Journal of Applied Psychology
1999, Vol. 84, No. 5, 651–660

0021-9010/99/$3.00

Leakage of Relevant Information to Innocent Examinees in the GKT: An Attempt to Reduce False-Positive Outcomes by Introducing Target Stimuli

Gershon Ben-Shakhar and Nurit Gronau
The Hebrew University of Jerusalem

Eitan Elaad
Division of Identification and Forensic Sciences,
Israel Police Headquarters

This study focused on the Guilty Knowledge Test (GKT)—a psychophysiological detection method based on a series of multiple-choice questions, each having one relevant and several neutral (control) alternatives. The study examined a new method designed to reduce false-positive outcomes due to leakage of relevant items to innocent suspects by introducing target items (i.e., items known to all examinees but unrelated to the crime) to which participants have to respond (e.g., by pressing a key) while answering the GKT questions. Informed innocent participants showed relatively larger electrodermal responses to the critical items than uninformed participants, but not as large as the responses made by guilty participants. No differences between informed and uninformed innocent participants were obtained with a respiration measure. The use of the target items tended to reduce the differences between informed and uninformed innocent participants. The results further demonstrated that electrodermal responding to the relevant items was correlated with memory of these items.

The Guilty Knowledge Test (GKT) is a method of psychophysiological detection that has been extensively researched and discussed in the literature (e.g., Ben-Shakhar, 1991; Furedy & Heslegrave, 1991; Lykken, 1974, 1981). The test uses a series of multiple-choice questions, each having one relevant alternative (e.g., a feature of the crime under investigation) and several neutral (control) alternatives, chosen so that an innocent suspect would not be able to discriminate them from the relevant alternative (Lykken, 1981). Typically, if the participant's physiological responses to the relevant alternative are consistently larger than the responses to all other alternatives, knowledge about the event (e.g., crime) is inferred.

This method protects innocent suspects, because as long as information about the event has not been leaked out, the relevant and control questions should be equivalent from their perspective. Under these conditions, the probability that an innocent suspect would show consistently larger responsivity to the relevant alternative than to the neutral ones depends only on the number of questions and the number of alternative answers per question, and this probability can be controlled such that maximal protection for the innocent is provided. Indeed, a relatively large number of studies conducted to assess the accuracy of the GKT have revealed that the rate of false-positive errors (innocent suspects classified as guilty) is around 5% (for a review of these studies, see Ben-Shakhar & Furedy, 1990; Elaad, 1998).

However, a successful implementation of the GKT depends on the validity of the assumptions underlying it and, in particular, on the assumption that the relevant information is not known to innocent suspects. This seems crucial because it has been assumed that knowledge of the "guilty information" is sufficient for creating psychophysiological differentiation between the relevant and the neutral items (e.g., Lykken, 1974). The concern about possible leakage of relevant information to innocent suspects is one of the reasons for the infrequent usage of the GKT as an aid in criminal investigations (e.g., Honts & Perry, 1992). In some cases, leakage of relevant information may not constitute a severe problem, because innocent suspects failing a GKT could explain how they became aware of the critical items

Gershon Ben-Shakhar and Nurit Gronau, Department of Psychology, The Hebrew University of Jerusalem, Israel; Eitan Elaad, Division of Identification and Forensic Sciences, Israel Police Headquarters, Jerusalem, Israel.

This study was based on a master's thesis conducted by Nurit Gronau under the supervision of Gershon Ben-Shakhar and Eitan Elaad. We thank Danna Ballas, Limor Bar, Erga Sinai, and Rotem Shelef for their help in the data collection. The article was written while Gershon Ben-Shakhar was on a sabbatical leave at Brandeis University. We thank the Psychology Department at Brandeis University for the facilities and the help provided during this period.

Correspondence concerning this article should be addressed to Gershon Ben-Shakhar, Department of Psychology, The Hebrew University, Jerusalem 91905, Israel. Electronic mail may be sent to mskpugb@mscc.huji.ac.il.

(e.g., they could cite a newspaper that mentioned this information while describing the crime). However, suspects can be exposed to some critical items during the interrogation without being aware of the circumstances in which this occurred and without being able to prove that the information has been leaked. The research on the effects of misleading information on eyewitnesses (e.g., Zaragoza & Mitchell, 1996) demonstrates the vulnerability of human memory and indicates that people cannot always account for the sources of their knowledge. Furthermore, some knowledgeable but innocent witnesses to a crime might refuse to admit their knowledge because of fear of reprisal by the culprit.

The present study focused on three major issues. First, we attempted to assess the robustness of the GKT to a violation of the critical assumption that only the guilty suspects have knowledge of the relevant information. Second, we hoped to learn about the mechanisms that underlie the phenomenon of differential autonomic responding to the relevant items. In particular, we were concerned with the assumption made by Lykken (1974), as well as by other cognitive approaches to psychophysiological detection (e.g., the dichotomization theory proposed by Ben-Shakhar, 1977, and by Lieblich, Kugelmass, & Ben-Shakhar, 1970), that knowledge of the relevant items is sufficient to create the pattern of differential autonomic responding to these items. Third, we attempted to reduce the rate of false-positive outcomes among informed innocent examinees by introducing target stimuli (i.e., items known to all examinees, but unrelated to the crime) to which examinees had to respond (e.g., by pressing a key) while answering the standard GKT questions. Such target stimuli were used by Farwell and Donchin (1991), who adopted the oddball paradigm for detecting information with ERP (event related potentials) measures, but they have not been used in typical GKT experiments with autonomic measures. Recently, Elaad (1997) showed that the introduction of target stimuli does not affect the correct detection of guilty examinees, based on autonomic measures. We hypothesized that these targets would differentially attract the attention of the informed innocent participants (thus attenuating their responses to the relevant items) and would have only a small effect on the guilty participants.

