

EFFECTIVE POLICING

Understanding How Polygraph Tests Work and Are Used

WILLIAM G. IACONO

University of Minnesota

Forensic applications of polygraph techniques rely primarily on the control or comparison question test (CQT). The author describes the CQT and its theoretical basis, and how it is used and evaluated by the polygraph professionals, and by scientists at arms length from the polygraph community. Because the CQT (a) has a weak theoretical foundation, making it unlikely that it can be as accurate as polygraph proponents claim, (b) is biased against the innocent, and (c) may be subject to countermeasures used by the guilty to appear truthful, CQT results cannot constitute evidence of either deception or truthfulness. In the absence of insight into brain mechanisms that underlie deception, it may be difficult to develop a valid lie detector. However, methods are available for detecting guilty knowledge, information that only the perpetrator of a crime and the police possess, which are ripe for further development as forensic applications.

Keywords: polygraph; control question technique; comparison question test; police interrogation; guilty knowledge test

To the dismay of a fearful public and the police, many crimes with a plausible suspect go unprosecuted due to a lack of hard evidence. It is possible that no amount of detective work will turn up the leads necessary to clear or indict the suspect. However, the evidence needed to resolve the case is present in the mind of the perpetrator. If only we could accurately access this information, more crimes would be solved, the innocent would rarely be falsely imprisoned, and the guilty liar would always be convicted. For more than half a century, the polygraph profession has claimed that we can resolve seemingly unsolvable cases by tapping into the physiological basis of a suspect's veracity. In this article, I will examine how well we can detect lying, using conventional police polygraph methods, and how law enforcement agencies can best use deception detection procedures as investigative aids. My focus will be the so-called control or comparison question test (CQT), a procedure that is widely used in forensic settings.

PSYCHOPHYSIOLOGICAL BASIS OF THE CQT

It is generally recognized that there is no physiological response that is uniquely associated with lying. Moreover, the brain mechanisms involved in lying are unknown, rendering difficult the development of techniques that can distinguish liars from truth tellers. Hence,

AUTHOR'S NOTE: Please address all correspondence to William G. Iacono, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455; e-mail: wiacono@tfs.psych.umn.edu.

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 35 No. 10, October 2008 1295-1308

DOI: 10.1177/0093854808321529

© 2008 International Association for Correctional and Forensic Psychology

detection of deception methods is designed to compare physiological reactions to different types of questions. The first commonly used lie detection procedure, the relevant-irrelevant technique, involved comparing the response to crime-relevant questions (e.g., a question in a case involving alleged child sexual abuse, "Did you insert your finger in Susan's vagina?") with irrelevant questions covering inconsequential topics for which the correct answer was readily apparent ("Is today Tuesday?"). This technique is seldom used in forensic settings because even innocent people are more reactive to the threatening accusation contained in the relevant question than to the innocuous content covered by the irrelevant question (Horowitz, Kircher, Honts, & Raskin, 1997).

The CQT was developed to address the obvious imbalance in the psychological significance of the two types of questions by replacing the irrelevant query with a comparison question. Historically referred to as the "control" question, this question is also accusatory, but it does not deal with a transgression related to the commission of the crime. Instead, it probes the integrity of a suspect by querying about possible past misbehaviors that one who committed such a crime might also have engaged in. For instance, for a sex abuse case, the control question might take the form "Have you ever committed a sexual act you were ashamed of?" or, because all relevant questions involve lying, "Have you ever lied to get out of trouble?"

The CQT is also sometimes referred to as the "probable lie test" because it assumes that the control questions are likely to elicit lies or at least nervous denials. CQT proponents argue that because the control questions are the only questions for which the innocent have any reason to be concerned, truthful people will respond more strongly to the control than to the relevant question. By contrast, liars are expected to find the relevant questions of greater concern, so they should respond more strongly to these questions.

CQTs are administered such that there are multiple relevant questions (typically three), each paired with a different control question. These questions are typically repeated three times, with the order of questions changed with each repetition. A CQT may also include other questions (a test often has 10 questions total), but only the relevant and control questions are evaluated to determine deception. Today's polygraph examiner is likely to use a laptop computer to digitize and record physiological reactions to the relevant and control questions. These reactions include skin conductance, monitored from the finger tips; respiration, recorded from belts around the chest and abdomen; and cardiovascular activity, tracked from a partially inflated blood pressure cuff placed around one arm. Computer software can be used to compare the magnitude of each type of physiological response to each control question and the relevant question with which it is paired, ultimately yielding an (unverifiable) estimate of the likelihood that an examinee is truthful or deceptive (e.g., "The probability of deception is less than .01"). Most examiners have been trained to process the physiological data manually to achieve the same goal. Because the algorithms on which computerized scoring is based are proprietary and thus cannot be evaluated, most examiners prefer to rely on their own scoring because they can defend it against outside scrutiny. Most tests end with conclusive verdicts of either deception or nondeception. However, examiners in private practice typically claim about 10% of CQTs end inconclusively, and Patrick and Iacono (1991b) found that police polygraph operators had an inconclusive rate of just more than 20%. If possible, suspects with such outcomes are retested until a definitive result is obtained.

