation of these stimuli was controlled by means of a four channel tape recorder (SONY TC-854-4) with four tapes having been pre-programmed for that purpose with an inter-stimulus-interval identical to the above. The presentation of the tone and light was counterbalanced in both groups. Continuous white noise of 70 dB at its source and about 60 dB
at the participants' level served to reduce any possible extraneous sounds during the PNP.

Procedure

When the participants arrived for the experiment, they were greeted by the experimenter and given general information concerning the nature of the experiment. They were told that the purpose of the experiment was to examine the reliability of a computer-assisted detection of deception (lie detection) and their task was, in an imagined crime situation, to play the role of a burglar who faced a polygraph test and tried to beat it in order to be released from the police department. The participants were also told that two electrodes would be attached to their left hands before starting the experiment but that they were only recording electrodes and they would not experience any discomfort from the electrodes. Finally, the participants were told that the experiment itself, that is apart from preparations and instructions, would last for approximately 40 minutes.

Following the general explanations of the experiment, each participant signed an informed consent form which stated that the participant was aware of the nature of the experiment and that he or she could terminate the experiment at any time with impunity.

The participants were then led into a room described as the "interrogation room," adjoining the one used for the physiological recording by the experimenter. The
participants were seated at a table facing the computer and then electrodes were attached for SCR measurement. Once the electrodes were in place, the participants were asked to rest the left hand on a soft pad and to keep it still and avoid unnecessary movements.

After the completion of the attachment of the electrodes, the participants were told that detailed instructions for the experiment and the description of the crime, for which the participants had to undergo a polygraph interrogation, would be given on the computer screen in written form and asked to initiate the computer program by hitting any key on the computer once they were ready. On the bottom line of the screen, a note prompted the participants to advance to the next screen by hitting any key so that they could read the instructions at their own pace (see Appendix B for the exact instructions presented on the computer screen).

On the first screen, each participant was instructed that the experiment had three segments, namely, a baseline period, reading a crime scenario, and a lie detection test. On the next screen, instruction about the baseline period was provided. The participants were told that they would receive several tones or lights for the 5 min baseline period and asked not to move their fingers with the electrodes too much. If the participants understood the instructions and agreed to start the baseline period, the experimenter turned off the computer screen and left the "interrogation room" and went into the adjoining room. Once the polygraph was calibrated,
SCR recording and the PNP were initiated. The first tone or light stimulus was presented approximately after 2 min from the onset of the recording. Half of the participants received the tone stimulus and the other half the light stimulus during this phase.

After the baseline period was completed, the experimenter returned to the "interrogation room" and turned on the computer screen. When the experimenter hit a key, the next instruction appeared. On the computer, it was explained to the participants that they were being asked to play the role of a burglar and later they would be requested in a crime scenario to choose one spot where they would have broken into a house and one item they would have stolen. On these chosen items, they would take a lie detection test because there was not enough material evidence for a conviction. They were further told that they should try to be a good liar during the detection phase and try not to reveal the chosen items in their physiological responding.

Then a demonstration of the synthesized voice was run in addition to a visual mode presentation of the question, "Did you steal the radio?" It was explained on the computer screen that the participants would go through a practice session where they deny every question orally by saying "no" and simultaneously hit the space bar on the computer keyboard. Four questions were presented during the practice session; they were then repeated if the participant failed to
respond to the last three practice questions in an average time of less than 100 ms or longer than 1000 ms.

After successful completion of the practice session, the experimenter returned to the experimenter's room and the participants were then given the crime scenario by hitting a computer key by themselves. In the scenario, the participants were asked to choose one of five spots to break in as well as one of five items that they could steal, by referring to two sheets which depicted a house layout and items (Appendix A-1, 2), respectively. The five alternatives for the spot were "1) Entry," "2) Family room," "3) Garage," "4) Kitchen," and "5) Living room," and for the items were "1) Bracelet," "2) Camera," "3) Purse," "4) Ring," and "5) Watch." The participants were requested to type the number key on the computer keyboard which corresponded to the selected alternatives. Immediately after the participants in the experimental group, but not in the control group, typed in the number of selected item, the alarm sound went off from the computer's two outside speakers. The alarm was 80 dB at the participants' level and composed of 1.2 kHz and 1 kHz sound provided alternatively at the duration of 800 ms each. The alarm could be terminated once the participant hit any key on the computer keyboard as directed on the next computer screen which appeared when the alarm went off. Thus, the instructions were different only in this part of the crime scenario between the experimental and control groups. On the final screen of the scenario, the participants were reminded
of the spot and item they had chosen. It was explained on the computer screen that the interrogation was about to begin and that the participants had to deny every question and hit the space bar on the computer keyboard at the same time. The computer also gave instructions to hit the space bar to initiate the interrogation once the participants were ready. When they hit the bar, the computer screen went blank and the first interrogative question was presented after 5 s had elapsed.

In the detection period, each question was presented orally by the computer synthesized voice as described above. The questions about the burglarized spot and stolen item were presented in the following way: "Did you break in through the entry?" with the spot being changed appropriately with the five alternative spots and "Did you steal the bracelet?" with the item being changed appropriately with the five alternative items. One series of questions was constructed either from the five spots or five items and each series was repeated four times alternatively. Half the participants received questions on the spot first and the other half received questions on the item first. The questions in the series were presented at a fixed interval of 20 s and with 40 s pause between each series. The questions on each item were presented at random in each series with the restriction that the first question in each series would not be the relevant alternative; thus the first question in each series was treated as a buffer and was excluded from the data analysis.
A follow-up telephone interview was carried out after approximately a month later without participants' prior notice. In the interview, the participants were asked which spot and item they had chosen in the experiment. The participants in the experimental group were also asked to describe their feeling when the alarm had sounded after selecting an item by choosing one of the following alternatives: a) startled, b) alarmed, c) aroused, d) interested, and e) nothing.

Finally, 6 participants chose Entry, 7 participants chose Family room, 23 chose Garage, 23 chose Kitchen and the remaining 5 participants chose Living room as an invasion spot. As a stolen item, 10 participants chose Bracelet, 16 chose Camera, 5 chose Purse, 20 chose Ring and the remaining 13 participants chose Watch.

**Dependent Measures**

The SCRs were defined and scored in the same way as in the Experiment I. For statistical purpose, however, the square root transformation for SCR amplitudes (Grings, 1974) was used to improve the SCR's distribution characteristics instead of the range correction method (cf. Lykken, 1972) used in Experiment I. This was due to the different manipulations for the experimental and control groups in the present experiment, where only the experimental participants received the alarm sound when they had chosen an item and the sound evoked the largest response in most of the experimental participants. In this case, the range correction method
would lead to smaller scores in the experimental participants because it divides every score obtained by a single participant with the maximum score obtained by that participant during the experiment. The square root transformation was chosen because it is the one that has been mostly applied to SCR amplitudes (cf. Boucsein, 1992).

