A Critical Review of the Control Questions Test (CQT)

Gershon Ben-Shakhar

The Hebrew University of Jerusalem

I would like to thank Professor Leonard Saxe for his helpful and constructive comments

Introduction

The Control Questions Technique (CQT) is the most widely used method of psychophysiological detection, and it is being applied for various purposes (e.g., criminal investigations, employee screening and selection) in several countries, most notably in North America and Israel. Although the scientific basis and validity of the CQT have been debated in the scientific literature (e.g., Ben-Shakhar & Furedy, 1990; Iacono & Lykken, 1997, 1999; Lykken, 1974, 1998; Raskin, 1986, 1989; Raskin, Honts, Amato, & Kircher, 1999; Raskin, Honts, & Kircher, 1997), its extensive use continues. The purpose of this chapter is to present a critical analysis of the CQT, which demonstrates that this method does not have a sound scientific foundation, and that decisions made on the basis of its results may suffer from several flaws and biases. This does not necessarily mean that the CQT cannot serve as an aid to law-enforcement agencies, which use other non-scientific investigative methods, but it definitely means that it should not be used as admissible evidence in legal proceedings.

The chapter opens with a very brief description of the CQT. The next sections focus on the major problematic aspects of the CQT: (1) Inadequate theoretical and logical rationale; (2) Inadequate standardization; (3) Lack of objective quantification procedures of the physiological responses; (4) The implications of contamination with other sources of information; (5) The problem of countermeasures. The empirical questions regarding the reliability and validity of the CQT will be discussed in the next section, and it will be argued that the available research is insufficient for estimating the accuracy of CQT outcomes that would be obtained in realistic criminal investigations. Finally, conclusions will be drawn regarding the various applications of the CQT.

A brief description of the CQT

Detailed descriptions of the CQT can be found in various sources (e.g., Raskin, 1989; Reid & Inbau, 1977; Saxe, Dougherty & Cross, 1983). For the purposes of this chapter a brief description of the CQT, as applied in forensic settings, is adopted from Elaad, Ginton, & Ben-Shakhar (1994). But before describing the CQT, I would like to note that this chapter focuses on the CQT as typically used by law-enforcement agencies, and therefore various variations on this technique will not be discussed.

The CQT is administered in several stages: First, the examiner becomes familiar with the facts of the case by reading the written report and by speaking directly to the police investigator who ordered the examination. Typically, relevant background information, such as the suspect's past criminal record, is made available to the examiner. During the next stage the examiner conducts a pre-test interview, in which the examinee is given the opportunity to talk about the offense and present his or her version of the case. The series of questions, to be asked later in the actual examination stage of the polygraph test, is formulated during this pre-test interview through an interaction between the examiner and the examinee. The examiner discusses the formulation of the questions with the examinee and ensures that he or she understands them and can give a direct "yes" or "no" answer to each question. The examiner explains the testing procedure and informs the examinee that the examination is voluntary. The next stage is the actual examination stage during which the examinee is attached to the polygraph. Typically, the examiner will attempt to convince the examinee that the polygraph is highly accurate. For this purpose a rigged card-test is usually administered either before conducting the CQT, or during an

intermission between CQT sessions (for a more detailed of this procedure, see Ben-Shakhar & Furedy, 1990; Saxe, 1991).

During the examination stage a series of questions is presented to the examinee while continuously measuring several physiological reactions. Typically, at least three physiological indices are used: Changes in respiration, obtained from a tube attached around the thorax and abdomen, changes in electrodermal activity, obtained from two electrodes placed on the palmer surface of two fingers, and changes in relative blood pressure, obtained from a partially inflated cuff placed around the upper arm. The questions are of three general types: (a) Relevant questions - directly crime-related questions of the "Did you do it?" type (e.g., "Did you break into Mr. Jones's apartment last Friday night?"). (b) Control questions - focusing on general, non-specific misconducts, of a nature as similar as possible to the issue under investigation (e.g., "Have you ever taken something that did not belong to you?"). (c) Irrelevant questions - focusing on completely neutral issues (e.g., are you sitting on a chair?). These are intended to absorb the initial orienting response evoked by any opening question, and to enable rest periods between the more loaded questions. Typically, the whole question series is repeated three or four times.

The inference rule used to derive the CQT's outcomes - the rule underlying inferences from a given pattern of physiological responding to a conclusion of guilt versus innocence, or truth versus deception - is based on a comparison of the responses evoked by the relevant and the control questions. Deceptive individuals are expected to show more pronounced responses to the relevant questions, whereas truthful individuals are expected to show the opposite pattern of responsivity (i.e., more pronounced responses to the control questions). Thus, a pattern of consistently larger responses to the relevant than to the control questions is taken as an indication

of deception ("deception indicated"); whereas a consistent pattern of larger responses to the control questions will lead to a conclusion that the examinee is answering truthfully ("no deception indicated"). An inconsistent pattern of responding or a pattern of no differences in the responses to the two types of questions will lead to an inconclusive decision.

The case against the CQT: Five major CQT features that make this method unscientific

As indicated earlier, the rationale and inference rule of the CQT have been debated in the scientific literature since the Seventies (e.g., Ben-Shakhar & Furedy, 1990; Furedy, & Heslegrave, 1991; Iacono, & Lykken, 1997, 1999; Kleinmuntz, & Szucko, 1984; Lykken, 1974, 1978, 1979, 1998; Podlesny, & Raskin, 1977, 1978; Raskin, 1982, 1986, 1989; Raskin, et al., 1997, 1999; Raskin, & Podlesny, 1979; Saxe, 1991; Saxe et al., 1985). The purpose of this section is not to reiterate this debate, but rather to discuss several features of the CQT, which highlight why this technique does not have scientific validity. As indicated earlier, this does not mean that the CQT is useless, as many unscientific methods are routinely used in criminal investigations. However, it is important to make a clear distinction between forensic techniques that have sound scientific foundations and those that are based on impressions and intuitions, but lack scientific validity. This distinction is important primarily when the legal status of CQT results is at stake, but the limitations of this method should be clear and explicit for all potential users, as well as the general public.

