Journal of Experimental Psychology: Applied
2014, Vol. 20, No. 2, 136–146

1076-898X/14/$12.00 DOI: 10.1037/xap0000015

Psychophysiological Detection of Concealed Information Shared by Groups: An Empirical Study of the Searching CIT

Assaf Breska and Daphna Zaidenberg
The Hebrew University of Jerusalem

Nurit Gronau
The Open University of Israel

Gershon Ben-Shakhar
The Hebrew University of Jerusalem

This study focused on the application of the Concealed Information Test (CIT) to situations in which the crime-related information is shared by a group of suspects but is not available to the investigators (a method known as the "searching CIT," or SCIT). Twenty-two groups, each comprising 4 to 7 participants (115 in total), planned 1 of 2 mock crimes (kidnapping or bank robbery). While planning the crime, each group decided on 5 crime-related critical items (e.g., the city in which the bank was located). Each critical item was chosen from a predefined set of 4 alternatives. At a second stage, the SCIT was administered individually and each participant was tested on the 2 crimes—the actual planned crime, in which the participant was "guilty," and the unplanned crime, in which the participant was "innocent." Two algorithms, adopted from Breska, Ben-Shakhar, and Gronau (2012), were applied to detect the critical items and to differentiate between "guilty" and "innocent" participants. Findings revealed that differentiation efficiency based on electrodermal and respiration measures was identical to that obtained with the standard CIT when applied to large groups, but lower, although significantly greater than chance, when applied to differentiate between small groups.

Keywords: Concealed Information Test, psychophysiological detection, group-level analysis, searching CIT

The use of physiological responses for the detection of deception has attracted the attention of law enforcement agencies and researchers since the beginning of the 20th century (see, e.g., Ben-Shakhar & Furedy, 1990; Marston, 1917; Raskin, 1989; Reid & Inbau, 1977), but it has considerably increased during the last two decades. This enhanced attention can be attributed to the September 11th terror attack in the United States and the subsequent terror activities in Europe, as well as to scientific developments in cognitive neuroscience and modern neuroimaging techniques (for reviews of recent research, see Ben-Shakhar, 2012; Rosenfeld, Ben-Shakhar & Ganis, 2012; Verschuere, Ben-Shakhar & Meijer, 2011). Furthermore, the increased need to detect suspects involved in planning and executing terror activities has raised new questions, as well as new detection methods, that require new research directions.

From a scientific perspective, there is only one method of psychophysiological detection that has been supported by research and theory. This method, which was traditionally labeled the Guilty Knowledge Test (see Lykken, 1959, 1960), but more recently has been termed the Concealed Information Test (CIT; see Verschuere & Ben-Shakhar, 2011), is designed to detect concealed knowledge, rather than deception. It utilizes a series of multiple-choice questions, each having one critical alternative (e.g., a feature of the crime under investigation) and several neutral (control) alternatives. The critical alternatives are significant only for knowledgeable (guilty) individuals, and there is ample evidence, mostly from psychophysiological research on orienting responses, indicating that significant stimuli elicit enhanced orienting responses (e.g., Gati & Ben-Shakhar, 1990; Siddle, 1991; Sokolov, 1963). Thus, if the suspect's physiological responses to the critical alternative are consistently larger than to the neutral (or irrelevant) alternatives, knowledge about the event (e.g., crime) can be inferred. As long as information about the event has not leaked out to innocent suspects,[1] the probability that an innocent suspect would produce consistently stronger responses to the critical than to the neutral alternatives depends only on the number of questions

This article was published Online First February 24, 2014.

Assaf Breska and Daphna Zaidenberg, Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel; Nurit Gronau, Department of Psychology & Cognitive Science Studies, The Open University of Israel, Raanana, Israel; Gershon Ben-Shakhar, Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel.

This research was funded by a Grant 408/09 from the Israel Science Foundation to Gershon Ben-Shakhar. We thank Gidi Aviram, Noa Choder, Lotem Elber, Yulia Levin, Daniel Magriso, Noam Markovitch, Tamar Pelet, Shai Sataran, and Reut Shilo for their assistance in this research. We are also grateful to Peter Rosenfeld and two anonymous reviewers for their constructive and helpful comments.

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

[1] The effects of information leakage on the outcomes of the CIT have been studied extensively, mostly by Bradley and his colleagues (for a review of these studies, see Bradley, Barefoot, & Arsenault, 2011).

and the number of alternative answers per question, and hence it can be controlled such that maximal protection for the innocent is provided. Clearly, the detection of concealed information does not necessarily imply that the suspect is guilty, as other explanations may be offered for the possession of guilty knowledge.

Typically, most research and applications of the CIT have been focused on the attempt to discriminate knowledgeable (e.g., guilty) from unknowledgeable (innocent) examinees. This usage of the CIT rests on the assumption that salient features of the crime are known to the investigators and can thus be used to formulate the CIT questions (e.g., the type of weapon used or other features of the crime scene). However, there are cases in which the precise crime details are not available to the investigators. For example, the Japanese Police, which uses the CIT extensively, applies, in some cases, a modified version of this method, termed “the Searching CIT” (SCIT), to retrieve information that is unavailable to the investigators, such as finding the location of a murder weapon (see, Osugi, 2011). If, for example, a certain location elicits enhanced responses consistently (across repetitions of the question and across physiological measures), it is classified as the true location. Meixner and Rosenfeld (2010) were the first to examine the SCIT with both “guilty” and “innocent” participants. They used the P300 component of the event-related brain potentials and compared the largest average P300 amplitude of each participant with the second largest response. The results revealed that 58% (21 out of 36) critical items were correctly detected. When detection was made at the individual participant level, detection rates were much higher: 10 out of the 12 knowledgeable participants were correctly detected, with no false positives, yielding an area under the receiver operating characteristic (ROC) curve of 0.98.

Recently, several studies (Breska, Ben-Shakhar & Gronau, 2012; Meijer, Bente, Ben-Shakhar & Schumacher, 2013; Meijer, Smulders & Merckelbach, 2010) have examined whether the SCIT can be applied to detect groups planning terror activities, in which the precise details of the plan are not available to the investigators (e.g., the planting of a bomb in a certain location unknown to the investigators). Clearly, all applications of the SCIT depend on some prior knowledge (e.g., through intelligence sources) about the possible critical items (e.g., possible locations of a bomb), such that it is highly likely that the correct one is among them. This requirement may limit the SCIT’s scope of application, although the Japanese Police experience indicates that it is usable in some cases (see Osugi, 2011). Note, however, that in the cases in which it can be applied (e.g., due to prior intelligence information), the SCIT may have significant implications for security and law-enforcement agencies. In particular, it may benefit from the fact that groups planning terror activities often share critical information, and consequently, when several examinees show a similar response pattern (e.g., an enhanced response to a particular item), it can be inferred with relatively high confidence that they belong to the terror group.