Another dependent variable was reaction time to each question offset and this was measured by making participants hit a computer key simultaneously when they give their oral answers. A computer clock gave 1 ms order accuracy of elapsed time.

Participants were classified as guilty or innocent by Lykken's (1959) method as in the Experiment I, except that the guilty score was set to 9 or more because there were a total of 8 repetitions of series of questions.
Results

Skin Conductance Responses in the Preliminary Novelty Paradigm (PNP)

Figure 5 shows mean electrodermal responses during the Preliminary Novelty Paradigm (PNP) for the experimental and control groups. To assess differences in electrodermal responsivity to simple stimuli between two groups, theirSCRs during the PNP were analyzed by means of a 2 x 11 (Group x Trials) ANOVA with repeated measures on trials. Only the Trials main effect was significant, $F(1, 10) = 21.92, p < .0001$ (epsilon = .818), indicating absence of group difference in electrodermal responsivity, at least in SCRs to simple tone or light stimuli.

With the stimulus pattern used in the PNP, the four novelty effects, namely, habituation, dishabituation, OR reinstatement or ORR, and super ORR effects, should be observed on SCRs during the PNP (e.g., Furedy, 1968, 1993b; Ginsberg & Furedy, 1974). To test the significance of these four novelty effects, SCRs between Trial 1 ($M = .54$) and Trial 9 ($M = .20$) were compared for habituation, Trial 11 ($M = .31$) and Trial 9 for dishabituation, Trial 10 ($M = .66$) and Trial 9 for ORR, and Trial 9 and Trial 1 for super ORR effect by using a 2 x 2 (Group x Trial) ANOVA with repeated measures on trials. All these effects were reliably shown in the sense that the former trial in each effect described above was significantly larger than the latter trial: habituation effect, $F(1, 62) = 61.78, p < .0001$; dishabituation effect,
Figure 5. Mean change in skin conductance response (SCR) during the Preliminary Novelty Paradigm.

F (1, 62) = 7.36, p < .001; ORR, F (1, 62) = 114.42, p < .0001; and super ORR, F (1, 62) = 12.97, p < .001. However, neither Group effect nor the Group x Trial interaction was significant, indicating again that both the experimental and control groups did not differ in terms of orienting responses.

Effects of the Alarm Manipulation on the Experimental Group

Only the experimental participants received the alarm sound when they selected an item during the scenario reading. This was carried out in the hope of adding a signal value onto the item which the experimental participants had selected. To check whether the manipulation caused a reliably larger reaction, SCRs which were initiated within 1
to 5 s after the spot and item selections were compared. As is shown in Figure 6, the alarm provided to the participants of the experimental group caused the largest SCRs. A 2 x 2 (Group x Category) ANOVA with repeated measures on categories was conducted on these SCR scores. Main effects of Group, $F(1, 62) = 16.47$, $M_s = 1.23$ and .84 for the experimental and control group, respectively; Category, $F(1, 62) = 38.68$, $M_s = .87$ and 1.20 for the spot and item category, respectively; and Group x Category interaction, $F(1, 62) = 35.63$, were all significant ($p < .0001$). Subsequent comparison revealed that SCRs after item selection by the experimental participants were significantly larger than SCRs after spot selection by the same group, $t (31) = 5.91$, $p < .0001$, $M_s = 1.56$ and .90 for the item and spot selection, respectively. The SCRs were also significantly larger than the SCRs the control group exhibited after spot selection ($M = .83$), $t (62) = 6.19$, $p < .0001$, and item selection ($M = .84$), $t (62) = 5.28$, $p < .0001$. The other combinations of comparison did not reveal any significant difference.

Because of the confound possibility that the alarm sound, per se could have produced an emotional effect, a follow-up telephone interview was conducted approximately a month later without the participants' prior notice. In the first phase of the interview, the participants were requested to identify which spot and item they had selected in the
Figure 6. Mean amplitude in skin conductance response observed in the experimental and control groups when the participants selected a spot and an item during the scenario reading. The alarm was provided only for the experimental participants when they selected an item.

experiment. A total of 54 participants could be reached. The accuracy of the memory of their own selection is summarized in Table 1. There was no significant difference between the groups in terms of the proportion of those who gave correct and incorrect answers for the spot selection, $X^2 (1, N = 54) = .43, p > .1$, and for the item selection, $X^2 (1, N = 54) = 1.03, p > .1$. This indicates that the alarm sound, which was presented only to the experimental group, did not affect memorial retention of the selected items.
Table 1.
Number of participants who correctly or incorrectly identified their own critical items during the telephone interview

<table>
<thead>
<tr>
<th></th>
<th>Spot</th>
<th></th>
<th></th>
<th>Item</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exp.</td>
<td>Cont.</td>
<td>Total</td>
<td>Exp.</td>
<td>Cont.</td>
<td>Total</td>
</tr>
<tr>
<td>Correct</td>
<td>23</td>
<td>23</td>
<td>46</td>
<td>21</td>
<td>22</td>
<td>43</td>
</tr>
<tr>
<td>Incorrect</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>

The participants in the experimental group were further asked to describe their feelings when they had heard the alarm sound immediately after their item selection by choosing one of the following alternatives: a) startled, b) alarmed, c) aroused, d) interested, and e) nothing. Twenty six out of 32 participants could be reached. Fourteen participants chose "startled," 4 participants chose "alarmed," 1 participant chose "aroused," 3 participants chose "interested," and remaining 4 participants chose "nothing." The result indicates that the alarm sound caused some emotional reaction in the experimental participants since only 4 participants chose "nothing" (p < .0001, binomial test).

Skin Conductance Responses during the Detection Period

During the detection period, each participant accumulated 4 SCRs to relevant questions and 12 SCRs to neutral questions for each spot and item category. These responses were averaged to provide one mean SCR score for relevant questions and one mean SCR score for neutral questions for each spot and item category. These data were
subjected to a $2 \times 2 \times 2$ (Group x Question Type x Category) mixed-design ANOVA with repeated measures on two item categories (spot and item) with a nested factor of question type, that is, relevant versus neutral questions.

Only Question Type produced a significant effect, $F (1, 62) = 144.20, p < .0001$, indicating that the relevant questions elicited reliably larger SCRs than the neutral questions, Ms = .85 and .66 for the relevant and neutral questions, respectively (see Figure 7). However, the alarm
treatment for the item category of the experimental group did not affect the GKT effect since neither Category effect nor Group x Category interaction was observed.

