

Trial by Polygraph: Reconsidering the Use of the Guilty Knowledge Technique in Court

Gershon Ben-Shakhar,^{1,2} Maya Bar-Hillel,¹ and Mordechai Kremnitzer¹

Polygraph test results are by and large ruled inadmissible evidence in criminal courts in the US, Canada, and Israel. This is well-conceived with regard to the dominant technique of polygraph interrogation, known as the Control Question Technique (CQT), because it indeed does not meet the required standards for admissible scientific evidence. However, a lesser known and rarely practiced technique, known as the Guilty Knowledge Test (GKT), is capable, if carefully administered, of meeting the recently set Daubert criteria. This paper describes the technique, and argues for considering its admissibility as evidence in criminal courts.

KEY WORDS: Guilty Knowledge Test (GKT); Daubert criteria; polygraph.

INTRODUCTION

Scientists and forensic experts have attempted for many years to develop instruments and methods for the purpose of detecting deception. One notable approach, based on measuring psychophysiological responses by a polygraph, has spawned several methods over the past century (see, e.g., Marston, 1917; Raskin, 1989; Reid & Inbau, 1977). The most common of these is the so-called Control Questions Test (CQT). It is also the one which earned the polygraph the colloquial sobriquet of “lie detector.” In some countries (primarily the United States, Canada and Israel) the CQT is widely used in criminal investigations. Yet courts of law have feared to tread where the police jumped in. United States federal courts have almost universally rejected polygraph evidence (e.g., *Commonwealth of Massachusetts v. Woodward*, 1998; *United States v. Cordoba*, 1998; *United States v. Scheffer*, 1998), as have their Canadian counterparts (see Furedy, 1989, for a review). An Israeli advisory committee, headed by a Supreme Court Justice (State of Israel, 1981), recommended that the results of a polygraph examination—or even whether a defendant had refused to have one—not be admissible evidence in criminal cases (see also Harnon,

¹Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel.

²To whom correspondence should be addressed at Department of Psychology, The Hebrew University, Jerusalem 91905, Israel; e-mail: mskpugb@mscc.huji.ac.il.

1982), except in proceedings of arrest and seizure. This legal skepticism is no doubt attributable to the scientific controversy over the CQT's validity, accuracy, and utility (e.g., Ben-Shakhar & Furedy, 1990; Iacono & Lykken, 1997, 1999; Lykken, 1974, 1998; but see Raskin, Honts, Amato, & Kircher, 1999; Raskin, Honts, & Kircher, 1997).

Two recent papers (Gallai, 1999; Saxe & Ben-Shakhar, 1999) examined the admissibility of the CQT in light of the guidelines set by the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals inc.*, 1993 (*Daubert*, for short). Both reached the same conclusion that, "the results of polygraph examinations, as they are practiced today [i.e., the CQT], should not be admissible evidence in federal courts" (Gallai, 1999, p. 88). We shall not reiterate the arguments presented in these papers in detail, but both showed that the CQT does not satisfy the *Daubert* criteria. Saxe and Ben-Shakhar (1999) focused their discussion on the concept of validity, which is essential in applying the *Daubert* criteria. They adopted Messick's definition of validity (Messick, 1989, 1995) and showed that the CQT lacks several critical components of construct validity (to use Messick's terminology, it meets neither substantive, or theoretical, validity, nor external validity). A particularly problematic feature of the CQT from the legal perspective is its lack of discriminant validity (see, Campbell & Fiske, 1959): the test outcomes can reflect various constructs, other than deception, such as surprise, fear and stress. Gallai (1999) focused on the four criteria set by *Daubert* (testability, known error rate, peer review and publication, and general acceptance) and demonstrated that the CQT does not meet these criteria, with the possible exception of the third. Gallai (1999) concluded that the dangers of admitting CQT evidence are great, although its benefits are few, if any, and that "The Federal Rules of Evidence are prepared to defend against admitting such evidence" (p. 116).

There exists, however, another method of polygraph investigation, called the Guilty Knowledge Technique or the Concealed Information Test (GKT, for short). Its lack of popularity in practice (Podlesny, 1993)—except, apparently, in Japan (see Ben-Shakhar & Furedy, 1990; Fukumoto, 1980; Nakayama, 2002; Yamamura & Miyata, 1990)—is probably due to its being so much harder to implement than the CQT. But unlike the CQT, it is based on sound scientific principles. Rather than picking up the physiological accompaniments of guilt and lying, as the CQT purports to do, it involves detecting stimulus significance and attention, from which guilt is then circumstantially inferred. Although this makes it a better candidate for being admissible evidence, most debates and discussions of polygraphy have focused exclusively on the CQT, or lumped the two techniques together. Thus, the familiar and well-conceived legal objections to the CQT were generalized to the GKT (see, however, the minority opinion in *State of Israel*, 1981), and GKT results, like those of the CQT, have not been presented in court.

The purpose of the present paper is to argue that the GKT, considered on its own merits, has much better prospects than the CQT to meet the *Daubert* criteria and therefore the question of its admissibility should be specifically considered. If it can be shown that the GKT can be adapted to meet the legal requirements from admissible evidence, it would be a boon to the trier of fact, in certain cases.