Meijer et al. (2010) tested 12 participants who were informed about the details of a planned terror attack, the details of which were not known to the investigator (though it was assumed that the terror-related details are among the different alternatives included in the test). Relying upon group average skin conductance responses (SCRs), the researchers were able to identify the correct alternative for each of the three SCIT questions used. However, this study is of limited external validity, because all participants were exposed to the critical items, whereas in most real-life cases, some suspects may be innocent (i.e., unaware of the critical items). In a subsequent study, Meijer et al. (2013) adopted a dynamic questioning procedure to examine the validity of the SCIT in identifying terror groups. Specifically, 20 groups, each comprising five individuals, participated in the experiment, and each group planned a mock terrorist attack to be executed at a certain street in a certain city and country. Each group had to choose a country (out of five possible European countries), then a city in that country (again out of five options), and finally, a street in the chosen city. To allow for a dynamic questioning approach, all five members of each group were tested simultaneously. They were first asked about the country and, depending on the group’s average SCRs to the five alternative countries, they were next asked about the city (i.e., if the average SCR to a certain country exceeded a predefined threshold, the participants were asked about the five optional cities of that country). The third question referring to street names was determined by the same procedure. The results revealed that 19 out of the 20 countries were correctly detected; in 13 out of these 19 countries, the city was correctly detected; and in seven of the 13 cities, the street name was also correctly detected.

Breska et al. (2012) adopted a different approach, which is also based on information shared by group members. They examined several algorithms designed to detect the critical items, as well as differentiate between knowledgeable and unknowledgeable participants, in the SCIT by reanalyzing three data sets from previously published CIT studies. Their analysis was based on the assumption that the critical items were unknown to the investigators but were included among the alternative items presented to the subjects. Specifically, they examined two classes of algorithms. The first was based on averaging responses across subjects to identify critical items, and then averaging responses across the identified critical items to identify knowledgeable subjects. The second class was based on the response profiles of all other subjects, computing the correlations between the response profile of each subject (i.e., the vector of responses across all items of all questions) with the response profiles of all the other subjects, and applying a principle component analysis (PCA) to decompose the correlation matrix into its principal components. A more detailed description of the two algorithms is provided in the Method section of the present study. The results of Breska et al. revealed that, in most cases, all critical items were correctly identified, and the efficiency of differentiation between knowledgeable and unknowledgeable subjects in the SCIT (indexed by the area under the ROC curve) approached that of the standard CIT, for both classes of algorithms. Importantly, the robustness of these results to variations in the number of knowledgeable and unknowledgeable subjects in the sample was also examined. This analysis demonstrated that the performance of these algorithms is relatively robust to changes in the number of innocent individuals examined in each group, provided that at least two (but desirably five or more) knowledgeable examinees are included. Although these results seem promising, they were derived from a post hoc analysis of existing CIT data sets. Clearly, the validity of the SCIT should be additionally examined in new experiments involving group planning, rather than executing, illegal activities.

The main purpose of the present study was therefore to examine the validity of the SCIT, using the algorithms developed by Breska et al. (2012) in differentiating between groups who actually planned a mock crime, and were thus aware of its details, and groups of innocent (unaware) examinees, when the precise mock crime details were unknown to the investigators. Specifically, the present study focused on the idea that groups sharing common information would display similar response patterns, which would facilitate the detection of the unknown crime details and the differentiation between guilty and innocent participants. It differs from the Meixner and Rosenfeld (2010) study, which focused on the detection of individuals, rather than groups, and used event-related potentials rather than autonomic measures, which are currently used in almost all applications of psychophysiological detection. It also differs from Meijer et al. (2010), who made no attempt to discriminate between knowledgeable (guilty) and unknowledgeable (innocent) participants. In addition, the present study differs from Meijer et al. (2013), who used a dynamic approach, which may be less applicable because it requires that all suspects are tested simultaneously. Finally, it differs from all these studies in attempting to experimentally validate the algorithms developed by Breska et al. (2012), which were tested on preexisting data and simulations.

To this end, groups of four to seven participants planned one of two mock crimes. Subsequently, all participants were examined individually with respect to *both* mock crimes (i.e., the planned as well as the unplanned crime), and for each of the crime scenarios, the algorithms were applied to identify the critical items and to discriminate between knowledgeable and unknowledgeable groups. Note that each participant was "guilty" in one of the scenarios (i.e., the planned crime) and "innocent" in the other scenario (the unplanned crime).

In addition, we examined the validity of the relevant–relevant comparison question test (RRT) proposed by Kircher and Raskin (as cited in Marchak, 2013).[2] This test consists of direct questions similar to those used in the Comparison Question Test (CQT). The CQT, which is the most frequently used polygraph technique in practice, has been severely criticized mainly because it relies upon improper control questions (e.g., Ben-Shakhar, 2002; Iacono & Lykken, 2002; Lykken, 1974). Typically, responses to relevant, crime related questions (e.g., "Did you plan a bank robbery?") are compared with responses to comparison questions, unrelated to the crime under investigation (e.g., "Did you steal something during the past 5 years?"). Unfortunately, these two types of questions are not equivalent, and both guilty and innocent suspects are likely to be much more concerned with the relevant questions directly related to the crime under investigation than with the comparison questions related to nonspecific issues. Consequently, the CQT is associated with a relatively large proportion of false-positive outcomes (e.g., Ben-Shakhar & Furedy, 1990). However, in the RRT, participants are asked about two mock crimes, the actual planned crime (in which the participants are "guilty"), and an unplanned crime (in which they are "innocents"), and therefore the unplanned crime details may serve as a valid control for the planned crime in the RRT. As the crime scenarios are of more or less equal valence, the comparison between responses elicited by direct questions related to one mock crime with responses to direct questions related to the other mock crime are more meaningful than the ambiguous, nonspecific comparison questions used in the CQT. Thus, the current experimental design presents an opportunity to apply the RRT, as participants are suspected of planning two mock crimes.

The RRT is similar to the "guilt complex technique" described in some polygraph textbooks (Reid & Inbau, 1977), but it has hardly received a systematic experimental investigation. So far, it has been examined by Cook et al. (2012), who applied it with ocular-motor measures (changes in pupil size and fixation time in response to true and false statements), and more recently by Marchak (2013). This more recent study is particularly relevant to the present context, as Marchak applied the RRT to detect mal intentions, rather than actual crimes. In two experiments designed to detect mal intentions, using a combination of eyeblink measures, Marchak reported sensitivity and specificity of 60% and 72.4%, respectively (in Experiment 1), and 78.1% and 68%, respectively (in Experiment 2).