<u>Weaknesses of the theoretical foundation and the logical rationale of the CQT</u>

An essential requirement of any technique, derived from scientific principles is that it would be linked to a theory that can be tested and validated. Indeed, modern views of the concept of validity (e.g., Messick, 1995) include the theoretical foundation as an essential element. To use Messick's terminology, validity is an overall evaluative judgement of the degree to which empirical evidence and theory support the adequacy and appropriateness of the interpretations and actions based on test scores. In other words, in the process of validating a test or a method, it is not sufficient to demonstrate that its results correlate with a relevant criterion (i.e., predictive, or criterion validity). A theoretical foundation is particularly important for validating the CQT because we need to generalize from experimental situations to realistic settings and this is virtually impossible without a theory.

Recently, Saxe and Ben-Shakhar (1999) analyzed the CQT, in light of Messick's (1989, 1995) approach and showed that it cannot be regarded as a valid test of deception. Specifically, there is no theory that can establish the relationships between physiological changes and deception. Furthermore, there is a general consensus, even among CQT proponents, that there is no specific "lie response". For example, Raskin (1986, p. 31) wrote: "No known physiological response or pattern of responses is unique to deception". Thus, Saxe and Ben-Shakhar (1999) argued that the two major sources of invalidity noted by Messick (1995) affect CQT polygraph testing. First, the construct of deception is underrepresented by the CQT results, because there is neither a theoretical rationale, nor empirical evidence to support the relationship between the physiological measures monitored during the CQT examination and deceptive behavior. Second, CQT results may reflect other constructs, such as surprise, anxiety, and stress. Consequently, it suffers from the second major threat to validity suggested by Messick, namely "construct-irrelevant

variance" - the assessment is too broad, and contains excess reliable variance associated with other distinct constructs.

But even if we abandoned the idea of psychophysiological detection of deception and restricted ourselves to more modest goals, such as detection of involvement with a criminal event, the CQT would still need to have a theoretical basis, or at least a logical rationale. Indeed, many CQT proponents (e.g., Raskin, 1986) have abandoned the notion of “a specific lie response”, but argued that inferences about truth or deception could be made by comparing the relative strength of the subject’s responses to relevant and control questions. However, no convincing rationale for such inferences has been given so far.

The major problem stems from the nature of the control questions used in the CQT. The phrase “Control Questions” gives the impression that true controls are being exercised. From a logical perspective, true controls require a perfect match between all factors other than the factor being tested (in this case, the factor of deception or involvement in a crime). Hence, the control questions ought to be just like the relevant questions in all details, though only the relevant questions should tie the suspect to the crime. In other words, from the perspective of an innocent suspect there ought to be no differences whatsoever between the two types of questions. Whereas this kind of control is exercised in another psychophysiological detection method - the Guilty Knowledge Technique (GKT) - it does not exist in the CQT. As explained earlier, in the CQT, the relevant questions are questions that relate directly to the specific event being investigated, while the control questions relate to general, non-specific crimes. Thus, in the CQT both relevant and control questions are relevant to all examinees, but to different and unknown degrees. Clearly, one cannot assume that these questions are equivalent. Because of these obvious differences, it has been

argued that the term "control" is misleading in this context (see, e.g., Ben-Shakhar & Furedy, 1990; Furedy & Heslegrave, 1991). The differences between the two types of questions are obvious, and even an innocent suspect can easily distinguish between a question that relates directly to the focal event around which the investigation revolves (the event that provided the impetus for conducting an investigation), and general questions related to hypothetical events from the examinee's past. Moreover, once an examinee is aware of the CQT rationale and inference rule, it becomes obvious to him or her that only the relevant questions pose a real threat. It should be stressed that the problematic nature of the control questions is not merely semantic, and calling these questions "comparison questions", as has been recently proposed by many CQT proponents, would not make the CQT's inference rule more reasonable and would not provide better protection against false-positive outcomes.

Supporters of the CQT claim that a skilled interrogator is capable of choosing control questions while interviewing the suspect, and of creating an atmosphere that leads innocent examinees to be more concerned with the control than the relevant questions, while guilty suspects become more concerned with the relevant than the control questions. The following citation from Honts and Perry (1992, p. 360) demonstrates how the logical basis of the CQT is explicated by its proponents: "It is assumed that all subjects will be concerned about the veracity of their denial to the control questions. Innocent individuals are expected to produce larger physiological responses to control questions than to relevant questions since they are sure of the veracity of their response to the relevant questions, but they are assumed to be either lying or at least uncertain about the veracity of their response to the control questions."

Unfortunately, other than the polygraphists' strong belief in this assumption, it has no grounding in psychological or psychophysiological research, nor is it convincing in its inner logic. Honts and Perry's reasoning rests on the assumption that "belief in the veracity of their answers to the relevant questions" is sufficient to guarantee that innocent suspects will show larger physiological responses to the control questions. This might have made sense, if such beliefs were the only factor determining physiological reactions. But as indicated earlier, many other factors affect physiological responsivity. Particularly, fear of being falsely classified as guilty and bearing the consequences of such an error is one salient factor that may cause strong reactions to the relevant, crime-related, questions among innocent suspects, even if they believe in the veracity of their answers to the relevant questions.

Indeed, numerous researchers (e.g., Ben-Shakhar & Furedy, 1990; Lykken, 1974, 1998) have expressed a concern that this technique is biased against the innocent suspect, because relevant questions could be readily perceived as more threatening and agitating than control questions to all examinees. Moreover, the rationale of the CQT creates an advantage for dishonest examinees. According to this rationale a control question elicits a strong reaction inasmuch as an examinee answers it falsely, or at the very least is unsure of the truthfulness of his answer. If we take this rationale at face value, the higher the chances that the examinee's answer is false, the higher his reactions to the control questions should be. According to this logic, a dishonest examinee that frequently steals, lies, or injures his fellow men, will show stronger responses to a control question regarding the aforementioned activities than a virtuous person. Therefore, paradoxically, the chance that a CQT test will incriminate the honest examinee (who doesn't tend to react to the control questions) is greater

than the chance that a dishonest examinee will be incriminated. Indeed, the very logic of the CQT points out the danger that the honest man's version will be judged untrue.