A Description of the GKT Through a Hypothetical Crime

Imagine the following scenario: At 2 a.m. on April 17, the central safe at the lower level of the Bank of the Rich was broken into (as indicated by a clock, which records every opening of the safe). An amount of 1.2 million dollars, in six bags of \$200,000 each, was stolen. The thief carried the bags out two at a time, dragging them on the floor and leaving trail marks. The thief fled from the scene of the crime in a stolen red Honda Accord, leaving behind in the safe a Wrigley's gum wrapper. The police apprehend a suspect, who is subjected to a polygraph interrogation.

For a prototypical GKT, a series of questions is prepared, which could pertain to various aspects of the crime. The idea is that each question should focus on a detail that can be known only to someone closely familiar with the target event, someone who has actually been at the scene of the crime or involved in it. The correct detail is embedded among a number of erroneous details, called "distractors," or "controls." Sample questions could be

1. The amount you stole is
1.0 million, 1.2 million, 1.4 million, 1.6 million, 1.8 million.
2. Your getaway car was
A Nissan, A Honda, A Toyota, A Ford, A Chrysler.
3. The theft took place at
10 p.m.; 12 p.m.; 2 a.m.; 4 a.m.; 6 a.m.

Usually, examinees are instructed to say "no" to all possibilities (although other procedures, such as allowing the suspect to remain silent, or repeating the item, also exist). The suspect's physiological responses to the correct details are compared to those elicited by the incorrect ones. The physiological measures typically monitored during polygraph tests include changes in electrodermal activity (e.g., skin conductance responses), changes in the pattern of respiration (e.g., respiration line length), and changes in cardiovascular measures, such as relative blood pressure and heart rate (for a detailed description of these measures, see Ben-Shakhar & Furedy, 1990; Podlesny & Raskin, 1977). A pattern of consistently stronger responses elicited by the correct details than by the distractors is interpreted as an indication that the examinee recognizes these details.³ Thus, it is the existence of so-called "guilty knowledge," not of guilt itself, that the GKT attempts to detect. Note that relevant items are correct details or facts, and false details are meant to serve as their controls. If the getaway car were, in fact, a Honda, then Nissan, Toyota, Ford, and Chrysler are the controls. To someone who does not know the make of the getaway car, there is nothing to signal which is the relevant detail. Ideally, no detail within a question is intrinsically more

³Various rules have been applied to classify individuals as "guilty" or "innocent" on the basis of their response pattern to the relevant and the control items. Clearly, we cannot describe all of them here, but to illustrate, we present the most widely used classification rule proposed by Lykken (1959). According to his proposal, the responses to all alternatives of each question are rank-ordered. If the relevant alternative elicits the largest response, a value of 2 is assigned to the question; if it elicits the second largest response, a value of 1 is assigned to the question; otherwise, a value of 0 is assigned. These values are then summed across all k questions (or repetitions) to produce a single detection score (ranging between 0 and $2k$) for each participant. A cutoff of k is then set on this detection score, such that a detection score larger than k yields a "guilty" classification, whereas a score of k or less yields an "innocent" classification.

threatening than the other, and the only thing to set apart the correct detail from the distractors is its correctness, which is detectable only to someone possessing knowledge of the crime details—the so-called “guilty knowledge.” A person professing to know nothing about the crime, but whose responses do, nonetheless, systematically distinguish among them (namely, a person whose responses to the relevant details are systematically more intense), has, at the very least, something to account for: From whence the guilty knowledge? Of course, a person who admits right away to knowing some details of the crime (e.g., an eye-witness) need not be questioned about them.

The Rationale of the GKT

The rationale behind the GKT is based on the theory and extensive research regarding orienting responses (ORs) and habituation processes in humans (e.g., Siddle, 1991; Sokolov, 1963, 1966). The OR is a complex of physiological and behavioral reactions evoked by any novel stimulus or by any change in stimulation (e.g., Berlyne, 1960; Sokolov, 1963). With repeated presentations of the stimuli, ORs undergo habituation, which is a gradual decline in response magnitude (Sokolov, 1963). In addition, stimuli that have a signal value for the subject (e.g., the subject’s own name) evoke enhanced ORs (Gati & Ben-Shakhar, 1990). Lykken (1974) was the first to note that this property of ORs endows them with the potential for disclosing guilty knowledge. He argued that: “. . . for the guilty subject only, the ‘correct’ alternative will have a special significance, an added ‘signal value’ which will tend to produce a stronger orienting reflex than that subject will show to other alternatives” (p. 728). The idea can be extended to other details that link the suspect to the culprit, not necessarily through the crime, but possibly through their biography (see the end of this paper for an example).