Note that the RRT also somewhat resembles a modification of the CQT, introduced by Bradley, MacLaren, and Black (1996) and further studied by Cullen and Bradley (2004). In this modified version of the CQT, the control questions are clear and direct questions, similar to those used in the CIT (e.g., if the relevant question was, "Did you steal $100?" a possible control would be, "Did you steal $150?"). The advantage of the RRT over the modified CQT is that it does not rest on the assumption that perpetrators notice and remember critical items and that these items are successfully concealed. On the other hand, the RRT is of limited usage because suspects must believe that they are truly suspected of committing two crimes rather than one. An example of a possible application of the RRT is in airport security, in which suspects of smuggling explosives can be questioned in addition about smuggling illegal drugs, and vice versa.

Method

Participants

One hundred fifteen Hebrew University of Jerusalem undergraduate students (76 females and 39 males) participated in the experiment for course credit or payment. Their mean age was 23.9 (SD = 3.7) years. Twenty-two groups, each ranging between four and seven participants, were recruited through ads placed on notice boards throughout the campus. However, a few participants who participated in the planning stage of the experiment did not show up for the second (i.e., SCIT) phase. In addition, some participants were eliminated because their physiological responses were not properly measured due to excessive movements during the SCIT. In particular, in one group, only three participants were left after subjects' elimination; consequently, this entire group was eliminated from the data analyses (to accommodate with our a priori decision that only groups of at least four participants would be analyzed). Thus, the data analyses are based on 21 groups, each comprising between four and six participants, for a total of 101 participants. All participants signed a consent form indicating that participation was voluntary and that they could withdraw from the experiment at any time without penalty.

[2] Kircher and Raskin did not publish their proposal of using the RRT format, but it was included in a research proposal cited by Marchak (2013).

Apparatus

Skin conductance was measured by a constant voltage system (0.5 V Atlas Researches, Hod Hasharon, Israel). Two Ag/AgCl electrodes (0.8-cm diameter) were used with a 0.05 M NaCL electrolyte (TD-246, Discount Disposables, St. Albans, VT). Respiration was recorded by a pneumatic tube positioned around the thoracic area. The experiment was conducted in an air-conditioned chamber, and an NEC CF-500 computer was used to control the stimulus presentation and compute skin conductance respiration. The stimuli were displayed on the computer monitor.

Design

The 21 groups planned a mock crime and were randomly assigned either a bank robbery (10 groups) or a kidnapping (11 groups) as the crime. Then all participants were tested individually about both mock crimes. Thus, one set of 11 groups had knowledge of the details of the kidnapping scenario (i.e., simulated guilty participants) and had no knowledge of the bank robbery scenario (i.e., simulated the innocents), and the other 10 groups simulated guilty and innocent participants in the robbery scenario and the kidnapping scenario, respectively.

Procedure

The experiment was conducted in two stages. In the first stage, each group met with Experimenter 1, who explained the general purpose of the experiment and instructed the group to plan a mock crime (either a bank robbery or a kidnapping). Each mock crime scenario consisted of five categories of information (e.g., in the robbery scenario—the bank's name, its location, and so on), and each one of these categories comprised six items in total (i.e., six bank names, six cities, and so on, for a specification of all the items in each category within each crime; see Appendix). However, each group was exposed at this stage to only four items, as one item was used as a buffer in each category and was therefore introduced to the participants only subsequently, during the CIT stage. In addition, for each group, we removed the items chosen by the previous group who planned the same crime. This procedure was adopted in order to avoid, as much as possible, the emergence of "favorite" items that are being repeatedly chosen. Consequently, the final choice was made between 4 items, but the buffer and the removed item were presented during the CIT.

The different categories, as well as the items within each category, were presented to the whole group simultaneously on a computer monitor. The five categories for the bank robbery were the name of the bank, the city, the day of the robbery, the type of the getaway vehicle, and the hiding place of the stolen money. The five categories for the kidnapping scenario were the occupation of the victim, the location of the planned kidnapping, the hour of the planned kidnapping, the hiding place of the victim, and the person to be contacted for ransom demand. The group had to reach a consensus regarding each choice, and when they completed their choices, all of the critical details of the mock crime were presented to the group on a computer monitor. In addition, each group member was asked to write down the chosen item of each category. At the end of the planning stage, the participants were told that the next stage would include a polygraph test, as well as a memory test.

In the second stage of the experiment, the SCIT was administered to each participant individually by Experimenter 2, who was unaware of the nature of the planned mock crime and its critical (chosen) items. For most participants, the administration of the SCIT took place either immediately after the first stage or during the next day. Few participants were tested 2 days after the planning stage. Experimenter 2 informed the participants that they were suspects in planning two crimes and that they would be tested with the polygraph to determine whether they were guilty or innocent. It was further explained that the experiment was designed to test whether they could cope with a lie detection test and convince the examiner that they are innocent of planning either of these crimes. To increase the participants' motivation to avoid detection, we used instructions similar to those employed by Gustafson and Orne (1963) and subsequently by many other researchers. Specifically, it was emphasized that "beating" this test is a difficult assignment, in which only highly intelligent people can succeed. In addition, participants were promised a bonus of 10 New Israeli Shekels (about $2.50) for successful performance of the task. This bonus was promised to each member of the entire group and was conditional upon the joint outcomes of the test (i.e., if the test would fail at classifying the group). Subsequently, the participant was attached to the electrodes and the pneumatic tube, and the SCIT examination was conducted.

The SCIT was preceded by an initial rest period of 2 min, during which skin conductance baseline was recorded. All examinees were questioned about the five categories of one crime scenario and then about the five additional categories targeting the second scenario. A question was formulated for each category (e.g., "On what day did the bank robbery take place?"). The order of the two sets of questions was counterbalanced across groups, yet all of the participants of a given group were tested using the same order of the two sets. The order of the five questions within each scenario was determined randomly for each participant. The questions and the items (i.e., possible answers, such as "Friday," "Sunday") were presented on the computer monitor, and the questions were simultaneously heard through the computer speakers. Each question was followed by a buffer item, designed to absorb the initial orienting response, and then by the other five items (including the item that was not among the choice options at the planning stage). Each question was presented twice and the order of the items within each question was randomly determined for each participant. Each question was presented for 5 s, and each item was presented for 5 s. The interstimulus interval (blank screen) ranged randomly from 8 to 12 s, with a mean of 10 s. Participants were asked to respond verbally, saying "no" to every item.