A number of studies examined the effects of exposing relevant information to innocent suspects on the validity of the GKT. The first two studies (Giesen & Rollison, 1980; Stern, Breen, Watanabe, & Perry, 1981) examined this issue by using mock-crime experiments in which the same relevant information was presented to all participants: Some received it in the context of a crime, whereas others received it in a neutral context. Thus, the critical items used in the GKT had a special meaning to all participants. Nevertheless, the results indicated that the crime-related items produced stronger differential responsivity than the same items in an innocent context. Moreover, it was possible to discriminate between guilty and informed innocent participants on the basis of the electrodermal measure, with very few false positive outcomes (none in the former study and 11% in the latter).

Subsequent studies conducted by Bradley and his colleagues examined this issue using somewhat different procedures. In the studies reported by Bradley and Warfield (1984) and Bradley and Rettinger (1992), both guilty and informed innocent participants received the relevant information in a crime context, but only the participants simulating the guilty actually performed the mock crime, whereas the informed innocent participants were just informed about the relevant details. In addition, these authors used a different version of the GKT, called the Guilty Action Test (GAT), in which suspects are asked if they actually committed the crime (e.g., "Did you murder the victim in . . . ?"), rather than if they knew about the crime (e.g., "Did the murder take place in . . . ?"). The results of these studies revealed much larger false-positive rates among the informed innocent participants (25% in Bradley & Warfield, 1984, and 50% in Bradley & Rettinger, 1992) than did the rates reported by Giesen and Rollison (1980) and Stern et al. (1981).

In a more recent study, Bradley, MacLaren, and Carle (1996) manipulated two important factors in addition to the leakage of the relevant information. First, they compared the standard GKT with the GAT variation. This comparison is important because it differentiates between deception and mere knowledge. Thus, although both guilty and informed innocent suspects know the relevant details, only members of the former group are deceiving in their answers to the GAT questions. Second, they manipulated the verbal responses to the GKT and GAT questions. The results of this study resembled those obtained by Bradley and Rettinger (1992) in showing a 50% false-positive rate among informed innocent participants under the GAT and a "No" response. In addition, they found a dramatic difference in the outcomes of the informed innocent participants between the two questioning formats (e.g., a 90% rate of false positives was obtained with the informed innocent participants under the "No" response condition in the GKT). The response mode also made a significant difference, and informed innocent participants who answered "No" to the questions were more frequently classified as guilty than those remaining silent and not responding verbally to the questions.

The results of Bradley et al. (1996) imply that deception plays an important role in creating the pattern of differential responding to the relevant items. This issue was examined by several other researchers, who either manipulated the verbal responses to the GKT questions (e.g., Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Gudjonsson, 1982; Horneman & O'Gorman, 1985; Kugelmass, Lieblich, & Bergman, 1967) or used the differentiation-of-deception paradigm (e.g., Furedy, Davis, & Gurevich, 1988; Furedy, Gigliotti, & Ben-Shakhar, 1994; Furedy, Posner, & Vincent, 1991). The conclusion that can be drawn from these results is that although deception is not a necessary condition for differential responding to the relevant questions, such a response pattern is typically increased by deception.

The purpose of the present study was to further investigate the effect of exposing the relevant information to innocent examinees on the validity of the GKT and to examine whether this effect can be mitigated by the use of the target-items task. In addition, we examined whether the differences between guilty and informed innocent participants could be accounted for by differences in the number of relevant items recalled by these two groups. We used the GAT version of questions presentation for two reasons. First, in the standard GKT procedure, informed innocent participants are differentiated from the uninformed innocent participants by both their knowledge of the critical items and by their deceptive responses to the questions. Under the GAT format, both informed and uninformed innocent participants are telling the truth, and knowledge of the critical items is the only factor differentiating those two groups. Indeed, the results reported by Bradley et al. (1996) suggest that the use of the GAT format reduces false-positive outcomes in that group as compared with the standard GKT format. Second, we believe that this formulation is particularly important in a simulation setup, because under the neutral and impersonal formulation of the GKT questions, participants in mock-crime studies might perceive the situation as irrelevant and not involving. The more personal formulation of the GAT questions may make these participants more concerned and more interested in the test and thus create a setup that is more similar to a realistic interrogation. It could be argued that the accusatory formulation of the questions used in the GAT procedure may cause innocent suspects to react more strongly to the relevant questions. However, even if the accusatory format increases the general arousal of innocent suspects, there is no reason that they would show stronger reactions to the critical alternatives than to neutral control alternatives (which are similarly formulated), as long as they do not possess the guilty knowledge.

Method

Participants

A total of 108 undergraduate students participated in the experiment for course credit or payment.

Instruments

Skin conductance was measured by a constant voltage system (0.5V ASR Atlas Researches), and two Ag/AgCl electrodes (0.8 cm diameter). Respiration was recorded by a pneumatic tube positioned around the thoracic area. The experiment was conducted in an air-conditioned laboratory and was monitored from a control room separated from the laboratory by a one-way mirror. A Macintosh II computer was used to control the stimulus presentation and compute the physiological responses.