Because the psychophysiological differentiation in the GKT is mediated through a mechanism of orientation, the enhanced responsivity to relevant items need not be attributed to deception, motivation, or fear of punishment. Indeed, Lykken (1974) said of an individual possessing the guilty knowledge: “Whether he is high or low in reactivity, whether he has confidence in the test or not, whether he is frightened and aroused or calm and indifferent, we can still expect that his response to this significant alternative will be stronger than to the other alternatives as long as he recognizes which alternative is ‘correct’” (p. 728). Ben-Shakhar and Furedy (1990) call this a cognitive approach to psychophysiological detection, because it relies on what one knows, rather than on one’s emotions, concerns, and conditioned responses. Research demonstrates that relevant information can be detected even when no motivational instructions are given to the subjects, and even when no verbal response is required (e.g., Ben-Shakhar, 1977; Ben-Shakhar & Lieblich, 1982; Elass & Ben-Shakhar, 1989).

Although the scientific theory behind ORs is sound, for the purpose of a criminal investigation additional conditions are essential. First, the details which are selected as relevant items must be such that any individual at the scene of the crime is certain to have noticed them, and, moreover, remembered them. Otherwise, they will not possess the signal value necessary for eliciting enhanced orientation, and a guilty

suspect may “pass” the GKT.⁴ Second, the details chosen for the test cannot have been leaked to potential examinees. It makes innocent suspects privy to the guilty knowledge, endowing the leaked crime details with signal value even for a person who wasn’t there, and has no intimate acquaintance with the crime. Leakage can be advertent or inadvertent, and occur through the media or in the course of interaction with the police. The problem of leakage is the most severe one with which the GKT as an admissible evidence-gathering tool must contend, because it is the most vulnerable to deliberate abuse. We will return to this issue later.

Contamination in GKT Polygraph Investigations

Contamination refers to the danger that impressions the polygrapher forms from other information is “read into” polygraph charts that may not have otherwise indicated them (see, Ben-Shakhar, 1991; Elaad, Ginton & Ben-Shakhar, 1994). Contamination can affect the outcomes of polygraph tests either through the choice of control questions, their formulation and presentation to the examinees, or through a biased interpretation of the polygraph charts. The use of contaminated polygraph examinations as admissible evidence in court is highly problematic from a legal perspective, because, among other reasons, they may be contaminated by inadmissible information (Ben-Shakhar, Bar-Hillel, & Lieblich, 1986). Ben-Shakhar et al. (1986) list several guidelines, which can help decontaminate polygraph tests. These guidelines, impractical if not impossible to implement for the CQT, can be readily adopted for the GKT. For example, the GKT can be protected against contamination by administering it in a blind procedure (i.e., by an examiner who does not even know the relevant details).

Countermeasures and Abuse in GKT Polygraph Investigations

Polygraph investigations, especially those carried out by the police, occur in a highly charged and emotional context. The suspect—whether guilty or innocent—has a powerful motive to pass the test. Unfortunately, the police, too, may sometimes have a vested interest not just in discovering the truth, but in bringing about a particular polygraph outcome. When the stakes are this high, both interrogator and interrogatee may wish to manipulate the polygraph results. When the suspect tries to manipulate the results, we speak of countermeasures. When the investigator tries to manipulate the results, we speak of abuse.

Countermeasures

It is possible, indeed quite easy, to train guilty examinees to “pass” a polygraph examination (e.g., Ben-Shakhar & Dolev, 1996; Elaad & Ben-Shakhar, 1991; Honts, Devitt, Winbush, & Kircher, 1996; Honts, Raskin, & Kircher, 1987, 1994; Kubis, 1962). Simple behavioral techniques, called “countermeasures,” can be acquired with

⁴Which details in the scene of a crime would almost certainly be noted and remembered by the perpetrator is not necessarily obvious (e.g., Loftus & Loftus, 1976).

little effort, and can cause strong reactions to the “control” items (remember—stronger reactions to these items get you off the hook). Some countermeasures are physical, such as biting one’s tongue; others are mental, such as recalling an exciting or frightening event, or counting backwards in leaps of seven. The latter are more pernicious, because they are hard to detect even for the most experienced examiners.

A series of experiments by Honts et al. (1987, 1994, 1996) demonstrated that the rate of mistakes made by CQT and GKT polygraphists testing examinees practicing countermeasures ranged between 50 and 70%. Of course, the mistakes can only increase false-negative outcomes (i.e., the proportion of guilty suspects who are erroneously classified as innocents—the number of guilty suspects classified as innocents divided by the total number of guilty suspects), but not false-positives (the ratio of innocents classified as guilty to the total number of innocents). Insofar as these rates generalize from laboratory studies to real examinations, they obviously restrict the usefulness of the polygraph.

The standard physiological measures taken in polygraph interrogations, such as electrodermal activity (EDA) and respiration changes, require an interstimulus interval of about 20–25 s, because there must be a return to the baseline before a new measurement can start. Honts et al. (1996) demonstrated that such an interval suffices to allow the implementation of mental countermeasures. Fortunately, recent studies have demonstrated that certain measures derived from electroencephalogram (EEG); i.e., Event-Related Potentials (ERPs), can be used successfully in the GKT (e.g., Allen, Iacono, & Danielson, 1992; Boaz, Perry, Raney, Fischler, & Shuman, 1991; Farwell & Donchin, 1991; Rosenfeld, Cantwell, Nasman, Wojdac, Ivanov, & Mazzeiri, 1988). Specifically, the P300 component of the ERP, which represents cognitive activity occurring within 300–500 ms after stimulus onset, has been used in the GKT (e.g., Farwell & Donchin, 1991). Because the P300 enjoys such a small latency, it can be obtained with very short interstimulus intervals (e.g., 2 s), too short to enable the use of countermeasures. Although GKT studies that used the P300 have not examined the effects of countermeasures, this rapid presentation suggests that it would be difficult to employ them under this setup.