The SCIT was followed by the RRT. This test consisted of six direct questions, presented in two sets, such that each set included one irrelevant question and two relevant questions (one for each crime). The first question in each set was an irrelevant question that required a "no" answer (e.g., "Are you standing up now?"), followed by two relevant questions, one targeting the bank robbery (e.g., "Did you participate in the planning of a bank robbery?") and the other targeting the kidnapping (e.g., "Did you take part in the planning of a kidnapping?"). Slightly different relevant questions were used in the two sets. The order of the two sets, as well as the order of the relevant questions within each set, was determined randomly for each participant. The questions' durations were 10 s and the ISIs were similar to those used in the SCIT. The same physiological measures used in the SCIT were also measured for each RRT question. At the end of this test, participants were detached from the electrodes and were informed that the polygraph investigation was completed. Then participants performed a recognition memory test, which consisted of the five questions related to their planned mock crime. Each question was presented on the computer screen along with six alternative answers, and participants were instructed to choose the correct answer. In addition, participants filled out a questionnaire regarding their performance in the experiment. Specifically, they were asked about their motivation to beat the tests, whether or not they used a strategy during the tests, and so forth. Finally, all participants were debriefed and compensated.

Response Scoring

For electrodermal responses, examinees' responses were transmitted in real time to the computer. The maximal conductance change obtained from the examinee, from 1 s to 5 s after stimulus onset was computed, using an A/D (NB-MIO-16) converter with a sampling rate of 20 per second. To eliminate individual differences in responsivity and permit a meaningful summation of the responses of different participants, each examinee's conductance changes were transformed into within-subjects standard scores (Ben-Shakhar, 1985). The *z* scores used in this study were computed relative to the mean and standard deviation of the examinee's response distribution within each question. Within-questions *z* scores were used because they are more resistant to habituation and are therefore more efficient (Elaad & Ben-Shakhar, 1997). Although all items within each question were used for the standardization, the buffer item and the item that was not presented to the participants at the planning stage of the experiment were not included in all subsequent analyses. This was done to avoid spurious responses to these novel stimuli. Finally, a final *z* score for each item was calculated by averaging the two *z* scores across the two presentations of the item.

Respiration responses were defined on the basis of the total respiration line length (RLL) during the 13-s interval following stimulus onset. Timm (1982) noted that the computation of the RLL from the curvilinear respiration pattern might be disproportionally affected by the starting point of measurement. For example, starting from a point in the rapidly ascending inspiration curve, and from a point at the end of the expiration curve, where changes are relatively slow, would produce different RLLs for equal time intervals. To deal with this problem, we followed the procedure used by Elaad, Ginton, and Jungman (1992), and defined each response as the mean of 10 length measures (0.1 s after stimulus onset through 13.1 s after stimulus onset, 0.2 s through 13.2 s after stimulus onset, etc.). In other words, 10 13-s windows were created, each beginning 0.1 s later than the previous window, and the RLL was defined as the mean of the 10 length computed for the 10 windows. Each RLL was computed using a sampling rate of 20 per second. A standardization transformation was applied for the RLL, similar to the one described earlier in relation to the electrodermal measure. But because guilty knowledge is reflected by smaller rather than larger RLLs, the RLL *z* scores were multiplied by -1 and are presented as positive values in all subsequent analyses. A combined measure was defined as a sum of the SCR and RLL Z scores.

For the RRT, the SCRs and the RLLs were standardized within each participant based on the six questions that were presented. Then, an RRT detection score was defined for each participant as the difference between the mean *z* score to the two bank-robbery-relevant questions and the two kidnapping-relevant questions. Thus, positive detection scores are expected for participants who planned the bank robbery and negative detection scores are expected for participants who planned the kidnapping.

Algorithms Used for SCIT Analysis

The data were analyzed using two of the algorithms proposed by Breska et al. (2012). Both algorithms rely on the assumption that as all guilty participants share the same critical information, their response to the critical item within each question will be systematically enhanced relative to their responses to all other items. On the other hand, for innocent participants, none of the items are expected to systematically elicit enhanced responses, and thus responses to all items within each question will reflect random variations. The output of both algorithms is the item in each question that is identified as the shared piece of information (termed "critical item"), and a continuous variable, termed "detection score," which reflects for each subject the likelihood that he or she is guilty. The detection score is then used to examine the efficiency of the algorithm in differentiating between those who actually planned the mock crime ("guilty participants") and those who did not ("innocent participants").

Algorithm 1 consisted of two stages. In the first stage, standardized responses to each item within each question were averaged across all participants (knowledgeable as well as unknowledgeable). The enhanced responses of the guilty participants to the same item within a given question should increase the average response to this item, relative to all other items of that question. Thus, the item producing the maximal mean response within each question was labeled as the critical item for that question. In the second stage, the *z* scores of the "critical items," identified in Stage 1, were averaged within each participant across all questions to create the detection score for this participant.

Algorithm 2 used PCA to analyze the similarities between the patterns of responses of all subjects across items, with the goal of identifying a pattern of responses that is common to a subgroup of subjects. According to basic linear algebra, the pattern of observed responses of each subject across all CIT items (a vector of values we refer to as "response profile") can be represented as a linear combination of a fixed set of latent variables termed "components" (similar to the factors extracted by factor analysis models), such that the response profile of each subject is a different weighted sum of the same set of components. The weight applied to each component in reconstructing the observed profile for each subject is termed "coefficient," and reflects the extent to which the latent component explains the observed response profile of that subject (similar to factor loading in factor analysis models).

PCA is a mathematical method that decomposes a group of vectors into a set of components and extracts the weights required to combine them into the original observed patterns, such that the first component explains the largest possible variance between all vectors, the next component explains the largest possible variance that was not explained by the first, and so on. Under the assumption that the difference between knowledgeable and unknowledgeable individuals is the only systematic source of variance in the data, it is expected to be the largest source of variance between the response profiles of all subjects. As all knowledgeable subjects are expected to show increased responses in the same items (the actual crime related items), their response profiles should be similar to each other. On the other hand, all unknowledgeable subjects are expected to show random responses in all items, and, as such, their response profiles should not resemble those of the knowledgeable subjects. Thus, the first component extracted by PCA should reflect the response profile shared by all knowledgeable subjects, and the coefficient of each subject on this component should reflect the extent to which the response profile of that subject is explained by this component. Therefore, the coefficient of each participant on the first component was used as the detection score. In addition, the values of the first component were standardized within each question, and the item with the largest absolute standard score in each question was identified as the critical item in that question.

The use of the algorithms and the entire data analyses were designed to achieve two major goals: (a) to identify the critical details in each planned scenario, and (b) to classify the groups according to the specific crime they planned. Consequently, the evaluation of the detection efficiency in the present context was based on two measures: (a) the proportion of correctly detected items, and (b) the degree of differentiation between those who planned a given mock crime ("guilty") and those who did not ("innocents"), as reflected by the area under the ROC, constructed by comparing the detection score distributions of "guilty" and "innocent" examinees. We first conducted a "global" analysis comparing the entire sample of "guilty" and "innocents" for each planned crime. However, as, in practice, much smaller groups of suspects are likely to be examined, we conducted an additional "individual group" analysis, comparing each of the 10 groups who planned the robbery mock crime with each of the 11 groups who planned the kidnapping and were unaware of the robbery details (for a total of 110 comparisons), and each group who planned the kidnapping with each of the other groups (again for a total of 110 comparisons). These global and individual group analyses were conducted for each algorithm and for each of the three measures. For the RRT, we also applied the global and the individual group analyses, but, in this case, only average ROC areas were computed, as the critical items cannot be detected with the RRT. The efficiency of the different algorithms was evaluated by comparing the obtained areas under ROC curves with a chance-level area (50%). ROC comparisons between different algorithms or measures were conducted using an appropriate statistical test for comparing different classifiers that are applied to the same sample (Hanley & McNeil, 1983).