Design

A mock-crime procedure was used, and participants were randomly allocated into the following three conditions (with 36 in each condition): (a) guilty examinees who actually performed the mock crime; (b) informed innocent participants who did not perform the mock crime, but were exposed to the relevant items in the mock-crime context; and (c) uninformed innocent participants who did not perform the mock crime and had no knowledge of the relevant information. An additional factor (the task of responding to target items) was manipulated within participants, with the following three levels: (a) no target items were included; (b) examinees were instructed to count the target items silently;[1] and (c) participants were instructed to press a key as soon as they heard a target item. Each target-item condition was presented with a different question (e.g., the count-item condition was coupled with the question about the amount of money stolen), and one of the neutral alternative answers to this question was defined as the target item.[2] The order of the three target-item conditions and the pairing of questions with conditions were counterbalanced across participants within each guilt condition.

Procedure

Participants simulating the guilty were instructed to enter a locked room using a key that was handed to them. In this room, they had to search for a colored envelope located in a certain specified place. They were instructed to "steal" this envelope, which contained a sum of money ranging from 3 to 8 New Israeli Shekels or NIS (about $1.20 to $3.20 at the time of the experiment) and an article of jewelry, to hide the money and the jewelry in their pocket, and to enter the examination room. Three classes of relevant items were used: the color of the envelope, the exact sum of money found in the envelope, and the type of jewelry found in the envelope. All participants simulating the innocent individuals were instructed to walk by the locked room downstairs to the examination room and to wait for about 1 min before entering it, but those who simulated the informed innocent participants were

[1] The purpose of the count instructions was to attract the attention of the participants to a neutral item, and they were not asked to provide any verbal report of their count during the interrogation. The target item was presented twice under each target-item condition, so counting was not a cognitively loaded task, but it forced participants to pay attention to the designated item. Indeed, all participants recalled the designated target items in this condition.

[2] This was done in all target-item conditions, but in the no-target condition, no target item was specified to the participants, and the target item was used only for the statistical analyses (e.g., for the manipulation check). Thus, from the participants' perspective there were only two target-item conditions (count and key-press).

informed about the relevant details prior to the polygraph interrogation.

An experimenter who was unaware of the type of the "stolen" envelope and the experimental condition to which the examinee was assigned, attached the polygraph devices and conducted the GKT examination. Participants were told that the experiment was designed to test whether they could cope with the polygraph test and to convince the examiner that they were innocent of stealing the envelope with the money and the jewelry. They were promised a bonus of 5 NIS for a successful performance of the task. All examinees were presented with three different questions, each focusing on a different feature of the mock crime (the color of the stolen envelope, the amount of money it contained, and the type of jewelry found in it). The questions, which were prerecorded, were presented through the Macintosh II computer and were formulated according to the GAT version (e.g., "Was the color of the envelope you stole . . . Blue?").

Five alternative answers were used for each question, and each of these five items was repeated twice following the presentation of the question. Examinees were requested to respond "No" to every item. A buffer item followed the initial presentation of the question, and thus there were a total of 11 alternative answers for each question (a buffer item followed by two repetitions of each of the five alternative items). The order of the five items presented following each question was randomly determined, and the interstimulus interval (ISI) ranged from 16 to 24 s, with a mean ISI of 20 s. The interval between the end of one question and the presentation of the next was used for a short rest, and, prior to the presentation of the new question, its topic was declared, and the appropriate instructions regarding the target items were given. Each question was presented with a different target task (i.e., no target items, counting target items, and pressing a key in response to the target items). The questions were presented in a predetermined order, counterbalanced across examinees within each guilt condition. At the end of the questioning session, all participants were asked whether they recalled the relevant items. In addition, they were requested to report the number of target items counted in the count-items condition. Finally, all participants were debriefed and paid.

Response Scoring and Analysis

Electrodermal responses. Responses were transmitted in real time to the Macintosh II computer. The maximal conductance change obtained from the examinee, from 1 s to 5 s after stimulus onset, was computed using an analog-to-digital (NB-MIO-16) converter with a sampling rate of 1000/s. To eliminate individual differences in responsivity and permit a meaningful summation of the responses of different examinees, each participant's conductance changes were transformed into within-examinees standard scores (Ben-Shakhar, 1985). To minimize habituation effects, within-questions standard scores were used (Ben-Shakhar & Dolev, 1996). Thus, the z scores used in this study were computed relative to the mean and standard deviation of the participant's responses to 9 of the 11 items (excluding the two repetitions of the target item) of each question.

Respiration. The respiration responses were defined as the total respiration line length (RLL) during the 15-s interval following stimulus onset. Each RLL was computed using a sampling rate of 20/s. Similar standardization transformation was used for the RLL as was described earlier in relation to the electrodermal measure, but in the case of respiration, guilty knowledge is reflected by smaller rather than larger RLLs.[3]

The variables of interest in the GKT paradigm are the relative responses evoked by the relevant stimuli (i.e., the features of the stolen envelope). Each guilty examinee actually took an envelope, and the features of that envelope defined the relevant stimuli for him or her. The features that were disclosed to the informed innocent participants defined the relevant stimuli for these participants. Finally, an envelope was randomly chosen for each uninformed innocent examinee, and its features defined the relevant stimuli for that examinee. The z scores corresponding to the responses evoked by these relevant stimuli were used as the dependent variables in all subsequent statistical analyses. A rejection region of $p < .05$ was used for all statistical tests.