A relationship between the GKT effect and the participants' memory correctness on their selected item was examined for each spot and item category separately. Since neither group difference nor group x category interaction was observed on the SCRs during the detection phase, both the experimental and control participants were combined and classified into two groups depending on their memory being correct or incorrect for each spot and item category (see Table 2 in previous section). A 2 x 2 (Question Type x Memory Correctness) mixed-design ANOVA with repeated measures on question type was conducted for each spot and item category. For the spot category, the only significant effect was the Question Type, $F(1, 52) = 38.94, p < .0001$, indicating that the relevant questions elicited larger electrodermal responses than the neutral questions, $M_s = .89$ and $-.68$ for the relevant and neutral questions, respectively. For the item category, the effect of the Question Type was significant, $F(1, 52) = 39.42, p < .0001$, again indicating that the relevant questions elicited larger electrodermal responses than the neutral questions, $M_s = .85$ and $.69$ for the relevant and neutral questions, respectively. Moreover, the effect of Memory Correctness was also marginally significant for the item category, $F(1, 52) = 4.00, p = .051$, indicating that participants who retained correct
Figure 8. Mean amplitude of skin conductance response (SCR) to relevant and neutral questions as a function of participants' memory correctness for their own critical items. The upper panel shows the mean SCR in the spot interrogation and the lower panel shows the mean SCR in the item interrogation.
memory of their stealing item had produced larger SCRs both to the relevant and neutral questions during the experiment, $M_s = .81$ and $.61$ for the correct and incorrect memory holders, respectively. However, two-way Memory Correctness x Question Type interaction was not significant. These data are displayed in Figure 8, with the upper panel showing the mean SCRs to the relevant and the neutral questions in the spot interrogation, and the lower panel showing the mean SCRs in the item interrogation.

Finally, a relationship between the GKT effect on SCRs and types of feeling that the experimental group participants experienced when the alarm went off during the pre-experimental scenario reading was examined. A 2 x 2 x 4 (Question Type x Category x Feeling) mixed-design ANOVA with repeated measures on two item categories that nested a factor of question type was conducted on SCR data. One participant who chose "aroused" as his feeling was excluded from the analysis. Figure 9 shows mean SCR amplitude to the relevant and neutral questions for each feeling category.

A main effect of Question Type was significant, $F (1, 21) = 74.91, p < .0001$, indicating that the relevant questions elicited reliably larger SCRs than the neutral questions, $M_s = .84$ and $.67$ for the relevant and neutral questions, respectively. Moreover, a two-way Feeling x Question Type interaction was also significant, $F (3, 21) = 4.31, p < .02$. Thus, a series of paired $t$-tests to compare the mean SCRs to the relevant and the neutral questions was
Figure 9. Mean amplitude of skin conductance response (SCR) to relevant and neutral questions as a function of feeling that participants in the experimental group described how they felt when the alarm sound went off during the scenario reading.

conducted in each Feeling group separately. Results showed that the relevant questions produced significantly larger responses than the neutral questions in the "startled" group, \( t(13) = 7.27, p < .0001, \) Ms = .96 (relevant) and .77 (neutral); the "interested" group, \( t(2) = 11.08, p < .01, \) Ms = .81 (relevant) and .53 (neutral); and the "nothing" group, \( t(3) = 3.70, p < .04, \) Ms = .69 (relevant) and .54 (neutral). However, the "alarmed" group showed only marginally significant difference, \( t(3) = 2.72, p < .08, \) Ms
= .74 and .69 for the relevant and neutral questions, respectively.

**Reaction Time to the Relevant and the Neutral Questions**

During the detection period, each participant accumulated 4 reaction times to relevant questions and 12 reaction times to neutral questions for each spot and item category. These responses were averaged to provide one mean reaction time score for relevant questions and one mean reaction time score for neutral questions for each spot and item category. These data were subjected to a 2 x 2 x 2 (Group x Question Type x Category) mixed-design ANOVA with repeated measures on two categories (spot and item) with a nested factor of question type, that is, relevant versus neutral questions.

The analysis revealed a significant effect of Question Type, \( F(1, 62) = 8.37, p < .01 \), indicating that the reaction time to the relevant question was shorter than that to the neutral question, \( M_s = 502.74 \) and 539.64 s for the relevant and neutral questions, respectively (see Figure 10).

Although an inspection of Figure 10 suggests that the reaction time measures of the participants in the control group were consistently shorter than that of the experimental group, neither Group effect nor any other interactions between factors were significant.
Figure 10. Mean reaction time to the relevant and neutral questions in the experimental and control groups for the spot and the item interrogation categories.

The relationship between the GKT effect on reaction time measures and the participants' memory correctness of their selected item was examined in the same way on SCR data with a 2 x 2 (Question Type x Memory Correctness) mixed-design ANOVA with repeated measures on question type. Neither main effect nor two-way interaction was significant for the spot category. For the item category, the only significant effect was a main effect of Question Type, $F(1, 52) = 11.29$, $p < .01$, indicating that the reaction time was shorter to the relevant questions than to the neutral questions, $Ms = 493.04$ s and 541.63 s for the relevant and neutral questions, respectively. No other significant effect was found.
Figure 11. Mean reaction time to the relevant and the neutral questions as a function of participants' memory correctness for their own critical items. The upper panel shows the mean reaction time in the spot interrogation and the lower panel shows the mean reaction time in the item interrogation.
These data are displayed in Figure 11, with the upper panel showing the mean reaction time to relevant and the neutral questions in the spot interrogation, and the lower panel showing the mean reaction time in the item interrogation.

Finally, a relationship between the GKT effect on reaction time measures and types of feeling that the experimental group participants experienced when the alarm went off during the scenario reading was examined. A 2 x 2 x 4 (Question Type x Category x Feeling) mixed-design ANOVA with repeated measures on two item categories with a nested factor of Question Type was conducted on the reaction time data. The participant who chose "aroused" as his feeling was again excluded from the data analysis. Figure 12 shows mean reaction time to the relevant and the neutral questions for each feeling category.

A main effect of Question Type was significant, $E(1, 21) = 13.62, p < .002$, indicating that the relevant questions elicited reliably shorter reaction time than the neutral questions, $M_s = 539.77$ and $592.07$ s for the relevant and neutral questions, respectively. Moreover, a two-way Feeling x Question Type interaction was also significant, $E(3, 21) = 4.71, p < .02$. Thus, a series of paired $t$-tests were conducted to compare the mean reaction time to the relevant and the neutral questions in each Feeling group separately. Only the "nothing" group showed significant difference in reaction time between the relevant and the neutral questions,
Figure 12. Mean reaction time to the relevant and the neutral questions as a function of feeling that participants in the experimental group described as they felt when the alarm sound had gone off during the scenario reading.

\[ t(3) = 4.33, \ p < .03, \ MS = 384.90 \text{ and } 492.72 \text{ s for the relevant and neutral questions, respectively.} \]

**Relationship between the Novelty Effects and the GKT Effect**

The four novelty effects during the PNP were defined as follows and calculated for each participant: Trial 1 minus Trial 9 for habituation, Trial 11 minus Trial 9 for dishabituation, Trial 10 minus Trial 9 for OR reinstatement or ORR, and Trial 10 minus Trial 1 for super ORR. As stated above, these four novelty effects emerged reliably in the
Table 2