Abuse

It is somewhat dismaying to have to consider police abuse of an investigative tool, but to be realistic, we must. Abuse occurs when the investigator manages to somehow cue the innocent suspect to give a stronger response to a relevant item. It is particularly pernicious when the innocent suspect is unaware of having been subtly informed which are the relevant items.

The Validity of Polygraph Investigations

The bottom line in debating polygraph admissibility is, of course, whether it can do the job. Can the GKT in fact distinguish between suspects who possess guilty knowledge and those who do not? This is an empirical question, apparently answerable through research and experimentation.

Unlike the CQT, the accuracy of the GKT can be readily assessed, because it does not purport to detect guilt, anxiety, or deception—variables, which are hard

to introduce into controlled laboratory studies. Ben-Shakhar and Furedy (1990) reviewed and summarized ten GKT mock-crime experiments, and showed that across these studies, 84% of 248 guilty examinees and 94% of 208 innocent examinees were correctly classified. Eaad (1998) analyzed 15 GKT studies and found similar results (81% accuracy with guilty subjects and 96% with innocents). More recently, Ben-Shakhar and Eaad (in press-a) conducted a meta-analytic review of 80 laboratory studies, which included 169 conditions, representing various versions of the GKT. On the basis of a subset of 10 studies that best approximate realistic applications of the GKT (i.e., mock-crime studies using motivational instructions, deceptive verbal responses and at least 5 GKT sets) and that relied only on the electrodermal measure, they estimated the validity of the GKT to be 0.79. Although these GKTs did not investigate real crimes, the rationale of the GKT depends only on the signal value of the test items, and not on the existence of real concern and anxiety. Hence, the generalization from these studies to real GKTs is not too worrisome. The generalizability (i.e., the external validity) of GKT experiments to real-life interrogations is further strengthened by results which show that high levels of stress, while elevating all psychophysiological responses, have no effect on the differentiation between relevant and neutral stimuli (Kugelmass & Lieblich, 1966).

Nonetheless, evidence from laboratory studies must be supplemented by evidence from field studies, because real interrogations differ from simulated GKT experiments in important ways. (1) Simulated GKT experiments have used very simple tasks in which it was ascertained that all subjects learned all the relevant items, and memory for these details was not a concern, because subjects were typically tested immediately after being exposed to the guilty information. In real life, the offender is faced with a complex scene, and may not in fact notice, process, or store all details in memory. Moreover, because suspects are rarely tested immediately after committing the criminal act, and sometimes only months later, forgotten details may lose their signal value. (2) The main thing that could jeopardize a GKT test—leakage—was missing in the simulated studies. In real interrogations, critical items may be leaked to innocent suspects, raising false-positive errors—especially if informed innocent suspects are unable to explain how they became aware of the guilty information. See more on this issue below. (3) Though mock-crime experiments give subjects motivation to “pass,” suspects in a real crime are obviously more motivated to use countermeasures.

Only two GKT field studies, Eaad (1990) and Eaad, Ginton, and Jungman (1992), have been published so far. They reported rates of false-positive errors as low as those reported in simulated studies so that in those studies leakage of critical information (point 2 above) did not play a detrimental role. The rates of false-negative errors they reported were, however, larger than in simulated studies, perhaps because in real investigations, it is harder to ascertain that the relevant items were noted and remembered (point 1 above). However, the use of the GKT in the criminal cases studied by Eaad (1990) and Eaad et al. (1992) was not optimal. In particular, the number of questions used in these field studies was rather small (the mean number of questions was 2 and 1.8, respectively). In addition, these two studies were based on GKTs that were administered immediately after a CQT, which might have attenuated the sensitivity of the physiological measures due to habituation. Clearly, more field

studies are required, and in particular field studies attempting to apply the GKT under optimal conditions.

Should GKT Results be Admissible in Criminal Courts?

The current guidelines for admissibility of scientific evidence in US federal courts were recently set by the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993). As indicated above, the CQT does not satisfy the major *Daubert* criteria (Gallai, 1999; Saxe & Ben-Shakhar, 1999). In this section, we examine whether the GKT meets these criteria. Recall that the GKT is a valid technique by the definition and standards set by Messick (1995). First, it stands on sound theoretical grounds, because the relationships between stimulus significance and the physiological responses monitored in the GKT are well established. Second, the GKT relies on proper control questions, which guarantees that inferences made on the basis of its results can be defended (e.g., alternative explanations to an observed pattern of consistently larger reactions to the relevant than to the control items can be checked and ruled out). Third, its empirical (predictive) validity has been supported by a large body of research (e.g., Ben-Shakhar & Elaad, press-a).