Results

An initial analysis was performed to check whether the target-item manipulation produced the expected response pattern. For this purpose, the mean standardized responses to the target items were computed across the two presentations of the target items and across participants within each condition. The mean skin conductance response (SCR) z scores, computed for the 3 target-items conditions, across guilt conditions revealed the expected order (0.13, 0.75, and 1.05 for the no-target, count-targets, and key-press conditions, respectively). The RLL mean z scores did not show the same pattern (the 3 means were 0.06, -0.09, and -0.06, for the no-target, count-targets, and key-press, respectively). Two 3×3 analyses of variance (ANOVAs) were conducted on these data, with the 3 guilt conditions serving as a between-participants factor and the 3 target-item conditions serving as a within-participants factor. A statistically significant main effect was obtained for the target-item factor with the SCR measure, $F(2, 210) = 74.25$, $\varepsilon = .98$. No significant main effect was obtained for the guilt factor, and no significant interaction between the two factors was obtained. The ANOVA conducted on the RLL measure did not reveal any statistically significant outcomes. These results suggest that the target-item manipulation did attract the participant's attention and produced the expected pattern of orienting responses, but this was reflected only by the electrodermal component of the orienting response and not by the respiration component.

To examine the main hypotheses of this study, the mean standardized response of each examinee on each measure (SCR, RLL, and their combination, defined as the differ-

[3] As a result of mechanical problems with the respiration-measurement equipment, the RLL data of several participants were lost (8 guilty, 12 informed innocent participants, and 5 uninformed innocent participants). Thus, all the analyses of the RLL and the combined measure are based on restricted sample sizes.

ence between the SCR z score and the RLL z score) was computed across the two presentations of the relevant item, within each target condition, and across target conditions. These means, which were averaged across participants within each guilt condition, are displayed in Table 1. A 3 × 3 ANOVA with repeated measures for the three target-item conditions was conducted on the data of Table 1 for each measure. In all cases, only the between-participants factor of guilt produced statistically significant effects, $F(2, 105) = 36.23$, for the SCR; $F(2, 80) = 6.13$, for the RLL; and $F(2, 80) = 30.48$, for the combined measure.

Post hoc comparisons, using the Scheffé method, conducted across target conditions revealed a statistically significant differentiation between the guilty and the informed innocent participants with all three measures, $F(2, 105) = 25.80$, for the SCR; $F(2, 80) = 10.94$, for the RLL; and $F(2, 80) = 30.90$, for the combined measure. In all these cases, guilty participants showed larger differential responsivity to the relevant items than the informed innocent participants. Statistically significant differentiation between informed (guilty and informed innocent participants) and uninformed participants was obtained with the SCR, $F(2, 105) = 47.30$, and the combined measure, $F(2, 80) = 32.70$, but not with the RLL. Finally, only the SCR measure significantly differentiated between the informed and uninformed innocent participants, $F(2, 105) = 10.70$.

In addition to group data, it is interesting to analyze and compare classification accuracy rates of individual participants. However, because we used a within-subjects manipulation of the target-item conditions in this study, each target condition was based on a single question. Because accuracy rates typically increase as a function of the number of GKT questions, it is clear that the accuracy rates obtained within each condition of this study are bound to be smaller than those typically obtained in GKT studies (which use at least four but often six or more different GKT questions). Nevertheless, we decided to present accuracy rates for each group of participants on the basis of each measure, in each experimental condition, as well as across target conditions.

Because detection accuracies obtained from a single condition (and a single question) and a single physiological measure were compared with accuracies obtained from combinations of three target conditions and two physiological measures, it was necessary to transform the various detection measures into a unified scale (otherwise, each detection measure would have a different variance, which would affect the probability of exceeding the cutoff score). To achieve this goal, we followed the procedure described in detail by Elaad and Ben-Shakhar (1997), and we standardized each detection measure across the entire sample. This allowed us to set a single cutoff score that had a similar meaning for all detection measures. A cutoff was set at 0.20, and each detection score of each examinee was compared with this cutoff score such that all participants scoring below it (above it in the case of RLL) were classified as innocent participants, and all other participants were classified as guilty. Table 2 presents rates of correct classifications of guilty and innocent participants on the basis of each measure, within each target condition, and across target conditions.

Inspection of Table 2 reveals that, as expected, detection rates based on a single question are rather low. However, when pooling the data across conditions and physiological measures, detection rates are much higher (90% of the uninformed innocent participants and 79% of the guilty participants were correctly classified). Moreover, only 58% of the informed innocent participants were correctly classified as innocent, and this rate is more or less similar to the findings of Bradley et al. (1996) with the GAT questioning format. A comparison of the rates of guilty classifications

Table 1
Means and Standard Deviations of the Mean Standardized Responses to the Relevant Items, Within Each Target Condition and Across Target Conditions

		No-target		Weak-target		Strong-target		Across-targets	
Condition	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
SCR									
Guilty	36	.70	.70	.65	.60	.68	.66	.68	.46
Informed innocent	36	.32	.73	.18	.65	.15	.60	.21	.37
Uninformed innocent	36	−.03	.58	−.25	.49	−.03	.66	−.10	.34
RLL									
Guilty	28	−.26	.69	−.40	.75	−.45	.58	−.37	.45
Informed innocent	24	−.17	.60	−.10	.53	.11	.61	−.06	.28
Uninformed innocent	31	−.13	.57	−.10	.53	−.11	.62	−.09	.30
Combined									
Guilty	28	.75	.78	.59	.68	.58	.55	.64	.44
Informed innocent	24	.34	.69	.06	.60	−.15	.62	.09	.39
Uninformed innocent	31	−.06	.58	−.19	.51	−.01	.61	−.10	.27

Note. For the respiration line length (RLL) measure, stronger responses are reflected by smaller values. SCR = skin conductance response.