Recently, a new variation of the CQT, called the Directed Lie Test (DLT), has been endorsed by Raskin and his colleagues (e.g., Horowitz, Kircher, Honts & Raskin, 1997; Raskin et al., 1997). In the DLT, the traditional control questions are replaced with directed lies (i.e., questions that both the examiner and the examinee agree are deliberate lies). However, even these control questions are clearly distinct from the relevant questions and therefore cannot function as true controls. Thus, the weaknesses of the CQT are not going to be resolved by the DLT. Furthermore, the DLT has been disputed even among CQT supporters. For example, Abrams (1999) recently argued that the excessive emphasis placed on the directed lies could lead to false-negative outcomes. A more general criticism of the DLT was made by Iacono and Lykken (1999) who argued that the directed lie questions may create an additional difficulty because explaining the purpose of these questions to the examinees clarifies the importance of giving strong responses to these questions, thereby making the test even easier to beat than the CQT.

Inadequate standardization

Many regard the polygraph examination as a test. However this expression (like the expression "control questions") is misleading when applied to the CQT. A review of the scientific literature dealing with psychological testing reveals that a basic requirement of a "test" is standardization. This requirement is essential to guarantee that all examinees undergo the same experience. Only when it is fulfilled do the resulting scores (or evaluations) have a uniform meaning, allowing comparisons between different people who took the test. The CQT procedure is poorly standardized. The pretest interview with the examinee, conducted by the interrogator,

is an essential part of every CQT examination. This part is completely subjective, but the control questions, which later form the basis of the polygraphist's inferences, are determined in its course. The selection of control questions depends solely on the interrogator's intuition, and the relationship that forms between him (or her) and the examinee.

In addition, the testing conditions may also be a function of the examiner and the relationship he or she formed with the examinee. For example, an examiner may present the questions in a different manner when he believes that the examinee is deceptive than when he believes he is testing a truthful suspect. This feature of the CQT has been acknowledged by several supporters of this technique. For example, Honts and Perry (1992, p. 372) wrote that, "…an examiner who was motivated to produce a deceptive result might ask overgeneral or provocative relevant questions, and spend a great deal of time on their review and presentation. Subsequently, this unethical examiner could ask very narrow, specific, or inappropriate control questions and spend very little time on their review and presentation. An examiner predisposed to produce a truthful result could take the opposite approach, overemphasizing the control questions and minimizing the relevant questions". Honts and Perry (1992) raised this possibility in relation to an unethical and dishonest examiner, but decades of research in social psychology teaches us that honest persons could be unintentionally affected by their prior beliefs (e.g., Chapman & Chapman, 1982; Klayman & Ha, 1987; Snyder & Swann, 1978a,b). More recently, Abrams (1999, p. 224) made a similar comment and wrote that, "…there is a delicate balance that exists between the comparison and relevant questions and many variables can tip this balance in either of those two directions. Too much discussion of one or the other during the pretest, a difference in inflection or loudness when the questions are being

asked, any discussion between charts that stresses either the relevant or comparison questions, or any mental activity on one question versus another can weigh the balance in the direction of that particular emphasis". These citations clarify the implications of the unstandardized nature of the CQT. It is therefore clear that, by and large, polygraph examinations conducted by different interrogators (even for a given case and suspect) are liable to be quite different from each other.

In contrast, consider tests of cognitive abilities, such as intelligence tests. Despite all the criticism leveled at these tests, it would be inconceivable for each examinee to be tested on questions created especially for him, and for each tester to construct the questions for a particular examinee based on the best of his intuitive judgment.

Because the CQT is poorly standardized, one can regard the conclusions of an interrogator who employs this technique as more or less on a par with the opinions of an interviewer, rather than objective results of a scientifically-based technique.

Lack of objective quantification of the physiological responses

Lack of standardization characterizes not only the choice and presentation of the CQT questions, but also the measurement and quantification of the physiological responses. This is rather surprising because the type of physiological responses monitored during a typical CQT test can be easily measured in an objective manner, using computerized procedures. Such an objective quantification is a routine procedure in psychophysiological experiments, and computer algorithms have been developed for measuring the responses in the CQT (e.g., Kircher & Raskin, 1988).

However, objective, quantified measurement procedures are rare in CQT practice. Some polygraph agencies rely on an overall evaluation of the polygraph charts. This approach is clearly impressionistic and subjective, and as such, vulnerable

to various judgement biases (e.g., the confirmation bias, which will be discussed in the next section). Others, use the semi-objective procedure proposed originally by Backster (1963). According to this procedure, two or three pairs of Relevant-Control questions are identified in each polygraph chart, and numbers (-3, -2, -1, 0, 1, 2, 3) are assigned to each pair for each physiological measure. The absolute value of the assigned number reflects the magnitude of the difference between the responses evoked by the two questions within the pair (e.g., -3 or +3 reflect a very large difference, -1 or +1 reflect a small difference and 0 reflects no difference) and the sign of the assigned number reflects the direction of the difference, in a way that positive numbers are associated with a pattern of larger physiological reactivity to the control question and negative numbers reflect the opposite pattern. These numbers are then summed up across question pairs, across physiological measures and across polygraph charts to yield a global score. Thus, if for example a polygraph examination is based on three charts and three physiological measures and if two pairs of Relevant-Control questions are identified for each chart, then the global score will range between -54 and +54. Typically, the following classification rule is used: If the global score exceeds +5, an NDI ("no deception indicated") classification is reached; if the global score is less than -5, the polygraph record is classified as DI ("deception indicated"); and if the global score ranges between -5 and +5, the record is classified as inconclusive.

While the Backster (1963) procedure is certainly an improvement over the overall evaluation approach, it is still subjective because it is often unclear whether a given pair of responses reflects a large, medium, or small difference between the responses to relevant and control questions. Thus, this approach too, may be vulnerable to judgement biases.