Four considerations that judges should apply in determining whether to admit scientific expert testimony were articulated in *Daubert*. An examination of the GKT in light of these four considerations leads to the conclusion that it can satisfy all of them.

1. *Testability*. The GKT can be tested and has been tested again and again (see Ben-Shakhar & Elaad, in press-a, for a review). Although these tests were almost exclusively based on mock-crime studies, we argued that their results can be generalized to realistic situations. The high external validity of the GKT studies stems from the nature of the GKT as a test of guilty knowledge, rather than a test of deception. The phenomenon of enhanced ORs to significant stimuli (i.e., guilty knowledge items) does not depend upon stress, anxiety, or deception (see, Kugelmass & Lieblich, 1966), and there is no reason to believe that it will not be demonstrated in realistic investigations, if proper items (i.e., salient features of the event) are chosen as the critical stimuli of the test. Similarly, there is no reason why innocent suspects, who have no guilty knowledge, would show a differential response pattern to the critical items if sufficient precautions are taken against information leakage.

2. *Known error rates*. The available validity studies clearly indicate that the false-positive rate is very low (around 5%). Furthermore, several studies (Ben-Shakhar & Elaad, in press-b; Bradley & Ainsworth, 1984; Bradley & Rettinger, 1992; Bradley & Warfield, 1984; Iacono, Boisvenu, & Fleming, 1984), which used relatively large numbers of GKT questions (at least 9), and obtained no false positives, indicate that false-positive rates can be further reduced by increasing the number of GKT questions. It should also be noted that similarly low false-positive rates were obtained in the GKT field studies reported by Elaad (1990) and by Elaad et al. (1992), which means that leakage of relevant items did not play a role in the those criminal investigations.

The false-negative rates associated with the GKT are higher (16% according to the review of Ben-Shakhar & Furedy, 1990, and 19% according to the review

of Elaad, 1998), but these rates could also be reduced by increasing the number of GKT questions. The average rate of false-negative errors calculated on the basis of 5 mock-crime studies that used at least 9 GKT questions (Bradley & Ainsworth, 1984; Bradley & Rettinger, 1992; Bradley & Warfield, 1984; Iacono, Cerri, Patrick, & Fleming, 1992; Iacono et al., 1984) was 13%. The false-negative rates obtained in the field studies were also relatively large, but as indicated earlier, a relatively small number of questions were used.

The notion that the accuracy of the GKT can be enhanced by increasing the number of questions was clearly demonstrated by a meta-analysis of GKT studies (Ben-Shakhar & Elaad, in press-a), in which the number of questions used turned out to be the most effective factor moderating the accuracy of the GKT. Specifically, this analysis showed that GKTs based on at least five questions produced an average effect size (differentiation between the electrodermal response distribution of guilty and innocents) of 2.35, as compared with an average effect size of 1.29 for GKTs based on a smaller number of questions. These estimates are based on just a single physiological measure (the electrodermal measure) and the accuracy of the GKT can be further improved if additional measures, such as respiration changes and ERPs, are added. It is important to note that from a legal perspective, false-positive errors are much more threatening than false-negatives, and the fact that the GKT can be designed such that the likelihood of a false positive outcome is minimized should be taken into account when considering the admissibility of this test.

3. *Peer review and publication.* An inspection of the reference list in the recent meta-analytic review of GKT studies (Ben-Shakhar & Elaad, in press-a) reveals a very large number of studies published in peer-reviewed journals which examine the validity of the GKT as well as various factors that may affect the outcomes of this test. The two major peer reviewed journals where GKT studies are regularly published are the *Journal of Applied Psychology* and *Psychophysiology*, both of which are prestigious journals with high impact factors and high rejection rates.

4. *General acceptance.* Although the CQT has been the focus of a heated debate and has aroused major objections from many researchers and experts (e.g., Ben-Shakhar, 2002; Furedy, 1989; Furedy & Heslegrave, 1991; Iacono & Lykken, 1997, 1999; Kleinmuntz & Szucko, 1984; Lykken, 1998; Saxe, 1991, 1994; Saxe & Ben-Shakhar, 1999), the GKT has been accepted with no objections. The only reservation regarding the GKT is related to the issue of its applicability, and some researchers have argued that it can be applied in only a relatively small number of cases (e.g., Podlesny, 1993).

Although meeting the *Daubert* criteria, any recommendation to introduce the GKT as admissible evidence in criminal courts depends on taking proper measures to deal with the major threat associated with this test, that of information leakage, and especially deliberate and malicious leakage. If a suspect is aware of having acquired the “guilty knowledge” and can account for it, there is no problem. But if guilty knowledge has leaked without the suspect’s awareness, such a suspect, though innocent, might incriminate himself. Of course, police abuse is a problem which is not specific to the polygraph, and can distort any investigation. Insofar as one can tell from the GKT field studies conducted so far, either leakage was prevented, or, at least, it did not compromise the test—but clearly more field studies are necessary.