Table 2
Correct Classification Rates of Guilty, Informed and Uninformed Innocent Participants, Under Each Target Condition and Across Target Conditions, Obtained With Each Physiological Measure and Their Combination

Condition	*N*	No-target	Weak-target	Strong-target	Across-targets
SCR					
Guilty	36	.67	.72	.72	.72
Informed innocent	36	.64	.61	.69	.61
Uninformed innocent	36	.83	.83	.75	.89
RLL					
Guilty	28	.36	.64	.68	.57
Informed innocent	24	.63	.67	.75	.67
Uninformed innocent	31	.55	.55	.58	.71
Combined					
Guilty	28	.68	.57	.71	.79
Informed innocent	24	.54	.71	.79	.58
Uninformed innocent	31	.87	.84	.74	.90

Note. SCR = skin conductance response; RLL = respiration line length.

among the guilty participants (correct classifications), and the informed innocent participants (false-positive classifications) reveals statistically significant[4] differences with the SCR measure (z = 2.82) and the combined measure (z = 2.74), but not with the RLL measure (z = 1.73). On the other hand, comparisons between correct classification rates of the two innocent groups revealed that with the SCR and the combined measure, informed innocent participants were significantly less accurately classified as innocent than the uninformed innocent participants (z = 2.69 for the SCR measure, and z = 2.76 for the combined measure). These results are more or less similar to those obtained with the continuous measures and show that at least with the SCR, the informed innocent participants fall in between the other two groups in terms of their responses to the critical items and their guilty classification rates (false-positive rates).

With regard to the target-items manipulation, the classification-rate analysis highlights two interesting features of the results. First, no indication was found that the introduction of target items interferes in any way with the detection of guilty participants. Surprisingly enough, there are even slight improvements in the detection rates of guilty participants (for both physiological measures) with the introduction of the target items. Second, inspection of Table 2 reveals also that at least when the key-press condition is compared with the no-target condition, the correct detection rates of informed innocent participants tend to improve. This trend seems stronger for the combined measure, where the correct detection rate of the informed innocent participants increased from 54% in the no-target condition, to 71% in the count condition, and to 79% in the key-press condition. However, these observations should be treated very cautiously, because the differences are not statistically significant.

Because the correct classification data depend on a single, arbitrary, cutoff point and because it is difficult to compare different experimental conditions on the basis of correct classification rates (i.e., gains in one type of classification are often accompanied by losses in the other classification category), an additional approach for describing and comparing detection efficiency was adopted from signal detection theory. Following the practice applied in previous studies (e.g., Ben-Shakhar & Gati, 1987; Elaad & Ben-Shakhar, 1989, 1997), a statistic describing detection efficiency by comparing the entire distributions of z scores to the relevant alternatives of guilty (or informed innocent) participants and uninformed innocent participants was computed for each measure within each experimental condition, as well as across target conditions. On the basis of these distributions, a receiver operating characteristic (ROC) curve was generated for each group (guilty and informed innocent participants) under each experimental condition. Figure 1 describes the ROC curves, computed on the basis of the combined measure, for the guilty and the informed innocent participants across target-item conditions. The figure demonstrates the differences between guilty and informed innocent participants, and shows that for any possible cutoff point, the correct detection rate of guilty participants exceeds the false classification rate of the informed innocent participants.

In addition, the area under each ROC curve was computed. The area under the ROC curve is a measure of detection efficiency, which does not depend on a single cutoff point, but rather reflects detection efficiency across all possible cutoff points. It assumes values between 0 and 1, such that an area of 0.5 means that the two distributions (the distributions of mean z scores, across the relevant items, of guilty and innocent examinees) are undifferentiated, and it is therefore impossible to use the responses for detecting whether an examinee is guilty or not. An area of 1 indicates that there is no overlap between the two distributions, and a perfect classification of guilty and innocent examinees would therefore be possible on the basis of the mean z scores. A more detailed description of signal detection theory and its applications can be found in several sources (e.g., Green & Swets, 1966; Swets, Tanner, & Birdsall, 1961).

Bamber (1975) described a method for estimating the variance of the area statistic and for computing confidence intervals for the true area. Using this method, we computed 90% confidence intervals for the area in each experimental

[4] For these comparisons, a 0.025 criterion for statistical significance was used because two dependent comparisons were made for each measure (comparing correct detection of guilty and informed innocent participants, and comparing correct detection rates of the two innocent groups).