The problem of contamination

Polygraphists who employ the CQT believe that it is vital that the same interrogator construct the questions and conduct the questioning. Often that interrogator also tallies up the results of the examination. This approach introduces contamination into the investigative process: Judgments made on the basis of the CQT are based on more information than is contained in the physiological measures alone, such as the examinee's criminal records and the information contained therein. In addition, during the investigation the polygraphist is in the position to watch and monitor the totality of a suspect's behavior, and not just his physiological changes. An experienced interrogator might well use these characteristics. While such rich information could enhance the accuracy of the polygraphist's final judgment, it also contaminates "objective" evidence with mere impressions.

The distinction between an objective lie detector based on physiological responses and the subjective impressions of human investigators must not be blurred. If we confuse the validity of the polygraphist's judgment with the validity of the polygraph, we are liable to overestimate the validity of the machine. Furthermore, contamination may introduce a bias into the polygraph examiner's final judgement. This bias was labeled "confirmation bias" (e.g., Ben-Shakhar, 1991a; Elaad et al., 1994) because the knowledge gathered prior to the polygraph investigation may induce certain expectations in the examiner. The polygraph investigation and chart interpretation may be biased in favor of these prior expectations.

Darley and Gross (1983) made a distinction between two types of confirmation bias, both of which may play a role in polygraph examiners' judgements: (a) The 'cognitive confirmation effect', which occurs in the absence of any interaction between the perceiver (in this case the polygraph examiner) and the

target person (the suspect). This effect is relevant for the chart interpretation phase of polygraph investigation, and its impact depends on the subjectivity of the chart interpretation process. (b) The 'behavioral confirmation effect', which refers to effect of expectations on the behavior of the examiner towards the suspect (e.g., the manner in which the relevant and control questions are presented to the suspect).

It should be noted that the "behavioral confirmation effect" is not eliminated even if the polygraph charts are evaluated by polygraphists who are "blind" to the case and the details of the investigation. Contamination could still enter the measurements through what psychologists call "interpersonal expectations" effect (Rosental and Rubin, 1978). This effect was first discovered in the context of psychological experiments, and is related to the experimenters often arousing in their subjects patterns of behavior that supports their expectations -- a sort of self-fulfilling prophecy. These effects are also known in medicine, for instance in the evaluation of the efficacy of new drugs. In the context of polygraphy, the effect occurs when an investigator, in the light of a previous examination for example, develops a certain hunch about the suspect's guilt or innocence, and can subconsciously influence the measurements to match this belief. It should be noted that the physiological measures recorded by the polygraph are very sensitive to changes in the volume of the interrogator's voice, emphases, speech pauses, and the like.

An interesting and impressive demonstration of the contamination effect and the type of bias to which it could lead, was presented in a television program produced by CBS in America. It was first shown in 1986,within the "60 Minutes" program, as an informal experiment conducted by its producers. As part of the experiment, they independently approached three polygraphists, with a request to conduct an investigation for a firm from which some photographic equipment had

allegedly been stolen. The polygraphists were told that only four employees had access to the equipment, and therefore only one of them could have stolen the equipment. They were also told which employee was suspected of being the thief, but that there was no evidence to support this suspicion. In truth, no equipment had been stolen, but each polygraph interrogator was given a different name as the name of the suspected thief (unbeknownst to the employees themselves). Each of the polygraph interrogators examined the four employees using CQT procedures, and each of the three investigators reached the confident conclusion that the employee that had been named to him had lied during the polygraph investigation, while the three other employees had spoken the truth. This demonstration gives a very vivid illustration of the confirmation bias that may result from contaminated CQT examinations.

Ethical as well as practical problems make it difficult to conduct controlled experiments on the effect of contamination on polygraph examiners' judgement. This is why only two studies that examined this issue have been published so far. Their results, however, are inconsistent. Elaad et al. (1994) showed that the cognitive confirmation effect does play a role in CQT chart interpretation, even when the semi-objective method proposed by Backster (1963) was adopted. They demonstrated that prior expectations of polygraph examiners affected the way they interpreted CQT polygraph charts, when in reality these charts were inconclusive. No significant effect was obtained in this study for conclusive charts. In a subsequent study Elaad, Ginton and Ben-Shakhar (1998) manipulated expectations and examined the effect of these expectations on the entire CQT examination. No effect was obtained in this study, but it is not entirely clear whether the manipulation was effective.

Ben-Shakhar, Bar-Hillel and Lieblich (1986) discussed the implications of contaminated CQT examinations for the legal usage of CQT results. They argued that

the problem with the judgments of CQT examiners rests not only in the subjectivity of their assessments, since subjectivity characterizes other categories of expert testimony as well. Rather, the laws of evidence sometimes deliberately constrain the evidence that can be presented in court. These constraints do not hold for polygraphists, who may have access to all the information that reaches the police. The ethical and legal regulations that guide and limit a polygraph examination are far less rigorous than those that guide court proceedings and there is no guarantee that polygraphists adhere even to these limited regulations. Unlike court actions, polygraph investigations are typically not conducted in the presence of a lawyer, and often provide only limited legal protection for the suspects. Therefore, the presentation of CQT results in court deprives the suspect of many of the legal protections provided to him in regular court proceedings. This is an opening for a type of "laundering" of inadmissible evidence, or evidence obtained through illegal means, without legal checks. Such evidence could accrue unknown weight through the influence it exerts on the opinions of the polygraphist, who has prior acquaintance with it.

This problem is especially acute when results obtained from CQT tests are presented as objective and scientific, when in fact the CQT is just a tool aiding the investigator in collecting impressions. The way a "lie detector" works carries an aura of mystery, but the true mystery is not in the function of the polygraph machine, but in the function of the mind of the human polygraphist behind the machine. Once we acknowledge this fact, we have to deal with the question of whether there is sufficient reason to believe that the polygraphist as a human lie-detector is superior and should be preferred over other people in general, and judges in particular.