Results

Global Analysis

For each algorithm and each mock crime, we constructed a ROC curve, based on a comparison of the respective detection score distributions of those who actually planned the mock crime and those who did not (i.e., who planned the other mock crime). We then computed both the proportion of correctly detected items and the area under each ROC curve, and averaged them across the two mock-crime scenarios (see Figure 1). This procedure was carried out for each physiological measure as well as the combined mea-

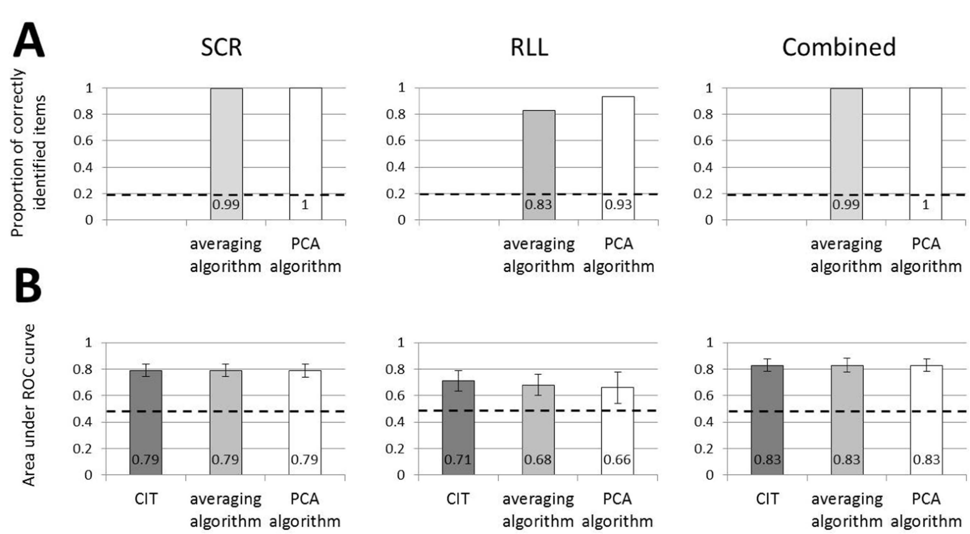

Figure 1. Global analysis results. Proportions of correctly detected items (Panel A) and areas under the receiver operating characteristic (ROC) curves (Panel B), computed for each mock crime, based on the entire sample of "guilty" and "innocent" participants, as a function of algorithm (means vs. principle component analysis) and physiological measure, as well as the corresponding ROC areas computed for the Concealed Information Test (under full knowledge of all critical items). Error bars represent 95% confidence intervals. The horizontal dashed lines (20% in Panel A and 50% in Panel B) represent chance-level performance.

sure.[3] In addition, we constructed a ROC curve for each physiological measure, assuming that all critical items are known (i.e., the standard CIT rather than the SCIT). The areas under these ROC curves can serve to evaluate the relative efficiency of the SCIT of each algorithm, defined as the ratio between the ROC area obtained for the SCIT and the area obtained when using the standard CIT with the same physiological measure. The relative efficiency of the SCIT for each algorithm and each physiological measure is presented in Table 1.

An inspection of Figure 1A reveals that, with the SCR and the combined measures, at least 99% of the critical items were correctly detected, and in most cases, all critical items were detected. The RLL was numerically less efficient than the SCR and the combined measure, but even with this measure, the proportion of correctly detected items was clearly above chance level of 20%, with a 95% confidence interval (CI) of 73% to 92% for the averaging algorithm and 86% to 100% for the PCA algorithm. Detection efficiency, as reflected by the area under the ROC curve, was significantly larger than chance-level performance of 50% for all measures and for both algorithms (all ps < 0.05; Figure 1B). Furthermore, an inspection of Table 1 reveals that when using the SCR or the combined measure, the areas under the ROC curves obtained for the SCIT were practically identical to those obtained for the standard CIT for both algorithms ($z = 0$ in all cases). With the RLL measure, the relative efficiency of the SCIT (as reflected by the ratio between the SCIT and the CIT areas under the ROC curves) ranged between 93% and 100%. Namely, the areas under ROC curves obtained with the SCIT algorithms were numerically smaller than the standard CIT; however, this difference was not statistically significant (averaging algorithm, $z = 0.7$; PCA algorithm, $z = 1.22$; $p > .1$ in both cases). Finally, as expected, comparing the detection efficiency of the three measures in the standard CIT revealed an improved detection efficiency with the combined measure relative to each of the other two measures, though this difference reached statistical significance only with the RLL measure ($z = 3.1$, $p < .05$; all other zs < 1.65, $p > .05$).

To test the validity of the RRT, for each physiological measure we constructed a ROC curve based on a comparison between the RRT detection score distributions of the entire groups of guilty and innocent subjects. The areas under these ROC curves were 0.61 (95% CI [0.51, 0.72]) for the SCR, 0.72 (95% CI [0.63, 0.81]) for the RLL, and 0.75 (95% CI [0.66, 0.84]) for the combined measure.

Table 1
Relative Detection Efficiency of the SCIT Compared With the Standard CIT (i.e., When Crime Details Are Known), Computed as the Ratio Between the ROC Area Obtained for the SCIT to the Respective Area Obtained for the Standard CIT

	Averaging algorithm		PCA algorithm	
	Global analysis	Group-wise analysis	Global analysis	Group-wise analysis
SCR	100%	89%	100%	84%
RLL	96%	82%	93%	72%
Combined	100%	90%	100%	77%

Note. The relative detection efficiency of the SCIT is presented for each algorithm, as a function of analysis type and physiological measure. SCIT = Searching Concealed Information Test; CIT = Concealed Information Test; ROC = Receiver Operating Characteristics; PCA = Principle Component Analysis; SCR = Skin Conductance Response; RLL = Respiration Line Length.

Individual Group Analysis

To evaluate the detection efficiency under more realistic conditions of small group sizes, we compared, for each mock crime each "guilty" group (consisting of participants who actually planned this crime) with each of the other "innocent" groups (who planned the other crime). For each comparison of two groups, we examined the proportion of items (out of five) that were correctly identified, and then averaged these proportions across all pairs of groups within each crime and for each physiological measure and each algorithm. In addition, for each comparison of two groups, we constructed a ROC curve on the basis of the detection score distributions of the two groups. We then computed the area under these ROC curves, as well as the average area across all pairs of groups within each crime, each algorithm, and each physiological measure. In addition, we constructed a ROC curve for each pair of groups, assuming that all critical items are known (i.e., the standard CIT). These areas and their average across all pairs of groups provide a measure of the optimal level of group discrimination in the present context, and allow for a comparison between the SCIT and the CIT (see Table 1). These results, averaged across the two mock-crime scenarios, are displayed in Figure 2.