Intercorrelations between novelty and the GKT effects (n = 64)

<table>
<thead>
<tr>
<th>Effects</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. GKT (SCR)</td>
<td>--</td>
<td>-.28*</td>
<td>.11</td>
<td>.04</td>
<td>.29*</td>
<td>.22</td>
</tr>
<tr>
<td>2. GKT (RT)</td>
<td>--</td>
<td>-.11</td>
<td>.04</td>
<td>.02</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>3. Habituation</td>
<td>--</td>
<td>.45**</td>
<td>.67**</td>
<td>-.40**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Dishabituation</td>
<td>--</td>
<td></td>
<td>.52**</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. ORR</td>
<td></td>
<td></td>
<td></td>
<td>.52**</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>6. Super ORR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.42**</td>
<td></td>
</tr>
</tbody>
</table>

** p < .01, * p < .05 by Fisher's r to z sense that the first term in the four equations significantly exceeded the second term. The GKT effect for each participant was obtained by subtracting mean responding to neutral questions from mean responding to relevant questions on SCR scores and reaction time measure. The Pearson's product-moment correlation coefficients (r) were calculated among these effects without splitting the experimental and control groups since there were no group differences either in the PNP or the GKT effects. The results are summarized in Table 2.

All novelty effects were significantly correlated to one another except the relationship between dishabituation and super ORR. Moreover, one of the novelty effects, OR reinstatement, showed a small but significant positive correlation with the GKT effect in SCR measure (r = -.29, p < .05). The GKT effect in SCR also correlated significantly with the GKT effect in the reaction time measure (r = .28, p < .05).
Classification Index

The field scoring method of Lykken (1959) was applied to the data to classify the participants as either guilty or innocent. In the experimental group, 25 out of 32 (78.1%) participants were classified as guilty. In the control group, 22 out of 32 (68.8%) were classified as guilty. A $X^2$ test was performed to test for significance of the different proportion of participants classified as guilty in two groups but no significant difference was revealed, $X^2 (1, N = 64) = .72, p > .1$.

Although the present experiment was totally different from Experiment I with regard to the mode of question presentation as well as other features, a comparison was made between two experiments on the number of participants who were classified as guilty. To do this, the numbers of participants in the two immediate-answer groups in the Experiment I were combined and this resulted in 16 out of 24 participants (66.7%) being classified as guilty. Similarly, the numbers of participants in both experimental and control groups in the present experiment were combined and this resulted in 47 out of 64 participants (73.4%) being classified as guilty. A $X^2$ test revealed that there was no significant difference on proportion of guilty-classified participants between the two experiments, $X^2 (1, N = 88) = .39, p > .1$. 
Discussion

Three major conclusions can be drawn from the present experiment. The first is that a significant GKT effect on SCRs, that is, SCRs to the relevant questions being larger than that to the neutral questions, was again obtained as in the Experiment I and as in the most of the GKT literature. However, the manipulation of signal value of the stimulus had no main or interactive effects on differential responsivity to the relevant questions. Second, a weak but significant relationship was observed between one of the four orienting response indices and the GKT effects in SCRs. Third, reaction time was significantly accelerated only to the relevant questions. Additionally, the computer-based oral presentation of questions seems to have functioned adequately in a sense that it produced the significant GKT effects.

In relation to the first finding about the signal value manipulation, one common account of the GKT effect is that signal value of the relevant item exceeds that of the neutral items (see, e.g., Lykken, 1974), thus bringing larger responsivity to the relevant items. A more recent formulation of this account is the noteworthiness hypothesis (Elaad & Ben-Shakhar, 1989), according to which the GKT effect reflects the degree to which the relevant items are more attended to than the neutral items--thus, the former is more "noteworthy." The aim of the alarm manipulation in the present experiment was to increase the degree to which the experimental condition, which consisted of role-playing a
crime scenario, was more realistic. It was hoped that this manipulation would add some signal value to the relevant questions and hence would produce a better GKT effect. It should be noted that this direct manipulation of the signal value itself of the relevant item differed from the more customary motivational manipulation by monetary incentives (e.g., Elaad & Ben-Shakhar, 1989; Davidson, 1968; Furedy & Ben-Shakhar, 1991) or by different instructions (e.g., Gustafson & Orne, 1963, 1965a; Horvath, 1979; Lieblich, Naftali, Shmueli, & Kugelmass, 1974).

The alarm was an effective stimulus, as it produced larger SCRs in the participants of the experimental group when they selected their own to-be-stolen items (see Figure 6). Such enhanced SCRs might be due to the participants' emotional reactions to the alarm, an idea which was supported by the post-experimental interviews. To this extent, therefore, it seems that the signal value of the scenario as a whole was enhanced. However, the alarm-associated relevant questions did not produce greater SCRs than the other questions, hence no signal value was added to the questions themselves. This means that the experimental manipulation was not successful to determining whether adding signal value would affect the detection of information, because the manipulation check by recalling stolen items indicated that no signal value was added to the questions themselves. It does appear, therefore, that increasing signal value of the relevant item is a difficult task in the laboratory context.
Accordingly, even though the noteworthiness hypothesis (Elaad & Ben-Shakhar, 1989) is consistent with the GKT outcomes, this theoretical consideration provides little help in developing procedures that maximize the accuracy of GKT detection, at least in the laboratory context.

An OR-reinstatement (ORR) obtained in the PNP indicated a small but significant positive correlation with the GKT effect ($r = .29, p < .05$). The ORR is a component of orientation which, after habituation, follows changes in stimulus intensity, modality, duration, frequency, sequence (i.e., duration and variability of the interstimulus intervals), complexity, information content, or stimulus significance (Boucsein, 1992). Moreover, Furedy and Ginsberg (1975) reported that electrodermal responses to a modally-changed stimulus tended to be larger than those elicited at the beginning of the habituation series. It should be noted that the contextual condition in which an ORR appears is similar to the condition in which an enhanced responsivity to the relevant item emerges in the GKT, at least in terms of stimulus significance. In other words, the repetitive stimulus and the modally-changed stimulus in the PNP correspond to the neutral item and the relevant item in the GKT respectively. Thus, an enhanced responsivity to relevant items obtained in the GKT paradigm might be due to an ORR component of orientation, and this is what OR-based account of the GKT such as Lykken's (1974) signal value hypothesis and Ben-Shakhar's (1977) dichotomization hypothesis predicts.
The significant correlation between the ORR and the GKT effect in the present study is the first correlative evidence for their predictions.