A recent study (Ben-Shakhar, Gronau, & Elaad, 1999) indicated, surprisingly, that the changes in respiration is a relatively accurate physiological measure even if some of the relevant information has been leaked to innocents. This study also demonstrated that certain techniques (e.g., introducing target items—items to which all subjects are required to respond by a key press—in addition to the usual items of a GKT) can reduce the risks of false-positive outcomes among informed but innocent subjects.

Prevention of leakage may require some administrative changes in police practices, but the Japanese experience suggests that it is possible (Fukumoto, 1980; Nakayama, 2002; Yamamura & Miyata, 1990). In order to follow the Japanese example and apply the GKT as a standard investigative tool in a large number of criminal investigations it will be necessary to modify police practices, such that critical features of an investigated event are identified and concealed at the outset of the investigation. Furthermore, GKTs should be conducted by investigators who are familiar with the scene of the crime and are trained to look for salient features that could be utilized as GKT questions. Admittedly, even if all these efforts are made, there will still be various criminal cases for which the GKT is not applicable. But we believe that the possibility of applying the GKT, even for a subset of criminal investigations, justifies these efforts.

In light of this analysis, we believe that the GKT, properly administered, could yield admissible evidence for criminal courts.

An Idealized GKT Investigation

We now describe an idealized GKT polygraph investigation, to highlight the possible drawbacks and complications that can arise when an actual GKT departs from this ideal.

Recall the bank robbery described in the second section. Ideally, when the first investigators arrive at the scene, they will make a list of possible items for a subsequent GKT, like those addressed in the example questions there (but also the brand of the chewing gum, the color of the wall-to-wall carpet in the saferoom, etc.). They can then set about to devise a GKT around those items even prior to apprehending any suspects, including in it as many items as possible. Appropriate distractors must be constructed for the correct items, and the completed test subjected to two kinds of checks—one with, and one without, physiological measures: (1) Asking someone ignorant of the details of the crime to guess the correct answers, to ascertain that the distractors were well chosen; (2) Measuring the physiological responses of an ignorant respondent when the options are read out, to ascertain that none of the answers are inherently exciting, for whatever unanticipated reason (for more detail see Lykken, 1998).

Once a suspect is being interrogated, the GKT should be administered sooner rather than later, and by a polygrapher who knows nothing about either the suspect or the crime. The suspect should be told which questions will be included, so that those he or she is willing to admit knowing the answer to can be discarded. Questions should be repeated several times, and the physiological measures should include at least the Skin Conductance Response (SCR) and Respiration Line Length (see Timm, 1982,

1987) that have been shown in many studies to produce optimal outcomes (e.g., Ben-Shakhar & Dolev, 1996; Ben-Shakhar & Elaad, in press-b; Ben-Shakhar et al., 1999; Elaad & Ben-Shakhar, 1997; Timm, 1982, 1987). The inclusion of ERP is also desirable because it can best deal with countermeasures, although this measure is presently not available in all polygraph laboratories. The entire interrogation, both before and during the polygraph test, should be videotaped. The chart should be scored mechanistically, preferably by computer, using algorithms.

From what we have said hitherto, the reasons underlying some of the above recommendations are apparent. Still, some are more essential than others. The most essential is that the polygrapher doing the test be ignorant of the case and of the suspect. Otherwise, the test is badly compromised by possible contamination, and might degenerate into mere clinical judgment. Similarly, computerized scoring is easy to implement and guards against biases, and so should be insisted upon.

TWO REAL CASES THAT COULD HAVE BENEFITED FROM A GKT

The polygraph's attraction for investigators has often been its potential ability to link a suspect to a crime when other evidence cannot satisfactorily do so. The GKT, as we emphasized throughout this paper, links a suspect to a crime through that suspect's cognitions, rather than through emotions, as in the CQT. But a link—or its absence—need not be to the crime in order to be probative in a criminal investigation.

When John Demjanyuk was convicted for Nazi war crimes in Jerusalem in 1988, his defense was that of mistaken identity. He was not, he claimed, Ivan the Terrible of Treblinka, in spite of some eyewitness identifications by some of that criminal's erstwhile victims. Materials uncovered after that verdict (ironically, by Demjanyuk's prosecutor), cast enough doubt on that identification to later cause a reversal of the guilty verdict by the Israeli Supreme Court. A creative use of the GKT, suggested by Lykken (1991), could have established very early whether or not Demjanyuk was Ivan the Terrible. Lykken noted that there are biographical details in each person's life which are probably remembered for life by that person, but are seldom known to strangers. For example, the name of the school which one attended, the name of a beloved pet, nanny or childhood friend, some life threatening childhood disease, etc. Presumably, some such items could have been recovered by investigators for Ivan the Terrible. An enhanced response to items such as the maiden name of the mother of Ivan the Terrible by John Demjanyuk would have linked the latter to the former's biography (though possibly through previous interrogations that Demjanyuk had been subjected to). But had Demjanyuk been unresponsive to Ivan the Terrible's biographical details, his claim of mistaken identification would have received some independent corroboration. Acquittal by the lower court would have saved Israel much time, money, and embarrassment.