Analysis of the performance of the SCIT algorithms revealed that, in contrast to the global analysis, both the average proportion of detected items (Figure 2A) and the average ROC area (Figure 2B) were lower than those obtained in the standard CIT, for all measures and algorithms. Specifically, whereas the relative efficiency of the SCIT (the ratio between the ROC area obtained for the SCIT and the area obtained when using the standard CIT) in the global analysis ranged between 93% and 100%, in the individual group analysis, it ranged between 72% and 90%. This reduction in detection efficiency of the SCIT relative to the standard CIT, found in the individual group analysis, was statistically significant for both algorithms (all $t_{(10)}$s > 2.5, all ps < 0.05). Notably, however, even with such small groups, performance of both algorithms was significantly larger than chance level for the SCR and the combined measure ((all $t_{(10)}$s > 1.85, all ps < 0.05, one-tailed), and the average efficiency of the averaging algorithm based on the combined measure was 90%.

The individual group analysis of the RRT revealed an average ROC area of 0.62 (95% CI [0.55, 0.69]) for the SCR, 0.72 (95% CI [0.64, 0.78]) for the RLL, and 0.73 (95% CI [0.68–0.80]) for the combined measure.

The results of the recognition test revealed that only five out of the 115 participants failed to recognize one item, and an additional participant could not recognize two items.

[3] All reported data analyses are based on four items within each question (excluding the "novel items" that were not presented in the planning stage but were presented during the CIT). However, we also analyzed the data based on all five items within each question and found that inclusion or exclusion of the novel items had no effect on the results.

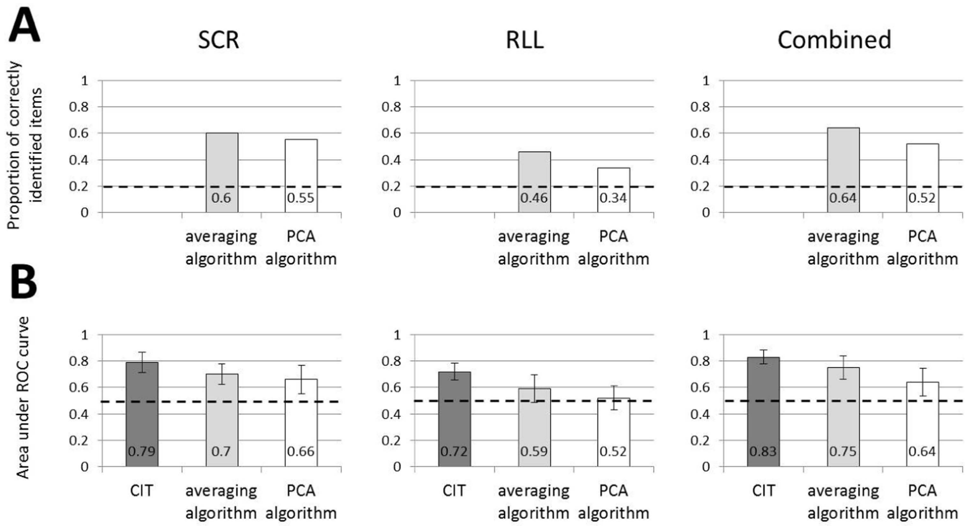

Figure 2. Individual group analysis results. Average proportion of correctly detected items (Panel A) and average receiver operating characteristic (ROC) area (Panel B) computed across all pairs of groups, for each crime, each algorithm (means vs. principle component analysis), and each physiological measure, as well as the corresponding ROC areas computed for the Concealed Information Test (under full knowledge of all critical items). Error bars represent 95% confidence intervals. The horizontal dashed lines (20% in Panel A and 50% in Panel B) represent chance-level performance.

Discussion

The current study examined the validity of the SCIT in detecting concealed information, when the critical items are not available, and in differentiating between guilty and innocent suspects. For this purpose, we utilized two algorithms proposed by Breska et al. (2012).

The results of the global analysis are quite impressive in revealing that the detection efficiency of the SCIT is almost as high as that obtained with the CIT under full knowledge of all critical items. These results are very similar to those reported by Breska et al. (2012), who applied these algorithms to three CIT data sets of previously published studies. Specifically, Breska et al. found, on the basis of global analyses, that, in most cases, all critical items were correctly identified and that the relative efficiency of the SCIT ranged between 89% and 100% (approaching 100% in most cases).

However, both global analyses (the current and the one applied by Breska et al., 2012) were based on relatively large and unrealistic numbers of both guilty and innocent examinees, and thus, for practical purposes, the analysis based on smaller groups would be more meaningful. Indeed, our individual group analysis indicates that SCIT detection efficiency of relatively small groups is lower than what can be achieved when all critical items are known. This result is also consistent with the simulations conducted by Breska et al. on small groups. Specifically, these simulations revealed that when applied to groups of about five guilty and five innocent examinees (which are comparable to our individual group analysis), the relative efficiency of the SCIT is considerably lower. We believe that in order to achieve a meaningful comparison of the present results with those reported by Breska et al., it would be most appropriate to rely on the data set adopted from a mock crime study by Nahari and Ben-Shakhar (2011), which produced CIT detection efficiency of 0.815 with the SCR, which is similar to the average ROC area obtained for the standard CIT in the present study with the same measure (0.79). When this data set was analyzed with only five guilty and five innocent examinees, the average ROC areas reported by Breska et al. (2012) were about 0.65 for the averaging algorithm, and only about 0.60 for the PCA algorithm. In fact, these areas are smaller than the respective areas obtained in the present experiment (0.70 and 0.65 for these two algorithms in the analysis of the SCIT, respectively). Although these differences are small and not statistically significant, they imply that the present results based on planning are not inferior to similar results (reported by Nahari & Ben-Shakhar) based on actual execution of a mock crime.

A comparison of the two types of algorithms reveals that although there were no consistent differences between them under the global analysis, the PCA algorithm was less efficient than the simple averaging algorithm when applied to smaller groups. This relative disadvantage of the PCA was reflected both by smaller numbers of correctly detected items and by smaller ROC areas obtained with each physiological measure. Once again, this result is consistent with the results reported by Breska et al. (2012). Based on the global analyses, Breska et al. demonstrated that, in all three data sets, all critical items were correctly detected with the PCA algorithm, but when applied to smaller groups, it was generally less efficient than the averaging algorithm. It is possible that the PCA algorithm requires relatively large data sets and does not function very well with smaller samples, and, consequently, its scope of application is limited.