However, the $r$ value, although statistically significant, seems to be too small in a usual sense for making a confident causal inference, as it accounts for less than 10% of the variance. Because psychophysiological measures such as SCRs are so reactive to all stimuli, variability of the response is large even in the same individual but at different points of time. Thus, it is usually difficult to obtain a high correlation between measures in psychophysiological studies. Another and more important reason for the low $r$ between the ORR and GKT effect may be due to a difference in number of stimuli presented prior to the stimulus change. Although it is difficult to specify the number of trials required to obtain a complete habituation of the OR because habituation depends on the nature of the stimulus, it is assumed that such habituation never occurred in a sequence of GKT questions in the present study, where only five items were presented in a random order. To obtain the GKT effect in a context of the ORR, one should provide the same number of neutral items prior to a relevant item as the number of repeated stimuli preceded a stimulus change in the PNP. If the GKT effect truly represents the ORR, a higher correlation may be expected between the OR pattern and the GKT effect.
The reaction time results suggest that the participant's attentional level was increased specifically to the relevant questions and that such attentional mechanism did play, at least, a partial role in the GKT's efficacy to detect guilt in the present experiment. Many psychological studies have been carried out with reaction time measure (see Welford, 1980, for an extensive review of reaction time studies), but it is difficult to specify the exact psychological construct that the measure provides. Nettelbeck (1980), after reviewing psychopathological studies which used reaction time as a measure, claimed that all the constructs provided to account for the results of the studies, such as orientation and selection, concentration, search, activation, and preparedness, were relevant to the concept of attention. In general, the more the stimulus is attended to, the more the reaction time to the stimulus is accelerated. In line with this, the acceleration of the reaction time which occurred only to the relevant questions in the present experiment can be interpreted as indicating that the participants attended more to the relevant than the neutral questions. This interpretation conforms to the OR-based hypothesis of the GKT effect, because orienting is considered by many researchers to reflect attentional process in organisms, and, in particular those processes that underlie passive attention to input (Siddle, 1991). In other words, both the acceleration of reaction time to the relevant items and the significant
correlation between the GKT effect and the present ORR mentioned can be explained in terms of focused attention.

Several students of the GKT have directly referred to attention as a possible mechanism which contributes the GKT effect (e.g., Day & Rourke, 1974; Elaad & Ben-Shakhar, 1989, 1990; Waid, Orne, Cook, & Orne, 1978; Waid, Orne, & Orne, 1981), but none have directly examined measures which might reflect the participants' attentional level. It should be pointed out, however, that the reaction time measure in the present study was rather indirect, as it consisted of asking the participants to hit a computer key simultaneously with giving their oral answers to the questions. A more direct measure of reaction time in the GKT context is voice latency. So far, only Cutrow et al. (1972) have used this measure in a GKT-like study in which they reported significantly shorter voice latency to relevant than neutral questions; however they did not discuss the implication of this finding, as their primary purpose was to compare usefulness of several psychophysiological indices. Both reaction time and voice latency have drawbacks as indices of detection of deception, because they are under voluntary control of the human subjects. However, they could certainly provide additional information in exploring the mechanism of detection of deception, especially in a laboratory setting.

In the present experiment, the questions were presented in auditory mode by a computer synthesized voice, in contrast to the visual mode in Experiment I. Visual presentation is
more convenient in the sense that it enables picture presentation and easy arrangement, but it may produce extra variability and even artifacts as a function of participants not attending maximally throughout the period of presentation. Auditory presentation is technically more complicated to arrange, but may be experienced as more similar than the visual presentation to a "natural" GKT procedure. To explore this speculation, correct detection rates were compared between visual (66.7%, when the two immediate-answer groups in Experiment I were combined) and auditory (73.4%, when both experimental and control groups in Experiment II were combined) presentation, but no significant difference emerged. However, Experiment I and II differed not only in the mode of presentation of the questions, but also in several other attributes, hence no definitive conclusion can be drawn from the above findings. Another issue concerning the accuracy of detection and the computer presentation of questions is the fact that the detection rates were lower in comparison with the mean accuracy rates of 84%, which Ben-Shakhar and Furedy (1990) provided in their review of 10 GKT studies conducted by human examiners. Because of procedural differences among these studies, it is also somewhat inappropriate to directly compare these data. Yet there was a significant difference when a comparison was made between the accuracy rates obtained in the present two experiments and the 10 GKT studies in the Ben-Shakhar and Furedy's review, \( \chi^2 (1, N = 312) = 6.28, p < .02 \), which
indicates that the accuracy rate in the present experiments is lower than in the previous ten studies. However, all this is mere speculation and only more systematic comparisons can indicate the effect of human versus computer presentation of questions. Such an attempt might provide implications for understanding the underlying mechanism of the GKT effect. Once a factor or factors which produce different accuracy rates are identified, the computer-based oral interrogation, which does away with the need for the examiner-examinee interaction, might be feasible to enhance the objectivity to the GKT procedure.
GENERAL SUMMARY AND DISCUSSION

In this study, two experiments were carried out to examine the accuracy of the GKT with SCR as a dependent variable in a laboratory situation where participants role-played a crime scenario. In both experiments, a computer was used to deliver the GKT questions rather than presenting them orally by a human examiner. The questions were delivered in a visual mode in Experiment I and in an auditory mode in Experiment II. Experiment I primarily focused on a comparison of detection efficiency between immediate- and delayed-answer conditions. In Experiment II, an underlying mechanism of the GKT effect was explored in terms of the signal value hypothesis (Lykken, 1974) by trying to manipulate the emotional impact of the questions.

Some of the major findings and conclusions of the study are as follows. First, a reliable GKT effect was obtained in both experiments with "relevant" questions evoking significantly larger SCRs in the participants than neutral ones. The effect was significant when the SCRs were examined in the interval immediately following question onset, irrespective of whether the question had been presented in the immediate- or the delayed-answer conditions in Experiment I. This indicates that, contrary to Nakayama et al. (1988), the immediate-answer procedure would fare better in field practice of the GKT than the delayed-answer procedure; furthermore the former is easier to deploy than the latter. Second, the manipulation of the signal value of the questions
in Experiment II failed to produce any differential responsivity to the relevant questions. Third, a weak but significant positive correlation was observed between the GKT effect and participants' orienting response patterns with respect to OR reinstatement, which is typically aroused by a sudden change of stimulus pattern or modality. This result suggests an attentional or orienting consideration of the GKT effect. Fourth, reaction time was significantly accelerated only to the relevant questions in both experiments, which lends further support to an attentional interpretation of the GKT effect. Fifth, no sex difference was observed in the GKT effect in Experiment I. Finally, all the observations described above were obtained by the computer-assisted interrogation in which no human examiner-examinee interaction was needed—a procedure that eliminated bias due to human subjective administration.

The mechanism underlying the differential, enhanced psychophysiological responsivity to the relevant questions has not been fully understood in the GKT, although it is even more obscure in the case of the CQT. Several hypotheses have been proposed to explain the way in which autonomic responses might be responsible in the detection of guilt or lying. Davis (1961) suggested three explanations: the conditioned response hypothesis, the punishment hypothesis, and the conflict hypothesis.