The problem of countermeasures

A number of experiments (e.g., Ben-Shakhar & Dolev, 1996; Elaad & Ben-Shakhar, 1991; Honts, Devitt, Winbush, & Kircher, 1996; Honts, Raskin, & Kircher, 1987, 1994; Kubis, 1962) have indicated that it is possible, indeed quite easy, to train guilty examinees and prepare them for a polygraph examination (either CQT, or GKT) in such a way that with a high probability they will be found truthful. This can be done by adopting some rather simple techniques, which can be picked up with little effort and can cause very strong reactions to the control questions. These techniques rely either on the use of physical means (such as biting one's tongue), or mental means (calling to mind an exciting or frightening event, or engaging in mental activities that require effort) each time a control question is asked. A series of experiments conducted by Honts and his colleagues demonstrated that the use of such counter measures could be most effective. They showed in different experiments that the rate of mistakes made by polygraphists testing "guilty" examinees who were using countermeasures ranged between 50 and 70 percent.

These countermeasures may increase false-negative outcomes (guilty suspects classified as "innocents"), but they have no effect on innocent examinees. Thus, from the legal perspective this problem may be less severe than the problematic nature of the control questions used in the CQT, because the legal system puts more weight on protecting the innocent than the guilty suspects. On the other hand, it is clear that if the results of the laboratory studies on the effects of countermeasures on false-negative outcomes generalize to realistic polygraph examinations, they will have a restricted utility. It should be pointed out that the type of countermeasures that are most detrimental for all psychophysiological-detection techniques are mental countermeasures, because mental manipulations cannot be detected even by the most experienced examiners. Two recent studies demonstrated that mental countermeasures

can be used effectively under both the GKT and the CQT (Ben-Shakhar & Dolev, 1996; Honts et al., 1996).

The Empirical Status of the CQT

The primary question that arises in all debates about the CQT is of course the question of criterion validity. That is to say, to what degree can one really distinguish between liars and truth tellers (or between examinees who are informed regarding a certain event to those who are not), based on their physiological responses to questions presented to them during a CQT examination. This may seem to be an empirical question that could be checked through research and experimentation; and if the results of such experimentation were to indicate a high degree of accuracy, then many of the previously raised concerns would become less important. In actuality, matters are more complicated, and it turns out that the research conducted to date cannot provide a simple, clear-cut answer to the question of the CQT's validity. In order to allow conclusions about the value of the CQT, as typically conducted in real-life conditions, an experiment should fulfill the following requirements (see, Ben-Shakhar & Furedy, 1990; Ginton, Dai, Elaad, & Ben-Shakhar, 1982):

1. The existence of a clear, conclusive and irrefutable criterion for the guilt or innocence of the research participants. Clearly, without such a criterion there is no way to determine whether the CQT interrogator was right or wrong in a particular case.
2. A representative sampling of examinees and of the situations in which CQTs are employed.
3. Independence between the criterion and the polygrapher's judgment (which may be affected by all the information at his disposal).

4. Testing conditions in the experiment, which resemble those of a true examination. In particular, it is important that the examinees be anxious about the consequences of the test and take it seriously, and that the lie or the transgression be real.

A review of the literature reveals that no existing experiments (perhaps with the exception of the Ginton et al., 1982 study) meet all these requirements. In particular, there are no experiments that simultaneously fulfill both the first and the last requirement. All the experiments providing a satisfactory criterion are simulations, so called "mock crimes", in which the subjects know that they are participating in a role-playing game. The subjects designated as "guilty" are asked by the experimenter to steal an envelope containing money, or some other item. Then, all the participants (both "guilty" and "innocent") are examined by a polygraph interrogator. Participants are asked by the experimenter to hide their connection to the "simulated crime". When the experiment is over, they are given a monetary reward, thanked, and sent on their way. Obviously these conditions do not begin to resemble the conditions of a real life criminal investigation. There is no deception in the conventional sense of the word, and there is no anxiety about the consequences of the test (either for the "guilty" or for the "innocent").

A different category of studies uses real investigations, but samples only cases in which the truth was later revealed when a suspect confessed to the crime. Even if we choose to ignore the possibility of a false confession, we are faced with the problem of a severely biased sample, since there is liable to be a direct causal relationship between the results of the polygraph examination and the confessions. It is well known that polygraph examinations are used not only to reveal the truth but also as an investigative tool and a lever for inducing suspects to confess (see, Furedy

& Liss, 1986). Chances are obviously higher that a polygraph interrogator will try to extract a confession from a suspect showing "signs of lying" in his polygraph chart, than from a suspect who has not shown such signs. Therefore a sample of confessors is liable to include an inflated representation of cases in which the responses revealed obvious signs of guilt, and to under-represent cases of judgmental error. This point was demonstrated by Iacono (1991) who showed that in principle, in a sample based on confessions it should be possible to get nearly perfect accuracy rates, even if the true accuracy of the polygraph is at chance level (see also, Iacono & Lykken, 1999). Even if the assumptions made by Iacono are not fully met, it is clear that samples of confessed suspects might be seriously biased. Thus, confession studies do not satisfy the second and third conditions mentioned above. In addition, confession studies, which are based on real-life CQT investigations, suffer from the problem of contamination, discussed above. Thus, it is doubtful whether a given accuracy rate obtained in a confession study can be attributed to the physiological information.

The studies, which used the confession criterion, yielded inconsistent results, but the overall picture indicates a rather high occurrence of "false incriminations". In at least two CQT field studies (Barland & Raskin, 1976; Horvath, 1977) the accuracy rate among the "innocents" was no higher than chance (i.e., the same accuracy that would have resulted from tossing a coin). The mock-crime studies attempting to validate the CQT usually lead to better results, however as previously stated these are hard to take seriously, because of the difficulty of generalizing from such artificial conditions to true life examinations.

It should be noted that the problem of generalization is particularly acute in the case of the CQT because of the danger that even innocent suspects, when being interrogated about a real crime with possible severe consequences, might be more

aroused by the relevant questions, which focus on the main issue under investigation, than by the control questions, that deal with much less threatening issues. This factor is completely missing from the simulated crimes.