The results regarding the relative efficiency of the two physiological measures are consistent with most previous studies (see a review by Gamer, 2011). The advantage of the SCR over the RLL was reflected both by larger proportions of correctly detected items and by larger ROC areas. In addition, the present results revealed, once again, that a combination of the two physiological measures, even when based on simple averaging of standardized responses (i.e., assigning equal weights to the two measures), improves detection efficiency beyond what can be achieved with the most efficient single measure. Thus, a proper evaluation of the SCIT potential validity in practical applications is best reflected by the averaging algorithm, based on the combined measure and applied to differentiate between groups of about five guilty and five innocent examinees. This application yields an average of 63% correctly detected items and an average ROC area of 0.75 relative to an area of 0.83 obtained when all critical items are known.

The present study differs from most previous CIT studies, including those analyzed by Breska et al. (2012), because it was based on planning rather than executing mock crimes. In principle, this difference may affect the results in various ways. On the one hand, execution of a crime may be associated with more significant crime details and with a deeper processing of these items compared with mere planning. On the other hand, planning may require a more focused attention to the crime details, but when executing a real crime, perpetrators may fail to notice some of the details. However, the only attempt to conduct a systematic comparison between committing and planning a mock crime revealed that the CIT with the SCR measure was similarly effective in both conditions (Meijer, Verschuere & Merckelbach, 2010). This conclusion was also supported by Meixner and Rosenfeld (2010), who demonstrated impressive detection efficiency of the P300-based CIT with participants who planned a mock terrorist attack. In the present study, participants had to reach a consensus regarding each item and then write down all of the chosen items. Consequently, memory for crime details was nearly perfect. Although the present study was not designed to systematically compare planning and execution, we showed that its results were very similar to those reported by Breska et al., on the basis of data generated from actual mock crime studies. This is consistent with the results reported by Meijer et al. (2010). Clearly, the issue of comparing planning to execution requires additional systematic research, but if the conclusion that the two procedures yield similar outcomes holds, it would strengthen theoretical approaches, such as orienting response theory, suggesting that mere knowledge of the critical items is sufficient for eliciting enhanced responses to these items.

The present results, based on ANS measures, produced considerably smaller detection efficiency, as reflected by the areas under the ROC curves, than the detection efficiency reported by Meixner and Rosenfeld (2010). Specifically, although the ROC areas in the global and individual group analyses of the present study were 0.83 and 0.75, respectively, the ROC area reported by Meixner and Rosenfeld was 0.98. These differences may be accounted for by the different physiological measures used in these studies. Indeed, a recent meta-analysis of CIT experiments conducted by Meijer, Klein Selle, Elber, and Ben-Shakhar (2013) revealed a general advantage of the P300 measure over three ANS measures included in the meta-analysis. It is also important to note that the present data set produced a ROC area of 0.83, even under the standard CIT (when all the critical items were known to the investigators). On the other hand, when considering the second goal of the SCIT, namely, detecting the unknown critical items, the present results are similar or even slightly better than those reported by Meixner and Rosenfeld. Specifically, although the averaging algorithm based on the combined measure produced averages of 64% and 99% correctly detected items in the individual groups and the global analyses, respectively, Meixner and Rosenfeld obtained 58% correctly detected items. The discrepancy between the moderate success in detecting the unknown items and the nearly perfect differentiation between guilty and innocent participants obtained with the P300 measure is unclear. It is more difficult to compare our results with those reported by Meijer et al. (2013) because both the procedures and the data analysis techniques used in these two studies differed considerably. We can only note that detection of the complete information (i.e., country, city, and street name) was obtained for 35% (seven out of 20) of the groups tested by Meijer et al. (2013). On the other hand, based on simulated innocents, Meijer et al. (2013) obtained perfect specificity.

In addition to the SCIT, the present study included a novel attempt to apply a direct questions test to differentiate between groups who planned a mock crime and those who did not. The RRT is a variation of the frequently applied, but highly controversial, CQT. As explained in the introduction of this article, it differs from the modified CQT introduced by Bradley and his colleagues (Bradley et al., 1996; Cullen & Bradley, 2004), which is based on control questions similar to those used in the CIT. Although we believe that the typical application of the CQT suffers from lack of sufficient control, and thus endangers innocent suspects (e.g., Ben-Shakhar, 2002; Lykken, 1974), the present application of the RRT relies on more appropriate, though not perfect, control questions. Specifically, when suspects are interrogated about two crimes of roughly equivalent valence, direct relevant questions about one crime can serve as control questions for the other, and vice versa. Consequently, the RRT may be applied in specific, yet rare cases in which suspects are interrogated regarding more than one crime. Although it is unlikely that the RRT could be applied to standard criminal investigations, because examinees must believe that they are suspected of committing two crimes (a real one and a fictitious one), it may aid law enforcement in other situations. For example, such a technique may be used in airports with examinees suspected of possessing explosive materials, using a question about illegal drugs as a control. Unfortunately, our attempt to examine the RRT did not yield impressive results, and although ROC areas were significantly larger than a chance area of 0.50, they were smaller than those observed with the SCIT. Specifically, the global analysis revealed an area of 0.75 for the RRT, compared with an average area of 0.83 for the SCIT. This difference was not statistically significant ($z = 1.3$), but if replicated, it may imply that, even under the best circumstances, the use of questions directly targeting deception is inferior to the use of indirect questions targeting concealed information.

It is a bit difficult to compare the present RRT results with those reported by Marchak (2013) because he did not rely on a signal detection analysis. However, applying the Grier (1971) formula to the sensitivity and specificity estimates obtained by Marchak yields ROC areas of 0.75 and 0.82 in Experiments 1 and 2, respectively. Although the results of the first experiment are very similar to our findings, Marchak's second experiment yielded somewhat better discrimination. Additional studies are required to determine whether these differences can be explained by the different physiological measures used, or by some procedural differences between the studies. For example, the present study focused primarily on the SCIT, and, consequently, the RRT was always presented following the SCIT. This might have resulted in habituation of the physiological measures, which has weakened their validity. Interestingly, our RRT data showed an advantage of the RLL over the SCR measure, and, once again, the combined measure was most efficient. It should be noted that this was just a first attempt to examine the RRT with the traditional ANS measures and this method definitely deserves more research.

Finally, some limitations of the present study should be pointed out. First, it is clear that the scope of applicability of both the SCIT and the RRT is rather limited. As mentioned earlier, application of the SCIT depends on some prior knowledge about the various options (e.g., possible locations of explosives). We believe that there are cases, especially in the antiterror combat, in which intelligence sources can supply such information. Likewise, the application of the RRT is also limited because suspects must believe that they are truly being suspected in committing, or planning, two crimes. In addition, the assumption that the two crimes are of equivalent valence does not necessarily hold. However, when judging the practical significance of these methods, one ought to consider not only the frequency of potential application but also its consequences. Detection methods that can be used very rarely, but may nevertheless serve as a powerful means to prevent a terror attack, should be seriously considered by law enforcement agencies.