The conditioned response hypothesis assumes that relevant items play the role of conditioned stimuli and evoke
enhanced responsivity to the relevant question with which they have been associated in the past (e.g., crime). However, as Davis (1961) has already noted, the simple conditioned response hypothesis could hardly be the whole explanation for the enhanced responsivity. For example, many laboratory experiments have demonstrated successful detection of a card number, picked by the participant before the interrogation procedure. In such cases, it seems unlikely that the strong association between a card number and enhanced psychophysiological response has been established. Although supplemental, some of the participants in Experiment II received an alarm sound when they picked up their own items of theft and the alarm caused the largest SCRs throughout the experiment. In a sense, the procedure resembles classical conditioning of SCR, with item as a conditioned stimulus and the alarm as an unconditioned stimulus. However, this one-shot conditioning apparently failed to produce larger SCRs since the GKT effect did not differ between items which were associated with the alarm and items which were not.

The punishment hypothesis suggests that the emotional, and thus enhanced, response to the relevant item is due to fear of the consequences of the participant's failure to deceive. However, this hypothesis does not explain why the specific larger response is produced only to the relevant questions. Moreover, in the usual laboratory experiments,
the participant is not punished in any way if he or she fails to deceive and this was also true in the present study.

The conflict hypothesis attributes the enhanced autonomic responsivity of the relevant item to conflicting tendencies between telling the truth and lying. The more a person tries to pass the detection test (intensive motivation), the more intense the conflict becomes. This hypothesis has been partially supported by studies which have shown that verbal lying enhances the rate of detection (e.g., Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Gustafson & Orne, 1965; Horneman & O'Gorman, 1985). On the other hand, these studies also demonstrated that significant detection of relevant information was possible even in a condition where the participants did not respond verbally to the question. Moreover, this hypothesis has difficulty in accounting for the results of a study by Kugelmass, Lieblich, and Bergman (1967), which compared participant's "No" response (lying) with a "Yes" response (confessing) to the relevant item. In this study, no differences were obtained between the two verbal responses--in both of which, the relevant item was detected at better than chance rates, using changes in skin resistance. Similarly, the results from the reaction time measure in the present study are at variance with the conflict hypothesis, as reaction time was accelerated when the participants gave their answers to the relevant questions.
In contrast to these early hypotheses and studies stressing emotional and motivational variables as the major factors determining differential response in the psychophysiological detection situation, later research has focused on cognitive factors. Lykken (1974) suggested that prior knowledge of the relevant items provides them with a signal value or stimulus significance that constitutes enhanced orienting responses (ORs) to the items. Electrodermal ORs have been extensively researched and, as a generally accepted result, the OR following a certain stimulus or a class of stimuli can be enlarged if those stimuli are given signal value for a cognitive or a motor reaction (Boucsein, 1992). Therefore, according to Lykken's hypothesis, neither deception (hence, conflict or fear) nor heightened motivation to avoid detection is necessary for detecting the information; the acquisition of the relevant information is sufficient to detect such information. Indeed, Kugelmass et al. (1967) demonstrated that detection was not affected by the act of lying, and several studies have demonstrated psychophysiological detection in a state of low motivation (e.g., Davidson, 1968; Kugelmass & Lieblich, 1966; Lieblich, Naftali, Shmueli, & Kugelmass, 1974).

This kind of orienting account for the GKT effect was partially supported by the result of Experiment II, that one of OR components, OR-reinstatement (ORR), indicated a small but significant correlation with the GKT effect. However, its coefficient of determination was less than 10% ($r^2 =$
hence factor or factors other than the OR ought to contribute the GKT effect. Moreover, Bradley and Warfield (1984) showed higher detection rates in a sample of guilty participants than in a sample of innocent participants who were only informed of the details of the crime, and who could, thus, recognize all the relevant items. They argued that although guilty knowledge was necessary for detection, it was not a sufficient condition for detection. These results raise difficulties for the theoretical approaches that base psychophysiological detection on guilty knowledge per se. In other words, as Ben-Shakhar and Furedy (1990) claimed, "no single theory or single theoretical approach is capable of providing a full account for the data. Rather, an integration of the different approaches is needed in order to achieve a better understanding of the mechanisms involved in psychophysiological detection" (p. 113).

The reaction time data in the present study indicate strongly that reaction time was significantly accelerated only to the relevant questions. It is noteworthy that reaction time was investigated in relation to detection of deception as early as at the turn of the twentieth century and applied in a real life criminal investigation. Among these pioneers was Jung (1906/1973, 1910), who applied his association-reaction method of studying mental processes to the detection of lying. In the association-reaction method, a series of words was presented to a person one by one, and the person was requested to respond to the word (stimulus-
word) as quickly as possible by orally answering the first word which came to mind in response to the stimulus-word. Response latency, namely, reaction time was measured for each word. On the basis of unusually long reaction times, in addition to the nature of the verbal response, emotional complexes were inferred. Jung's primary interest in the method lay in the diagnosis of the emotional complexes underlying hysteria and neurasthenia (Jung, 1906/1973). In a crime-related use of the association-reaction method, a list of stimulus-words contained several words relating to the crime. If the reaction time to these critical words was found to be long, the examinee was thought to be trying to deceive and thus being guilty. Jung (1906/1973) provided a case of real embezzlement in which he investigated a suspect which he judged as guilty, and who later confessed to the crime. Several researchers followed his method in experimental situations and found that reaction times tended to be longer whenever deception was practiced (e.g., Henke & Eddy, 1909; Leach & Washburn, 1910; Yerkes & Berry, 1909).

Marston (1920), who is best known for his psychophysiological study of deception and for introducing the notion of "specific lie response" into this area, studied reaction time under laboratory conditions, employing a task which was different from the association-reaction method. In the task, the participants were instructed to either obey (honest) or disobey (lying) requests to perform arithmetical calculation on sets of numbers. The time required to answer
each calculation was measured and it was predicted that reaction times would be longer when the participants tried to deceive the experimenter. Marston (1920) found three types of participants on the basis of their reaction times when they were lying: Those whose responses were consistently long; those whose reaction times were inconsistent from day to day; and those whose reaction times were consistently short. However, Goldstein (1923), in an attempt to replicate Marston's experiment, failed to find support for the existence of the third type of participants whose reaction time was shortened when they practiced lying, arguing that the participants with short reaction time were simply not involved in the task of deceiving the experimenter. Goldstein (1923) concluded, mostly by depending on introspective reports by the participants, that the consciousness of deception appeared as feeling of strain, self-consciousness, hesitation, conflict of impulses, and emotional disturbances and this led to longer reaction times when lying was practiced. The whole argument raised considerable controversy (e.g., English, 1926; Goldstein, 1923; Marston, 1925; Rich, 1926), but, as a general conclusion, whenever considerable affect is generated with deception, reaction times would probably increase (Orne, Thackray, & Paskewitz, 1972).