As indicated earlier, the study reported by Ginton et al. (1982) is the only study that came close to fulfilling all four requirements. However, no generalizations can be drawn even from this study because the realistic setup resulted in a large rate of dropouts, particularly of guilty participants. Most of the guilty participants either refused to take the polygraph test, or confessed just before taking it. Thus, at the end there were only two guilty participants who actually took the CQT, and clearly this sample does not allow for any generalizations.

The Ginton et al. (1982) study demonstrates the difficulties in estimating the validity of the CQT as a tool for detecting deception in real criminal investigations. But from a psychometric perspective, there is another important and necessary, though not sufficient condition, that any test must meet, namely reliability. Reliability, which can be estimated more easily than validity, refers to the consistency (or reproducibility) of the test scores. A test is considered reliable if its results tend to be replicated when the same individuals are tested several times under similar circumstances. In classical reliability theory, reliability is estimated by a correlation between two sets of equivalent measurements, such as administering the same test twice to the same group of subjects or administering two equivalent forms of the same test. In some circumstances reliability is estimated by correlating two sets of scores obtained from two independent observers (or judges) evaluating the performance of a given group in a specified condition. The different types of reliability estimates focus on different sources of inconsistencies, or measurement errors, and the choice of an appropriate reliability coefficient depends on the purposes of the specific

measurement and on the desired range of generalizability. Sometimes more than one type of reliability estimate will be required.

Reliability of polygraph-based scores, whether expressed by numbers (e.g. the "quantified" method suggested by Backster, 1963) or by qualitative categories (guilty, innocent, or inconclusive), refers to the degree to which these scores tend to be stable across measurement situations. Stability could be estimated by using one of two approaches: (a) testing the same individual twice on the same issue, using the same polygraphic method, but having two independent examiners administer the test; or (b) testing the subject just once, but letting two independent experts score the charts. Clearly, the second method yields reliability estimates of very limited use for evaluating the CQT because two independent examiners could, in principle, reach a complete agreement (especially if they were trained by the same polygraph school and if they used a quantified scoring method) despite a very low test-retest consistency. In other words, the second approach relates to just one source of measurement error - errors in chart scoring and interpretation. But the crucial question is not whether two CQT examiners could be trained to read a given polygraph chart consistently, but whether the procedure as a whole (including the most critical stage of constructing proper relevant and control questions) is reliable. In order to obtain proper estimates of the polygraph's reliability - estimates of measurement errors related to the test as a whole - one must use the first approach and administer the whole polygraph-based interrogation twice, using two independent examiners.

Unfortunately, reliability studies of polygraph-based classifications are scarce and, incredibly enough, those that have been conducted have employed only the between-examiners agreement approach (e.g., Barland, 1975; Horvath and Reid,

1971). Thus, it is impossible to conclude from the available data whether a given subject interrogated twice by independent examiners will be similarly classified.

Summary and conclusions

Six major arguments, indicating basic flaws in CQT polygraph examinations were raised. These flaws undermine the scientific basis of this technique:

1. There is a basic flaw in the rationale of the CQT; a flaw related to the selection of "Control Questions", and their lack of equivalence to the Relevant Questions. This flaw is liable to introduce a systematic bias, which might increase the rate of false-positive outcomes (innocent suspects classified as "guilty") into CQT examinations. It should be stressed that this bias is particularly damaging from a legal perspective because false-positive errors are just that type of mistake which criminal courts are particularly anxious to avoid.
2. The CQT is not a standardized test, and therefore it is unjustified to use it as a basis for comparisons of any sort. Thus, its results should be regarded as no more than subjective-impressionistic conjecture.
3. The CQT does not rely upon objective methods of quantifying the physiological measures. This feature of the CQT opens the door to biases in interpreting CQT outcomes and in drawing conclusions from them.
4. The CQT contains an element of "contamination". In other words it is possible that the conclusions made by a CQT examiner are based on information that was in his hands prior to conducting the examination, rather than the physiological measurements recorded by the machine. This feature of the CQT, which might introduce a confirmation bias (i.e., a bias favoring the prior hypothesis, over the alternatives), is particularly damaging from a legal perspective. The influence of

apriori information and of behavioral impressions could result in the polygraphist reaching conclusions on the basis of elements that would not be admissible as court evidence, and that are now introduced under the cover of a “truth machine”.

5. The CQT is vulnerable to both physical and mental countermeasures. These manipulations, that are relatively easy to learn, may increase the false-negative error rate. Mental countermeasures are particularly problematic because they cannot be detected by the examiners.
6. Finally, there is neither sound empirical, research-based evidence that indicates any kind of validity for CQT polygraph tests, under the realistic interrogative situation, nor are there good reliability studies demonstrating consistent outcomes derived from independent CQT examinations.

From all the above it appears that one cannot consider a CQT as an objective test, or tool based on scientific principles. In fact, there is no evidence that the CQT is more than another aid for interrogators who form subjective impressions based on many factors. Moreover, it is far from clear that the impressions of a CQT investigator are superior to those of any other investigator who questions suspects without the benefit of the polygraph. In particular, the use of the CQT can cause several judgmental biases, some of which mitigate against innocent suspects and against truthful respondents. A CQT polygraph test depends on the interrogator and his interaction with the examinee and could be influenced by personal inclinations of the interrogator and the premises of his investigation.

Once all the problematic aspects of the CQT are acknowledged, we must turn to the question of its applications. Currently, applications of the CQT can be classified into the following three general classes: (1) The CQT can be introduced as an admissible evidence in courts; (2) It can be used by law-enforcement agencies as an

aid in their investigations of specific acts (e.g., criminal acts); (3) It can be used for selection and screening of employees. The final section of this chapter is devoted to a brief discussion of each of these possible applications.