In a second crime brought before Israeli courts, Moshe Azaria, a mildly retarded man of 21 years, was convicted of the murder of a young boy, based largely on his own confession, and a later reconstruction of the crime. Evidence uncovered much later suggested that he was beaten during his interrogation, and confessed to some crimes which did not even occur, and that he was directed into, rather than directing,

the later reconstruction. But shortly after the confession, and prior to a long day of investigation in which he apparently was fed the details of the crime and coached into accepting them, he was subjected to a GKT test. He was asked about the mode of the murder, and responded more strongly to the incorrect “a rag stuffed into the victim’s mouth,” than to the correct “a rope tied around the victim’s neck.” The GKT test results, as well as the polygrapher’s summary (“This does not indicate that the suspect committed the act”), were not brought to the attention of the court which convicted Azaria. One cannot but wonder how the verdict would have been affected had the court been told of this GKT. For a critique of the Azaria case see Sanjero and Kremnitzer (1999).

In an attempt to obtain a retrial subsequent to the discovery of this evidence, and other, that was withheld from the original court, Azaria’s lawyer argued for the acceptance of the GKT results not just because of the scientific reasons laid out in the present paper, but also because it could be used to exonerate, rather than convict, the suspect.

SUMMARY AND CONCLUSIONS

The purpose of this paper was to reconsider the admissibility of GKT polygraph test results as an aid in criminal courts. We attempted to show that when properly administered, the GKT meets the four major *Daubert* criteria: testability, known error rates, peer review and publication, and general acceptance. In addition, we argued that it meets the validity definition of Messick (1995) because it stands on solid theoretical grounds, it enjoys a considerable empirical validity, and inferences made on the basis of its results can be defended. We also discussed the major threats to the validity of the GKT, namely leakage of relevant information, as well as the application of countermeasures and abuse, and pointed out several means that should be taken to minimize the effects of these threats. Finally, we described the ideal way in which a GKT should be applied. Our recommendation following this discussion was that if properly administered, the GKT could be used as admissible evidence in criminal courts. Even if this usage may be limited in its scope (because in many situations it is not easy to identify many salient features of the crime that can be used as proper GKT questions), it will be a significant aid to the trier of fact.

ACKNOWLEDGMENT

We thank Eitan Elaad, and several anonymous reviewers for commenting on an earlier version.

REFERENCES

- Allen, J. J., Iacono, W. G., & Danielson, K. D. (1992). The development and validation of an event-related-potential memory assessment procedure: A methodology for prediction in the face of individual differences. *Psychophysiology*, 29, 504–522.

- Ben-Shakhar, G. (1977). A further study of dichotomization theory in detection of information. *Psychophysiology*, *14*, 408–413.
- Ben-Shakhar, G. (1991). Clinical judgment and decision making in CQT polygraphy: A comparison with other pseudoscientific applications in psychology. *Integrative Physiological and Behavioral Science*, *26*, 232–240.
- 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: Academic Press.
- Ben-Shakhar, G., Bar-Hillel, M., & Lieblich, I. (1986). Trial by polygraph: Scientific and juridical issues in lie detection. *Behavioral Sciences and the Law*, *4*, 459–479.
- Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge test: The effects of mental countermeasures. *Journal of Applied Psychology*, *81*, 273–281.
- Ben-Shakhar, G., & Elaad, E. (in press-a). The validity of psychophysiological detection of deception with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*.
- Ben-Shakhar, G., & Elaad, E. (in press-b). Effects of questions' repetition on the efficiency of the guilty knowledge test: A reexamination. *Journal of Applied Psychology*.
- 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., Gronau, N., & Elaad, E. (1999). Leakage of relevant information to innocent examinees in the GKT: An Attempt to reduce false-positive outcomes by introducing target stimuli. *Journal of Applied Psychology*, *84*, 651–660.
- Ben-Shakhar, G., & Lieblich, I. (1982). The dichotimization theory for differential autonomic responsivity reconsidered. *Psychophysiology*, *19*, 277–281.
- Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. New York, NY: McGraw-Hill.
- Boaz, E., Perry, N. W., Raney, G., Fischler, I. S., & Shuman, D. (1991). Detection of guilty knowledge with event-related potentials. *Journal of Applied Psychology*, *76*, 788–795.
- Bradley, M. T., & Ainsworth, D. (1984). Alcohol and the psychophysiological detection of deception. *Psychophysiology*, *21*, 63–71.
- 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.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. Commonwealth of Massachusetts v. Woodward (D. Boston, 1998), SJC-07635. Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. Supp. 2786 (1993).
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigation. *Journal of Applied Psychology*, *75*, 521–529.
- 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 level and verbal response type on psychophysiological detection in the guilty knowledge test. *Psychophysiology*, *26*, 442–451.
- Elaad, E., & Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *International Journal of Psychophysiology*, *11*, 99–108.
- Elaad, E., & Ben-Shakhar, G. (1997). Effects of items' repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology*, *34*, 587–596.
- Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, *7*, 279–292.
- 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 potentials. *Psychophysiology*, *28*, 531–547.
- Fukumoto, J. (1980). A case in which the polygraph was the sole evidence for conviction. *Polygraph*, *9*, 42–44.
- Furedy, J. J. (1989). The North American CQT polygraph and the legal profession: A case of Canadian credulity and a case for cultural concern. *The Criminal Law Review*, *31*, 431–445.
- 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, & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (Vol. 4). London: Jessica Kingsley Publishers Ltd.
- Gallai, D. (1999). Polygraph evidence in federal courts: Should it be admissible? *American Criminal Law Review*, *36*, 87–116.
- 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.