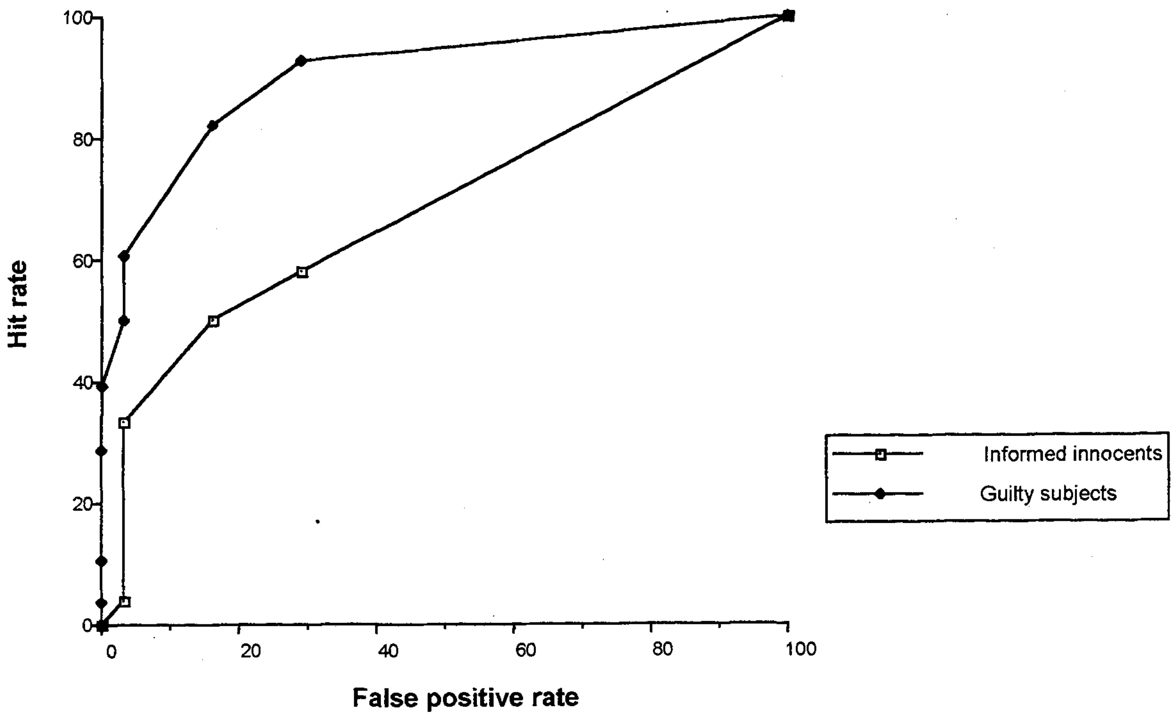

Figure 1. ROC curves, computed on the basis of the combined measure, for guilty and informed innocent participants across target-item conditions.

condition with each physiological measure. The areas under the ROC curves along with the corresponding 90% confidence intervals are displayed in Table 3.

Inspection of Table 3 reveals that although no systematic reduction in the areas under the ROC curves of guilty examinees was obtained as a result of introducing the target items, the areas under the ROC curves of the informed innocent participants tended to decrease with the introduction of these items. In particular, although the SCR areas for guilty examinees were identical under the key-press task and the control, no-task conditions, the areas obtained with the informed innocent condition tended to decrease, and under the key-press condition it did not significantly differ from a chance area of 0.50. The RLL areas were generally smaller, but it is important to note that for the informed innocent participants, they did not differ significantly from a chance area in all three target conditions.

The most remarkable results were obtained with the combined measure. For guilty examinees, the areas were practically the same under all target-item conditions, but for the

Table 3
Area Under the ROC Curves and 90% Confidence Intervals (CIs) Computed Within Each Target Condition and Across Target Conditions

		No-target		Weak-target		Strong-target		Across-target	
Condition	*N*	Area	90% CI	Area	90% CI	Area	90% CI	Area	90% CI
SCR									
Guilty	36	.78	(.69, .87)	.87	(.80, .95)	.78	(.69, .87)	.92	(.87, .96)
Informed	36	.64	(.53, .75)	.69	(.58, .79)	.57	(.46, .68)	.75	(.66, .84)
RLL									
Guilty	28	.56	(.43, .68)	.66	(.54, .78)	.68	(.56, .79)	.70	(.59, .81)
Informed	24	.51	(.38, .64)	.49	(.36, .62)	.42	(.29, .54)	.48	(.36, .60)
Combined									
Guilty	28	.80	(.71, .90)	.82	(.73, .91)	.77	(.67, .87)	.94	(.89, .98)
Informed	24	.69	(.57, .81)	.61	(.49, .74)	.45	(.33, .58)	.65	(.52, .78)

Note. For the skin conductance response (SCR), respiration line length (RLL), and combined measures, the numbers of uninformed innocent participants were 36, 31, and 31, respectively. ROC = receiver operating characteristic.

informed innocent participants, a gradual reduction was observed from 0.69 under the no-target condition to 0.61 under the count condition, and 0.45 under the key-press condition. This means that under the key-press condition, the informed innocent participants were undifferentiated from the uninformed innocent participants.

Finally, we examined whether the physiological responses to the relevant items were related to memory for these items. It turned out that all guilty participants correctly recalled all three relevant items, so no physiological differentiation could be found within this group as a function of recall. However, some variability in recall was observed among the informed innocent participants, and, in this group, recall of the relevant items seemed to be related to the SCRs elicited by these items. Table 4 displays mean SCR z scores and areas under the ROC curves as a function of the number of relevant items recalled by the informed innocent participants. The table reveals that SCR responses to the relevant items increase as a function of recall. On the other hand, the RLL showed no relationship to the number of items recalled.

The within-participants manipulation of the target-items factor raises a possibility of carryover effects. Specifically, it could be argued that if the count-target condition followed the key-press condition, participants could guess that the target would be repeated twice, and therefore their cognitive load would be reduced as compared with participants for whom the count condition preceded the key-press condition. To examine this possibility, we analyzed the responses to both the relevant and target items under the count-target condition and examined whether they were affected by order (i.e., whether the count condition followed or preceded the key-press condition). We also examined whether the memory scores obtained by the informed innocent participants were affected by the same order variable. The results of these analyses revealed that no order effects were obtained, neither with the physiological measures nor with the memory scores.