In contrast to the above, reaction time in the present study was accelerated only to the relevant question, about which, operationally, the participants were lying.
Therefore, it seems that the participants in the present study were not conscious of deceiving when they responded to the relevant questions. In other words, a considerable affect such as conflict and emotional disturbances (e.g., Davis, 1961) is not necessary to produce larger responses to relevant questions, at least in the laboratory GKT. Rather it appears that the participants' attention to the relevant questions contributes a reliable GKT effect—as witnessed, partially, by the significant negative correlation between reaction time and the SCR in the GKT effect ($r = -.28$, $p < .05$). In a real life detection of deception, reaction time to questions may not be useful because it is under voluntary control by humans. Under laboratory conditions, however, it provides additional information to understand the mechanism underlying the GKT effect, if participants are not aware that their reaction time is being covertly measured.

Recently, measures of brain electrical activity, commonly referred to as event-related brain potentials (ERPs), have been claiming the attention of researchers in the detection field (Bashore & Rapp, 1993). Several GKT laboratory studies have been carried out with a P300—one component of ERP—as a dependent measure and these studies have provided promising data on detection efficiency (e.g., Farwell & Donchin, 1991; Rosenfeld, Cantwell, Nasman, Wojdac, Ivanov, & Mazzeri, 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987). The P300 is typically generated under the "oddball paradigm," in which a participant is required to
monitor a series of stimulus events comprising two stimuli. One of the stimuli, often referred to as the "standard" stimulus, has a .8 to .9 probability of occurring during each trial, while the other, the "target" stimulus, occurs with probability between .1 and .2 and is interspersed among the standard stimuli. When the participant is simply asked to count the occurrence of the infrequent--hence "odd"--target stimulus, the P300 is observed as a large positive deflection following the presentation of the target stimulus. Therefore, the oddball paradigm in many ways resembles the GKT paradigm, in which, in terms of stimulus probability, the relevant and neutral questions serving as the target (odd) and standard stimuli, respectively.

To make the oddball paradigm more suitable for the detection of deception, Farwell and Donchin (1991) employed three categories of stimuli in their P300-based GKT study, namely, probe, irrelevant, and target stimuli. The probe stimuli were crime-relevant items in a crime scenario and had a .17 probability of occurring during the test. The irrelevant and target stimuli were crime-irrelevant but indistinguishable from the probe stimuli if the participants were innocent, and had a probability of .66 and .17, respectively of occurring during the test. The participants were requested to press one response button when the target stimuli appeared and press the other when the irrelevant and probe stimuli appeared. Note, however, that the probe stimuli were not distinguished in the instructions given to
the participants—they were only informed of the target and nontarget distinction. Therefore, if a participant elicited a larger P300 to the probe stimuli, it could be argued that the participant had discriminated between the probe and irrelevant stimuli and hence he or she possessed guilty knowledge. With this cleverly designed method, Farwell and Donchin (1991) reported correct classification rate of 87.5%.

One potential advantage of the ERP index over peripheral responses for detecting deception is its shorter onset latency. As the name implies, the P300 peaks around 300 ms after stimulus presentation, provided the participant is actively attending to that stimulus—far shorter time than the onset latency of peripheral responses such as electrodermal activity, which is usually measured in 1 to 5 s time-window after stimulus presentation. Therefore, the P300 may reflect a purely cognitive process of concealing information or deception, and, thus, the P300 could become a reliable index for exploring the mechanism involved in the GKT. However, from a practical point of view, there are enormous problems of software expertise and costs in developing an ERP-based test and equipment which would be usable in the field (Ben-Shakhar & Furedy, 1990). Moreover, the ERPs are vulnerable to artifacts caused by subtle body movements and even by eyeblinks; thus obtaining reliable ERP records seems to be difficult, if not impossible, without the examinee's cooperative attitude toward the test. In fact, after conducting GKT on real life criminal suspects by
concurrently measuring the P300 and the conventional autonomic responses, Miyake, Mizutani, and Yamamura (1993) concluded that the conventional peripheral measures were sufficient for detection because the P300s were too noisy to make correct classification in the field. In their study, 18 suspects were given the test and 12 of them were classified as guilty and 6 as innocent, on the basis of autonomic measures. The accuracy rate of the classification was 100%, as later confirmed by confessions or arrests of the real criminals. However, only 44% of the suspects were classified correctly on the basis of the P300 index. Moreover, recording ERPs had to be abandoned halfway in 5 cases because of artifacts caused by eye movements and eyeblinks. Accordingly, the ERP-based GKT at this stage seems suitable only to laboratory exploration of the underlying mechanism of the GKT.

Finally, the laboratory-to-field generalization problem is especially complex in this area (Furedy & Ben-Shakhar, 1991). For example, it is possible that the delayed-answer condition is superior to the immediate-answer condition in the field, where an examinee's motivation to deceive, if guilty, is surely higher than that of participants in laboratory experiment. With heightened motivation, the examinee may try to contaminate his or her physiological responses by physical countermeasures accompanied by oral answers— the strategy which may deteriorate the physiological responses especially in the immediate-answer condition.
Therefore, the only way to evaluate these practical efficacy issues is in the field itself. However, efforts to study the GKT under controlled laboratory circumstances are still useful for the standardization of the GKT and for furthering understanding of its underlying mechanism, because such parametric investigations can provide information which leads to formulation of the application-related hypotheses. The computer-assisted GKT seems advantageous both in the laboratory experiments and field practice, in the sense that it will provide GKTs that are more objectively administered than with a human examiner.
REFERENCES


APPENDIX B

The following instructions was presented on a computer screen in front of a participant before the experiment start. The bordered modules appeared one by one and the participant was required to hit a computer key to proceed to the next screen. Note that the exact format of the instruction was different due to the size of the computer screen.

Thanks for coming! The session will start soon. Before starting, I will give you some instructions about this experiment on behalf of the experimenter. If you are not sure about my explanation, please ask the experimenter anytime. He can give the answer for your question.

Hit any key to proceed, please!

This lie detection experiment can be divided into three parts. They are
1) a baseline period
2) reading a crime scenario, and
3) actual interrogation procedure,
and it takes about an hour to complete.

Hit any key to proceed, please!

The baseline period will begin now and will take about 5 min. to complete. This is to check the recording equipment and to record your basic skin conductance responses. Some lights and tones will be presenting during this period. All you have to do is just to sit quietly and relax. Please try not to move your fingers with the electrodes too much.

After the baseline period is over, the experimenter will tell you what to do next.

If you have any question, please ask the experimenter now! If you are ready, please say so to the experimenter!

[The experimenter turned off the computer screen and left the room to start recording.]
[After the baseline period, the experimenter entered the room and turned on the computer screen.]

As you already know, this experiment is about lie detection. The room you are now in is an interrogation room in a police station. Here, I would like you to play the role of a burglar, who actually committed a burglary and has been under arrest for that reason.