- Harnon, E. (1982). Evidence obtained by polygraph: An Israeli perspective. *The Criminal Law Review*, 329–348.
- 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., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*, 1, 241–247.
- Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252–259.
- Iacono, W. G., Boisvenu, G. A., & Fleming J. A. (1984). Effects of Diazepam and Methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology*, 69, 289–299.
- Iacono, W. G., Cerri, A. M., Patrick, C. J., & Fleming J. A. (1992). Use of anti-anxiety drugs as countermeasures in the detection of guilty knowledge. *Journal of Applied Psychology*, 77, 60–64.
- Iacono, W. G., & Lykken, D. T. (1997). 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* (pp. 582–618). St. Paul, MN: West Publishing.
- Iacono, W. G., & Lykken, D. T. (1999). Update: 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* (pp. 174–184). St. Paul, MN: West Publishing.
- Kleinmuntz, B., & Szucko, J. J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist*, 39, 766–776.
- Kubis, J. F. (1962). *Studies in lie detection: Computer feasibility considerations*. Technical Report #62-205, prepared for the Air Force Systems Command. Contract No. AF 30 (602)-2270, project No. 5534, Fordham University.
- Kugelmass, S., & Lieblch, I. (1966). Effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, 50, 211–216.
- Loftus, G. R., & Loftus, E. F. (1976). *Human memory: The processing of information*. Hillsdale, NJ: Erlbaum.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385–388.
- Lykken, D. T. (1974). Psychology and the lie detection industry. *American Psychologist*, 29, 725–739.
- Lykken, D. T. (1991). Why (some) Americans believe in the lie-detector while others believe in the guilty knowledge test? *Integrative Physiological and Behavioral Science*, 26, 214–222.
- Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York, NY: Plenum.
- Marston, W. M. (1917). Systolic blood pressure changes in deception. *Journal of Experimental Psychology*, 2, 143–163.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 49–86). New York: Academic Press.
- Podlesny, J. A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest*, 20, 57–61.
- Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin*, 84, 782–799.
- Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence*. New York: Springer-Verlag.
- Raskin, D. C., Honts, C. R., Amato, S. L., & Kircher, J. C. (1999). Update: The scientific status of research on polygraph techniques: The case for the admissibility of the results of polygraph examinations. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Vol. 1, pp. 160–174). St. Paul, MN: West Publishing.
- Raskin, D. C., Honts, C. R., & Kircher, J. C. (1997). The scientific status of research on polygraph techniques: The case for 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*. St. Paul, MN: West Publishing.
- Reid, J. E. & Inbau, F. E. (1977). *Truth and deception: The polygraph ("Lie Detection") technique*. Baltimore: Williams and Wilkins.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeiri, L. (1988). A modified event-related potential-based guilty-knowledge test. *International Journal of Neuroscience*, 42, 157–161.

- Sanjero, B., & Kremnitzer, M. (1999). Retrial: Reality or dream? On how justice loses to the principle of judicial finality (case 6148/95 Moshe Azaria vs. The State of Israel, ruling 54(2) 334). *The Ramat Gan Law Review (in Hebrew)*, 1(1).
- Saxe, L. (1991). Science and the CQT polygraph: A theoretical critique. *Integrative Physiological and Behavioral Science*, 26, 223–231.
- Saxe, L. (1994). Detection of deception. *Current Directions in Psychological Science*, 3, 69–73.
- Saxe L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. *Psychology, Public Policy and Law*, 5, 203–223.
- Siddle, D. A. T. (1991). Orienting, habituation, and resource allocation: An associative analysis. *Psychophysiology*, 28, 245–259.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York, NY: Macmillan.
- Sokolov, E. N. (1966). Orienting reflex as information regulator. In A. Leontyev, A. Luria, & A. Smirnov (Eds.), *Psychological research in the U.S.S.R.* (pp. 334–360). Moscow: Progress Publishers.
- State of Israel. (1981). *A report regarding the uses of the polygraph*. Jerusalem: Ministry of Justice (in Hebrew).
- Timm, H. W. (1982). Effects of altered outcome experiences stemming from placebo and feedback treatments on the validity of the guilty knowledge test. *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.
- United States v. Cordoba, 158 (D. California, 1998), aff'd, SA CR 95-39-GLT[SF].
- United States v. Scheffer, 118 S. Ct. Supp. 1261 (D. Washington, 1998), aff'd, USCA Dkt. No. 95-0521/AF (US Court of Appeals for the Armed Forces).
- Yamamura, T., & Miyata, Y. (1990). Development of the polygraph technique in Japan for detection of deception. *Forensic Science International*, 44, 257–271.