Table 4
SCR Mean Standardized Responses to the Relevant Items and ROC Areas, as a Function of the Number of Items Recalled by the Informed Innocents

No. of items	N	M	SD	ROC area	90% CI
0–1	14	.09	.37	.68	(.55, .81)
2	11	.15	.37	.69	(.55, .83)
3	11	.44	.30	.88	(.81, .96)
Differences between 0–1 and 3 items		$t(23) = 2.53$*		$z = 2.23$*	

Note. SCR = skin conductance response; ROC = receiver operating characteristics.
* Statistically significant.

Discussion

The results of this study raise several issues related to both theoretical and applied aspects of psychophysiological detection. First, the present results are consistent with those obtained in previous studies (e.g., Bradley et al., 1996; Bradley & Rettinger, 1992; Bradley & Warfield, 1984), in demonstrating that informed innocent participants showed relatively larger electrodermal responses to the critical items than uninformed innocent participants, but smaller than guilty participants (across target conditions, the areas under the ROC curve were 0.92 for guilty examinees and only 0.75 for the informed innocent participants). This finding implies that mere knowledge of the relevant information in itself is not sufficient to create the differential response pattern to the relevant items that is typically observed in guilty participants. In other words, additional factors are affecting differential responding to the relevant information. This conclusion is consistent with the results of several studies demonstrating that both verbal deception and motivation to avoid detection may increase differential responsivity in the GKT paradigm (e.g., Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Horneman & O'Gorman, 1985) and that deception affects electrodermal responsivity under the differentiation of deception paradigm (e.g., Furedy et al., 1988, 1991, 1994). These studies showed that although some psychophysiological differentiation between guilty and innocent participants can be demonstrated, even without deception (e.g., when no verbal responses are required or when "yes" responses are given to all questions) and under low motivational conditions, deceptive responses and increased motivation contribute to increases in differential responding to the relevant items and thus increase the efficiency of psychophysiological detection. Thus, mere knowledge without an actual involvement with the event may be analogous to a low motivation and nondeceptive context.

When the respiration measure was used, no differences between informed and uninformed innocent participants were obtained. It is not clear why the RLL measure is not affected by knowledge of the relevant information, but it is important to note that this measure was less affected than the electrodermal measure by other manipulations. For example, two recent studies reported that mental countermeasures affected psychophysiological detection with the electrodermal measure but not with the respiration measure (Ben-Shakhar & Dolev, 1996; Honts, Devitt, Winbush, & Kircher, 1996). In addition, the target manipulation used in this study affected the SCRs elicited by the targets, but not the RLLs.

The second goal of this study was to examine the possibility that introducing an additional task would draw the attention of the informed innocent participants from the relevant items without affecting the guilty participants. The present results provide an initial demonstration that this manipulation might indeed be effective. The use of the target items tended to reduce the differences between informed and uninformed innocent participants. For example, the average SCR z score to the relevant items decreased in the informed innocent condition from 0.32 in the control, no-target condition to 0.18 and 0.15 in the count and key-press conditions, respectively. Similarly, the RLL z scores increased from -0.17 in the control condition to -0.10 and 0.11, in the two target conditions.

The ROC results showed no consistent trend with the electrodermal measure, but showed a gradual decrease with the combined measure (from 0.69 under the no-target condition, through 0.61 under the count condition, to 0.45 under the key-press condition). These results are consistent with the notions that knowledge of the relevant information is not sufficient for differential autonomic responsivity and that introducing another task may differentially attract the attention of informed innocent participants. This conclusion may be significant from an applied perspective, because if it could be generalized to the real-world context, it might reduce the risk of false-positive outcomes in informed innocent examinees. Furthermore, a successful application of the target-items procedure may enhance the use of the GKT, instead of less standardized and controlled methods of psychophysiological detection.

A third issue that was examined in this study was the relationship between memory for the relevant items and relative responding to these items. The findings indicate that such a relationship was demonstrated for the electrodermal measure. This result is not surprising and is consistent both with previous findings and with theoretical accounts relating electrodermal reactivity to attention (e.g., Corteen, 1969; Waid, Orne, Cook, & Orne, 1978; Waid, Orne, & Orne, 1981). It is interesting to note, however, that even when memory was controlled for (by looking only on those who recalled all three items), larger mean responses to the relevant items were observed in the guilty examinees than in the informed innocent participants. This result indicates that memory alone is not sufficient to account for the differential autonomic responsivity to the relevant items. Furthermore, it supports our conclusion that knowledge and awareness of the relevant information are not the only factors determining differential autonomic responding to the relevant items and that additional factors, such as personal involvement with the event, verbal deception, and motivation to avoid detection, may increase differential responsivity in the GKT (or GAT) paradigm.

Finally, this study provides further information about the relative efficiency of different physiological measures in psychophysiological detection. In this respect, the present results are inconsistent with several recent studies that demonstrated that the RLL measure is as efficient as the electrodermal measure (e.g., Ben-Shakhar & Dolev, 1996; Elaad & Ben-Shakhar, 1997; Elaad, Ginton, & Jungman, 1992; Timm, 1982, 1987). The efficiency of the RLL for differentiating between guilty and innocent participants was considerably less than that of the electrodermal measure. However, the RLL does contribute to the combined measure, which seems to perform slightly better than the SCR for the detection of guilty participants. Moreover, the RLL contributes to reduce false-positive outcomes in the informed innocent participants (across target conditions the ROC area for differentiating informed and uninformed innocent participants with the combined measure is only 0.65, as compared with 0.75 for the SCR). Furthermore, if only the strong target condition is considered, the informed and uninformed innocent participants are undifferentiated with the combined measure, whereas the guilty participants are detected as efficiently as they are with the electrodermal measure.