TEST ADMINISTRATION: THE CQT INTERVIEW AND INTERROGATION

The ideal lie detector test would be administered something like a medical X-ray assessment. When a trained X-ray technician administers the standard protocol, the results are independent of the procedure followed. For instance, were two different technicians to obtain X-rays on the same person, the resulting images would be highly similar. However, this type of standardization does not exist for the CQT, and it is important to understand how this can affect the outcome of the test.

The CQT is not so much a single test as it is a collection of procedures that have in common the inclusion of relevant and control questions. Whatever specific form it takes, the CQT is more accurately characterized as a polygraph-assisted interview than a test. This interview has three phases. It begins with a pretest component during which the examinee is interviewed by the examiner and the questions are formulated. This phase of the interview can cover a lot of ground but is likely to involve asking test subjects about their views regarding their honesty, the nature of the punishment that would be appropriate for the crime, the factual evidence surrounding the crime, the subject's explanation or interpretation of the facts, and how best to word the relevant questions so they can be understood and answered unambiguously "yes" or "no."

An important goal of this interview phase is to convince the subject that the control questions are just as important as the relevant questions. This is accomplished by getting the subject to agree that the kind of person likely to commit the particular crime would also be likely to engage in other types of misbehavior, such as deceiving the police. Without ever making a distinction between the control and relevant questions, the examiner then introduces possible control questions that the subject also agrees are reasonably worded and can be answered unambiguously; for example, "Have you ever lied to a person in a position of authority?" Because the examiner admonishes the subject that the test will only work if the subject is completely truthful, many subjects, when confronted with a control question like this, will indicate that they have in fact lied to an authority figure; for example, to a parent or teacher. At this point, the control question is reworded so that it begins with this question: "Other than what you have told me about, have you ever lied to a person in a position of authority?" If the subject says, "yes," then the examiner leads the subject to believe that making such admissions is somewhat incriminating thereby hoping that the subject will refrain from revealing any additional misdeeds. No matter how many times an innocent person qualifies the coverage of a control question, the question nevertheless is still assumed to elicit a probable lie or at least significant concern when asked. It is noteworthy that the examiner must pick control questions that are appropriate, given the character of the subject. A repeatedly convicted felon who probably has a life history of lying to authority figures is likely to be asked a control question that covers more serious transgressions; for example, "Have you ever committed an undetected felony?"

The procedure followed to develop questions during the pretest interview is often characterized as establishing the "psychological set" of the examinee, the unsubstantiated notion that the individual's attention is involuntarily channeled to whatever question holds the greatest immediate threat to his or her well-being (Abrams, 1989; Matte, 1996). For the guilty, this means that the examiner has effectively ensured that their concern is directed to the relevant question. The psychological set of the innocent is appropriately established when the control question conveys the greater threat.

The second phase of the CQT involves asking the questions while the physiological reactions are recorded. Because the inflated blood pressure cuff becomes uncomfortable and about 20 s must elapse between each question to provide ample time for the physiological response to register and recover, a typical 10-question CQT lasts about 3 min. Each run through the question list produces a chart (or the computer equivalent of a chart) recording of the physiological reactions to the questions. Often, three such charts are obtained with the order of questions altered for each. Between each pair of charts, the subject is given the opportunity to alter the wording of questions that, with additional reflection, may seem unclear or confusing.

Some examiners also include a procedure designed to convince subjects they will be caught if they lie. The subject may be asked to choose a card from a marked deck or to write down a number between 1 and 10 and place it face up so both the subject and examiner can see it. The examiner then asks the subject if he has selected any of a series of cards or numbers, including the chosen one, to see if the physiological reactions are strongest to the subject's pick. This test, sometimes called a stim test (intended to stimulate the subject to increase detectability) or acquaintance test (acquainting the subject with the inevitability of detection), is interpreted such that the subject's pick is always detected.

For subjects who are deemed to have deceptive outcomes, the third and final phase of the CQT involves confronting the subject in an attempt to obtain an admission or outright confession. Most examiners are expert interrogators who are adept at using what they know about the subject's character and demeanor during the interview, the crime facts, and the CQT outcome to convince subjects that it is in their interest to confess.