References

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in psychophysiological responsivity. *Psychophysiology, 22,* 292–299. doi:10.1111/j.1469-8986.1985.tb01603.x

Ben-Shakhar, G. (2002). A critical review of the Control Questions Test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103–126). New York, NY: Academic Press.

Ben-Shakhar, G. (2012). Current research and potential application of the Concealed Information Test: An overview. *Frontiers in Psychology, 3,* 342. doi:10.3389/fpsyg.2012.00342

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective.* New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-3282-7

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 187–199). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511975196.011

Bradley, M. T., MacLaren, V. V., & Black, M. E. (1996). The Control Question Test in polygraphic examinations with actual controls for truth. *Perceptual and Motor Skills, 83,* 755–762. doi:10.2466/pms.1996.83.3.755

Breska, A., Ben-Shakhar, G., & Gronau, N. (2012). Algorithms for detecting concealed knowledge among groups when the critical information is unavailable. *Journal of Experimental Psychology: Applied, 18,* 292–300. doi:10.1037/a0028798

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S. D., Woltz, D. J., & Kircher, J. C. (2012). Lyin' eyes: Ocular-motor measures of reading reveal deception. *Journal of Experimental Psychology: Applied, 18,* 301–313. doi:10.1037/a0028307

Cullen, M. C., & Bradley, M. T. (2004). Positions of truthfully answered controls on Control Questions Tests with the polygraph. *Canadian Journal of Behavioral Science, 36,* 167–176. doi:10.1037/h0087227

Elaad, E., & Ben-Shakhar, G. (1997). Effects of items' repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology, 34,* 587–596. doi:10.1111/j.1469-8986.1997.tb01745.x

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77,* 757–767. doi:10.1037/0021-9010.77.5.757

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 27–45). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511975196.003

Gati, I., & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology: General, 119,* 251–263. doi:10.1037/0096-3445.119.3.251

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin, 75,* 424–429.

Gustafson, L. A., & Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology, 47,* 408–411. doi:10.1037/h0041899

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

Iacono, W. G., & Lykken, D. T. (2002). The scientific status of research on polygraph techniques: The case against polygraph tests. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony,* (Vol. 2, pp. 483–538). St. Paul, MN: West.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43,* 385–388. doi:10.1037/h0046060

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology, 44,* 258–262. doi:10.1037/h0044413

Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist, 29,* 725–739. doi:10.1037/h0037441

Marchak, F. M. (2013). Detecting false intent using eye blink measures. *Frontiers in Psychology, 4,* 736. doi:10.3389/fpsyg.2013.00736

Marston, W. M. (1917). Systolic blood pressure changes in deception. *Journal of Experimental Psychology, 2,* 117–163. doi:10.1037/h0073583

Meijer, E. H., Bente, G., Ben-Shakhar, G., & Schumacher, A. (2013). Detecting concealed information from groups using a dynamic questioning approach: Simultaneous skin conductance measurement and immediate feedback. *Frontiers in Psychology, 4,* 68. doi:10.3389/fpsyg.2013.00068

Meijer, E. H., Klein Selle, N., Elber, L., & Ben-Shakhar, G. (2013). *Memory detection with the Concealed Information Test: A meta-analysis of skin conductance, respiration, heart rate, and P300 data*. Manuscript submitted for publication.

Meijer, E. H., Smulders, F., & Merckelbach, H. (2010). Extracting concealed information from groups. *Journal of Forensic Sciences, 55,* 1607–1609. doi:10.1111/j.1556-4029.2010.01474.x

Meijer, E. H., Verschuere, B., & Merckelbach, H. (2010). Detecting criminal intent with the concealed information test. *The Open Criminology Journal, 3,* 44–47. doi:10.2174/1874917801003020044

Meixner, J. B., & Rosenfeld, J. P. (2010). A mock terrorism application of the P300-based concealed information test. *Psychophysiology, 48,* 149–154. doi:10.1111/j.1469-8986.2010.01050.x

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology, 48,* 733–744. doi:10.1111/j.1469-8986.2010.01148.x

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben Shakhar, & E. Meijer (Eds.), *Memory Detection: Theory and application of the concealed information test* (pp.

253–275). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511975196.015

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 247–296). New York, NY: Springer.

Reid, J. E., & Inbau, F. E. (1977). *Truth and deception: The polygraph ("lie detector") technique* (2nd ed.). Baltimore, MD: Williams & Wilkins.

Rosenfeld, J. P., Ben-Shakhar, G., & Ganis, G. (2012). Detection of concealed stored memories with psychophysiological and neuroimaging methods. In L. Nadel & W. Sinnott-Armstrong (Eds.), *Memory and law* (pp. 263–306). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199920754.003.0011

Siddle, D. A. T. (1991). Orienting, habituation, and resource allocation: An associative analysis. *Psychophysiology, 28,* 245–259. doi:10.1111/j.1469-8986.1991.tb02190.x

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York, NY: Macmillan.

Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology, 67,* 391–400. doi:10.1037/0021-9010.67.4.391

Verschuere, B., & Ben-Shakhar, G. (2011). Theory of the concealed information test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 128–148). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511975196.008

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511975196

Appendix

A Specification of All the Items in Each Category of Each Crime That Were Used in This Study

		Items					
Crime types	Category	Buffer item	1	2	3	4	5
Bank robbery	Bank's name	Leumi	Otsar Ha-Hachyal	Hapoa'lim	Mizrahi-Tefahot	Discount	First International
	Bank's location (city)	Hadera	Ashdod	Haifa	Be'er-Sheba	Rishon-Letziyon	Kfar-Saba
	Day of robbery	Friday	Sunday	Monday	Tuesday	Wednesday	Thursday
	Getaway car	Truck	Motorbike	SUV	Tow truck	Family car	Sports car
	Money hiding place	Your own account	Inside your home	Foreign bank	The woods	Relatives' house	A safe at the office
Kidnap	Kidnapee's occupation	Banker	CEO	Diamond dealer	Oil baron	Real-estate entrepreneur	Art merchant
	Place of kidnap	Park	Residence	Office	Country club	Restaurant	Mall
	Kidnap hour	10:00AM	09:00AM	12:00PM	15:00PM	19:00PM	22:00PM
	Kidnapee hiding location	Construction site	Abandoned house	Bomb Shelter	Hotel	Marina	Warehouse
	Ransom contact	Brother	Wife	Son	Business partner	Lawyer	Parents

Received July 24, 2013
Revision received December 28, 2013
Accepted December 30, 2013 ■