References

Bamber, D. (1975). The area under the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology, 12*, 378–415.

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

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in psychophysiological responsivity. *Psychophysiology, 22*, 292–299.

Ben-Shakhar, G. (1991). Future prospects of psychophysiological detection: Replacing the CQT by the GKT. In J. R. Jennings, P. K. Ackles, and M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 4, pp. 283–289). London: Jessica Kingsley.

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the Guilty Knowledge Technique: The effects of mental countermeasures. *Journal of Applied Psychology, 81*, 273–281.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Springer-Verlag.

Ben-Shakhar, G., & Gati, I. (1987). Common and distinctive features of verbal and pictorial stimuli as determinants of psychophysiological responsivity. *Journal of Experimental Psychology: General, 116*, 91–105.

Bradley, M. T., MacLaren, V. V., & Carle, S. B. (1996). Deception and nondeception in Guilty Knowledge and Guilty Actions Polygraph Tests. *Journal of Applied Psychology, 81*, 153–160.

Bradley, M. T., & Rettinger, J. (1992). Awareness of crime-relevant information and the Guilty Knowledge Test. *Journal of Applied Psychology, 77*, 55–59.

Bradley, M. T., & Warfield, J. F. (1984). Innocence, information, and the Guilty Knowledge Test in the detection of deception. *Psychophysiology, 21*, 683–689.

Corteen, R. S. (1969). Skin conductance changes and word recall. *British Journal of Psychology, 60*, 81–84.

Elaad, E. (1997). Polygraph examiner awareness of crime-relevant information and the Guilty Knowledge Test. *Law and Human Behavior, 21*, 107–120.

Elaad, E. (1998). The challenge of the concealed knowledge polygraph test. *Expert Evidence*, *6*, 161–187.

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology*, *26*, 442–451.

Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the Guilty Knowledge Test. *Psychophysiology*, *34*, 587–596.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal Guilty Knowledge Tests. *Journal of Applied Psychology*, *77*, 757–767.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, *28*, 531–547.

Furedy, J. J., & Ben-Shakhar, G. (1991). The roles of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. *Psychophysiology*, *28*, 163–171.

Furedy, J. J., Davis, C., & Gurevich, M. (1988). Differentiation of deception as a psychological process: A psychophysiological approach. *Psychophysiology*, *25*, 683–688.

Furedy, J. J., Gigliotti, F., & Ben-Shakhar, G. (1994). Electrodermal differentiation of deception: The effect of choice vs. no choice of deceptive items. *International Journal of Psychophysiology*, *18*, 13–22.

Furedy, J. J., & Heslegrave, R. J. (1991). The forensic use of the polygraph: A psychophysiological analysis of current trends and future prospects. In J. R. Jennings, P. K. Ackles, and M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 4, pp. 157–189). London: Jessica Kingsley.

Furedy, J. J., Posner, R., & Vincent, A. (1991). Electrodermal differentiation of deception: Memory-difficulty and perceived accuracy manipulations. *Psychophysiology*, *28*, 163–171.

Giesen, M., & Rollison, M. A. (1980). Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality*, *14*, 1–11.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gudjonsson, G. H. (1982). Some psychological determinants of electrodermal responses to deception. *Personality and Individual Differences*, *3*, 381–391.

Honts, C. R., Devitt, M. K, Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, *33*, 84–92.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior*, *16*, 357–379.

Horneman, C. J., & O'Gorman, J. G. (1985). Detectability in the card test as a function of the subject's verbal response. *Psychophysiology*, *22*, 330–333.

Kugelmass, S., Lieblich, I., & Bergman, Z. (1967). The role of "lying" in psychophysiological detection. *Psychophysiology*, *3*, 312–315.

Lieblich, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR detection of information as a function of stimulus set size. *Psychophysiology*, *6*, 601–608.

Lykken, D. T. (1974). Psychology and the lie detection industry. *American Psychologist*, *29*, 725–739.

Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector*. New York: McGraw-Hill.

Stern, R. M., Breen, J. P., Watanabe, T., & Perry, B. S. (1981). Effects of feedback of physiological information on responses to innocent associations and guilty knowledge. *Journal of Applied Psychology*, *66*, 677–681.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. C. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.

Timm, H. W. (1982). Effects of altered outcome experiences stemming from placebo and feedback treatments on the validity of the Guilty Knowledge Technique. *Journal of Applied Psychology*, *67*, 391–400.

Timm, H. W. (1987). Effect of biofeedback on the detection of deception. *Journal of Forensic Science*, *32*, 736–746.

Waid, W. M., Orne, E. C., Cook, M. R., & Orne, M. T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of information. *Journal of Applied Psychology*, *63*, 728–733.

Waid, W. M., Orne, E. C., & Orne, M. T. (1981). Selective memory for social information, alertness and physiological arousal in the detection of deception. *Journal of Applied Psychology*, *66*, 224–232.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, *7*, 294–300.

Received June 1, 1998
Revision received November 3, 1998
Accepted November 5, 1998 ■