Although there have been various attempts to introduce CQT results as admissible evidence in U.S. courts (e.g., Commonwealth of Massachusetts v. Woodward, 1997; United States v. Cordova, 1998; United States v. Scheffer, 1998), the present analysis indicates that this application is highly problematic. We have stressed that at least two features of the CQT are particularly problematic from the legal perspective. The nature of the control questions, and the inference rule used to classify examinees as deceptive or truth-tellers, might increase the risk of false-positive errors, and this in itself is a sufficient reason to prohibit the use of the CQT in criminal trials. Contamination is the second problematic feature of the CQT from the legal perspective. As indicated by Ben-Shakhar et al. (1986) this feature of the CQT might open a back door for all kinds of inadmissible evidence, which may enter the courtroom unchecked and unchallenged through their influence on the knowledgeable CQT examiner.

In addition, the non-standardized nature of the CQT and the fact that it is not based on scientific principles make it a poor candidate for assisting the trier of fact. Recently, Saxe and Ben-Shakhar (1999) analyzed the admissibility of CQT results in the Federal Courts of the United States, in light of the new guidelines set by the U.S. Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). This analysis demonstrated that the concepts of reliability and validity as commonly used by behavioral scientists (e.g., Brennan, 1992; Messick, 1995) are not applicable to the CQT. In other words, the CQT does not satisfy the major Daubert criteria (testability

and known error rate, as well as reliability and validity) and therefore its outcomes should not be used as admissible evidence.

Some have suggested that CQT results might be ruled admissible in civil cases, provided both sides have agreed to this in advance. However, even this application is very problematic, because the phrase "lie detector" and the mystery it evokes could mislead many people and cause them to agree to a polygraph test without having any idea of its real nature. People could agree to the test under the impression that it is indeed a serious scientific tool for determining truth or lies, when in fact, as I have tried to show, this is not at all the case. The very suggestion that someone should undergo a "lie detection" test might be misleading and unfair. This practice could undermine innocent people.

The second possible application of the CQT as an aid to the police and other law-enforcement agencies in their investigations differs drastically from the legal usage of the CQT. First, police investigations do not need to rely on scientifically-based tools, and many other investigation methods that are not necessarily scientific are being constantly used. Second, the consequences of police interrogations differ drastically from those of legal procedures. Typically, police interrogations end in a decision either to press charges, or not. As long as CQT results are inadmissible in Criminal Courts, the police cannot rely on it, as the only evidence, when deciding to press charges against a suspect. Consequently, the emphasis on false-positive errors, which characterizes the legal system, does not hold for police investigations, and therefore many of the problematic features of the CQT, which may increase false-positive errors, are less damaging in the context of police interrogations. A more detailed discussion of these issues, from a decision-theoretic perspective can be found in Ben-Shakhar, Lieblich and Bar-Hillel, (1982).

Thus, from a psychophysiological perspective, there are no compelling reasons to object to the use of the CQT as an aid in police investigations, provided that it would be acknowledged that CQT results constitute no more than an opinion of an investigator. In addition, it should be pointed out that an alternative method of psychophysiological detection, which does rest on sound scientific principles is available. I am referring to the GKT, which was originally endorsed by Lykken (e.g., 1974, 1998), and subsequently by many others (e.g., Ben-Shakhar, 1991b; Ben-Shakhar & Furedy, 1990; Elaad, 1998; Elaad & Ben-Shakhar, 1997). Currently, the GKT is rarely applied by law enforcement agencies in North America and Israel (see, Podlesney, 1993). It has been argued that it is difficult to apply this technique, because it requires that salient features of the event under investigation will be concealed from the suspects. On the other hand, the GKT has been used for many years by Japanese law enforcement agencies as the preferred method of psychophysiological detection (Ben-Shakhar & Furedy, 1990; Fukumoto, 1980; Yamamura, & Miata, 1990). This fact implies that the feasibility of the GKT is greater than the current beliefs of North American and Israeli experts, and it is highly recommended that greater efforts will be made to replace the CQT by the more standardized and better controlled method of the GKT. Recently, Ben-Shakhar, Bar-Hillel and Kremnitzer (1999) discussed the possibility of using the GKT as admissible evidence, but this is beyond the scope of this chapter.

The third possible application of the CQT, as a tool for personnel selection and screening is entirely different from the previously discussed applications. Contrary to criminal investigations, selection and screening does not revolve around a specific and known event (e.g., crime). Rather, an attempt is being made to detect dishonest behavioral tendencies. Such an attempt is based on an additional assumption, namely

that dishonest behaviors exhibit a cross-situational consistency. This assumption has been questioned by several researchers (e.g., Michel & Peake, 1982).

Ben-Shakhar (1989) labeled this kind of application (e.g., screening) as an event-free usage of the CQT, as opposed to the more typical event-related usage. He argued that all the problematic features of the CQT become even more severe under the event-free application, because this method cannot be used in a straightforward manner. Recall that the relevant questions used in the CQT pertain to a specific event (crime). In order to use it for detecting hypothetical (or future) crimes, control questions (which relate to general misdeeds) must play the role of the relevant questions. In other words, enhanced physiological reactions to the typical control question (e.g., 'Have you ever stolen from your employer?') are now taken as an indication of deception and a consistent responding to those questions might mean that an applicant to a certain job will be rejected on the grounds of failing the polygraph test. But to make such inferences, one must compare the responses to those new relevant questions with the responses to equivalent control questions. Unfortunately, it is impossible to construct such control questions, because they must relate to other hypothetical crimes of similar importance, and naturally a consistent responding to those questions is not going to make the applicant more attractive as a future employee.

In addition, there is no empirical evidence to support the event-free application of the CQT. Thus, it is not surprising that the United States Congress outlawed the use of polygraphs for screening of employees in the privet sector (The Employee Polygraph Protection Act, 1988). It is similarly not surprising that even the keen supporters of the event-related applications of the CQT, including its use as

admissible evidence, have opposed to its use for screening and selection. (e.g., Honts, 1991, 1992).

References

Abrams, S. (1999). A response to Honts on the issue of the discussion of questions between charts. Polygraph, 28, 223-228.

Backster, C. (1963). Polygraph professionalization through technique standardization. Law and Order, 11, 63-64.

Barland, G.H. (1975). Detection of Deception in Criminal Suspects: A Field Validation Study. Ph.D. dissertation, University of Utah.