But you are a tough criminal and you have not confessed your crime during normal interrogation procedures. Although several investigators in this case are convinced that you are the burglar, it is also true that there is not enough material evidence to show that you are guilty.

So, the irritated chief investigator decided to make you take a lie detection test, and you have agreed. That's why you are here. If you can beat this lie detection test, you will be released.

Do you understand your role? If not, please ask the experimenter now. Next I'll explain the crime you committed (I know, I know, you are innocent. But just play the role!). You don't have to memorize the details. Just try to grasp the situation.

Hit any key to proceed, please!
The crime concerned here is, of course, the burglary.

The other day, at midnight, someone broke into a house in a Toronto suburb. After the police investigations, it was revealed that the burglar broke into the house and stole something. Fortunately, no one in the family was hurt because they were on a trip to South Florida (gorgeous, eh?). From the way in which the burglary had been carried out, the police identified and arrested you as the burglar. You had three records of burglary before.

But you were smart enough not to leave any material evidence this time. As I said before, several investigators are convinced that you are the burglar, but it is also true that they have not enough evidence.

As I said before, you will be released if you can beat the following lie detection test. If you fail, you will still be kept in custody. Moreover, the result of the test will be used in your trial as an evidence.

**Now, try to beat the following test!!!**

Next, I'll explain the rest of experimental procedures. If you have any question about the instructions, please feel free to ask the experimenter at any time.

Hit any key to proceed, please!

Afterward, I will present a scenario that depicted the crime you had committed. Please read the scenario carefully and follow the instruction in it. During the course of reading, I will ask you to make your own decision twice, that is to say, to choose the **spot** where you would have broken into the house and the **item** you would have stolen as a burglar. But keep your choice to yourself and don't tell the experimenter about the spot and the item you choose because the experimenter's mission is to identify them based on your physiological recording.

Hit any key to proceed, please!

Following the scenario reading, the interrogation will start. During the interrogation period, I'll present one question at a time at the center of my screen. The question looks like this (hit Y key on the keyboard!):

**Did you break in through the bedroom?**
At the same time, this question was also presented aurally by computer synthesized voice.

Surprised? I can speak English and I think my English is better than the experimenter's one. Agree?

Anyway, when each question is asked, you must always reply by saying (orally),

No!

Immediately after the end of question, and hit the space key at the same time. Your task is to deny every question in a same way.

Please hit any key to proceed!

Now, let's have some practice. Please keep in mind that you must respond as soon as possible after you hear the question. Also, don't forget to hit the space key at the same time.

If you are ready, please hit any key!

[Here, the participant was given some practice. If the participant's reaction time fell between 100 and 1000 msec, the computer said "Good!" and then prompted the participant to proceed to the next screen. If the reaction time was not proper one, or s/he did not give his or her oral answer, the computer or the experimenter instructed her/him so and gave more practice.]

Did you break in through the patio?

Did you steal the radio?

Good!! Let's proceed to the next instruction by hitting any key!

During the interrogation, I'll ask you 10 different questions and each question will be repeated 4 times. So you have to respond to 40 questions during this period.

There are several seconds between questions and you might become irritated. But please stay calm. That might be a tactic of the experimenter.

During the interrogation, please restrain from extensive body movements, especially left hand finger movement! If you don't do so, the record of your skin conductance response will be distorted and you give an impression that you use such strategy to avoid detection.
That's all the instructions I can give you now. If there's something you are not sure about, please ask the experimenter now. OK? If you're ready, let's proceed to the crime scenario!

Please hit any key to proceed, please!

**SCENARIO**

It was Saturday midnight in late winter and you were on your way to commit a burglary. The target was a house in a Toronto suburb. The house attracted your attention because it was off the main road, and you knew that there was no traffic at that time of night. Also, you knew the occupants of the house had been away on vacation. So your work that night looked easy.

You drove your car behind a stand of trees, where it was hidden from view, and turned off the engine, listening attentively for any sounds to make sure that you were alone. Nothing disturbed the silence apart from a lone squirrel that you seemed to have waken up. As it ran up the nearest tree you put your gloves on and began to approach the house.

Hit any key to proceed, please!

You walked cautiously around the house while you tried to figure out where it would be easiest to break into and soon you decided on the spot.

Was the spot
1) Entrance,
2) Family room,
3) Garage,
4) Kitchen, or
5) Living room?

Please choose just one spot by hitting a corresponding numerical key on the computer keyboard.

OK, you chose [spot], right? Let's return to the scenario now by hitting any key!

[If the participant belonged to the experimental group, then s/he received the following SCENARIO (a). If the participant belonged to the control group, then s/he received the following SCENARIO (b).]
SCENARIO (a-1)

Breaking in was even easier than you had expected but searching with your flashlight there didn't seem to be anything worth taking on the ground floor, so you proceeded to the second floor. It seemed that you were going to have bad luck that night as nothing worth stealing turned up until you, suddenly, noticed a chest of drawers in the master bedroom. You approached the chest and pulled out the top drawer, but there was nothing to be found until you inspected the last drawer. There, lying in the corner, was the worthy item. Then you reached for the item.

Was the item a
1) Camera,
2) Bracelet,
3) Purse,
4) Ring, or
5) Watch?

Please choose just one item now by hitting a corresponding numerical key on the computer keyboard.

[When the participant hit the numerical key, the computer produced a siren-like sound. The sound continued until the participant hit any key again following the instruction given in the next screen.]

SCENARIO (a-2)

Damn it!! The alarm!! I must get out of here now!!
You grabbed the item, put it in your pocket, and ran away.

(*PLEASE HIT ANY KEY TO TERMINATE THE SOUND*)

Luckily, you managed to get away before the police arrived.

Please hit any key to proceed, please!
SCENARIO (b)

Breaking in was even easier than you had expected but searching with your flashlight there didn't seem to be anything worth taking on the ground floor, so you proceeded to the second floor. It seemed that you were going to have bad luck that night as nothing worth stealing turned up until you, suddenly, noticed a chest of drawers in the master bedroom. You approached the chest and pulled out the top drawer, but there was nothing to be found until you inspected the last drawer. There, lying in the corner, was the worthy item. You grabbed it and put it in your pocket. Then you ran downstairs, and left the house.

Was the item a
1) Bracelet,
2) Camera,
3) Purse,
4) Ring, or
5) Watch?

Please choose just one item now by hitting a corresponding numerical key on the computer keyboard.

Please hit any key to proceed, please!

So, that's the end of story. During the burglary, you entered the house from [the spot the participant would choose] and stole [the item s/he would choose], right?

And now you're in an interrogation room. Soon, the interrogation will start. Your task is to beat the polygraph. Remember?

You have to deny every question and hit the space key after you hear the question.

Are you ready? When you hit the space key, the interrogation will start.

Good luck!