Barland, G. H., & Raskin, D. C. (1976). Validity and reliability of polygraph examinations of criminal suspects (Report No. 76-1, Contract No. 7599-0001). Law Enforcement Assistance Administration: United States Department of Justice.

Ben-Shakhar, G. (1989). Non conventional Methods in Personnel Selection in: P: Herriot (Ed). Handbook of Assessment in Organizations: Methods and practice for recruitment and appraisal. Chichester: John Willy & Sons, 469-485

Ben-Shakhar, G. (1991a). 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. (1991b). Future prospects of psychophysiological detection: Replacing the CQT by the GKT. In J.R. Jennings, P.K., Ackles and M.G.H. Coles (Eds.), Advances in Psychophysiology, 4. Greenwich, Conn: JAI Press.

Ben-Shakhar, Bar-Hillel, M. & Kremnitzer, M. (1999). On the admissibility of the Guilty Knowledge Technique in court. Manuscript in preparation.

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 Technique: The Effects of Mental Countermeasures. Journal of Applied Psychology, 81, 273-281.

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

Ben-Shakhar, G., Lieblich, I. and Bar-Hillel, M. (1982). An evaluation of polygrapher's judgments: A review from a decision theoretic perspective. Journal of Applied Psychology, 67 , 701-713.

Brennan, R. L. (1992). Generalizability theory. Educational Measurement: Issues and Practice, 11, 27-34.

Chapman, L. J., & Chapman, J. P. (1982). Test results are what you think they are. In: D. Kahneman, P. Slovic, & A. Tversky (Eds.). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, pp. 239-248.

Commonwealth of Massachusetts v. Woodward (D. Boston, 1998), SJC-07635.

Darley, J.M., & Gross, P.H. (1983). A hypothesis-confirmation bias in labeling effects. Journal of Personality and Social Psychology, 44, 20-33.

Daubert v. Merrell Dow Pharmaceuticals. 113 C. Ct. Supp. 2786 (1993).

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

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., & Ben-Shakhar, G. (1998). The Role of Prior Expectations in Polygraph Examiners Decisions. Psychology, Crime and Law, 4,1-16.

Employee Polygraph Protection Act of 1988. Public Law 100-347, 29 U.S.C. 2001 (1988).

Fukumoto, J.(1980). A case in which the Polygraph was the sole evidence for conviction. Polygraph, 9, 42-44.

Furedy, J.J. & Heslegrave, R.J. (1991). The forensic use of the polygraph: A psychophysiological analysis of current trends and future prospects. In: J.R. Jennings, P.K. Ackles and M.G.H. Coles (eds.), Advances in Psychophysiology, 4, Jessica Kingsley Publishers Ltd.

Furedy, J. J., & Liss, J. (1986). Countering confessions induced by the polygraph: Of confessionals and psychological rubber hoses. The Criminal Law Quarterly, 29, 92-114.

Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real life situation. Journal of Applied Psychology, 67, 131-137.

Honts, C. R. (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. Forensic Reports, 4, 91-116.

Honts, C.R. (1992). Counterintelligence scope polygraph (CSP) test found to be poor discriminator. Forensic Reports, 5, 215-218.

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

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

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.

Horowitz, S.W., Kircher, J.C., Honts, C.R., & Raskin, D.C. (1997). The role of comparison questions in physiological detection of deception. Psychophysiology, 34, 108-115.

Horvath, F. S. (1977). The effect of selected variables on interpretation of polygraph records. Journal of Applied Psychology, 62, 127-136.

Horvath, F. S., & Reid, J. E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. The Journal of Criminal Law, Criminology and Police Science, 62, 276-281.

Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In J. R. Jennings, P. K. Ackles & M. G. H. Coles (Eds.), Advances in psychophysiology, 4, 1-101. London: Jessica Kingsley.

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. St. Paul, MN: West Law.

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. Volume 1. Pocket Part. St. Paul, MN: West Publishing, pp. 174-184.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

Klayman, J. & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211-228.

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.

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

Lykken, D. T. (1978). The psychopath and the lie detector. Psychophysiology, 15, 137-142.

Lykken, D. T. (1979). The detection of deception. Psychological Bulletin, 86, 47-53.

Lykken, D. T. (1998). A tremor in the blood: Uses and abuses of the lie detector. New York, NY: Plenum.

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.

Mischel, W. & Peake, P.K. (1982). Beyond deja vu in the search for cross-situational consistency. Psychological Review, 89, 730-755.

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.

Podlesny, J.A. & Raskin, D.C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology, 15, 344-359.

Raskin, D.C. (1982). The scientific basis of polygraph techniques and their uses in the Judicial process. In: A. Trankell (Ed.). Reconstructing the past: The role of psychologists in the criminal trial. Stockholm: Norsted & Soners.

Raskin, D.C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding applications and acceptance of polygraph evidence. Utah Law Review, 60, 29-74.

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. Volume 1. Pocket Part. St. Paul, MN: West Publishing, pp. 160-174.

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 Law.

Raskin, D. C., & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. Psychological Bulletin, 86, 54-58.

Reid, J.E. & Inbau, F.E. (1977). Truth and Deception: The Polygraph ("Lie Detection") Technique. Baltimore: Williams and Wilkins.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. Behavioral Brain Sciences, 1, 377-415.

Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. American Psychologist, 46, 409-415.

Saxe L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. Psychology, Public Policy and Law, in press.

Saxe, L., Dougherty, D., and Cross, T. P. (1983). Scientific validity of polygraph testing (OTA-TM-H-15). (Report for the U.S. Congress Office of Technology Assessment). Washington, DC: U. S. Government Printing Office.

Saxe, L., Dougherty, D., & Cross, T. P. (1985). The validity of polygraph testing: Scientific analysis and public controversy. American Psychologist, 40, 355-366.

Snyder, M., & Swann, W. B., Jr. (1978a). Behavioral confirmation in social interaction: From social perception to social reality. Journal of Experimental Psychology, 14, 148-162.

Snyder, M., & Swann, W. B., Jr. (1978b). Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology, 36, 1202-1212.

United States v. Frank Javier 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.