It should be apparent from the above description that the CQT is not a standardized, objective procedure. Even when two examiners are trained to administer the same CQT protocol, the resulting CQTs will be different. CQTs are likely to vary in the choice of questions formulated to serve as relevants and controls, the wording and ordering of questions, the inclusion of other questions besides the relevants and controls, whether to include a stim test, the manual and/or computer scoring and interpretation of the resulting physiological data, and the degree to which the examinee buys the idea that all of the questions are equally determinative of test outcome. Although modern CQTs are computerized, computerization has no effect on any of these method variations; computers only ensure that the digitized physiological data are recorded for review and analysis. Computer scoring ensures reliability by reducing scoring inconsistency, but scoring algorithms have not been made available for scientific evaluation, have unknown validity, and are entirely dependent on the quality of the examiner's question development and administration of the CQT. Most examiners agree that the key ingredient for a competently administered CQT is the examiner. Hence, successful lie detection rests with the examiner, not the polygraph or computerized scoring procedure.

SCIENTIFIC SKEPTICISM REGARDING CQT THEORY

Evaluation of a test's efficacy begins with an appraisal of its theoretical foundation. For the CQT, this involves evaluating the intended function of relevant and control questions. It is easier to understand why guilty suspects can be expected to fail a CQT than it is to understand why innocent suspects can be expected to pass. Because the relevant question has obvious import to the perpetrator of a crime and the typical control question is likely to

seem inconsequential by comparison, larger responses can be expected in the category of relevants than in control category questions. However, the relevant question is just as relevant for the innocent suspect who fears false detection. One of the reasons why physiological deception detection seems plausible to us is that we can all recall an occasion when, caught in a lie, our face flushed, our heart rate picked up, and our hands became clammy. Field polygraphs are quite capable of registering these physiological changes; indeed, they work as well as sophisticated laboratory polygraphs for this purpose (Patrick & Iacono, 1991a). However, these same physiological reactions are likely to arise with the passionate denial that occurs when false charges are leveled against us. The fear of detection is indistinguishable from the fear of false detection; the physiological reactions recorded by the polygraph under these two circumstances cannot be discriminated one from the other.

Other factors also undermine our confidence that control questions work as intended. If an innocent person is not lying in response to a control question, cannot think of a lie associated with the control question, or has divulged all instances of lying during the review of control questions carried out in the pretest interview (i.e., the subject actually believes the examiner's assertion that the CQT cannot be passed if any lie is held back), the test will be failed. Hence, the CQT is biased against innocent suspects. As Patrick and Iacono (1991b) demonstrated, police polygraph examiners appear to understand this, because virtually the only time they disregard their own physiological data collected during a CQT, the data indicate deception. Hence, sometimes charts that clearly indicate deception are countermanded in examiner-written reports that conclude the result was inconclusive or truthful. By contrast, nondeceptive charts are almost never interpreted by examiners as indicating they believe the subject is actually deceptive, nor are such charts reclassified by examiners as inconclusive.

Even for the guilty we have no way of knowing for certain that the CQT works as intended. As noted elsewhere (Iacono & Patrick, 1987), there are many reasons why a guilty person may respond more strongly to a control than to a relevant question. For instance, in the case of sex offenders, many of their crimes are undetected, but these crimes are likely to be covered by a typical control question, for example, "Have you ever committed a sex act you were ashamed of?" A guilty sex offender who is worried that the polygraph will lead to suspicion in other crimes by virtue of concern directed at control questions has ample reason to respond disproportionately more strongly to these questions.

These criticisms of the CQT have not been effectively addressed by the polygraph profession. For instance, to circumvent problems associated with probable lie control questions, a CQT alternative has been proposed, the directed lie test (DLT). In a DLT, the probable lie control question is replaced with a question that the suspect agrees will be answered with a deliberate lie, for example, "Have you ever made even one mistake?" The suspect is asked to think of the mistake when responding to the directed lie question, ensuring that the question elicits a lie. However, for the innocent, the directed lie question still does not match the relevant question in psychological significance, and the fact that it does not is obvious, making it analogous to the irrelevant question in the relevant-irrelevant technique. For the guilty, the directed lie could deal with a matter of equal or greater significance than that covered by the relevant question. Although used forensically, the DLT does not have a strong following, in part because the purpose of the directed lie questions is transparent, making it easier for individuals who use countermeasures to understand that it is this question to which they should attempt to augment their physiological response.

Another way the polygraph community has dealt with the criticisms directed at the use of probable lie controls has been to stop referring to them as such. Hence, in 1999, after almost 50 years of reference to these questions as *controls*, the American Polygraph Association renamed the CQT as the *comparison question test*.

CQT ACCURACY

The debate concerning CQT accuracy has raged for decades and is too broad a topic to consider fully here, but has been hashed out thoroughly elsewhere, especially in a legal text that presents comprehensive detailed arguments supporting both the case for (Honts, Raskin, & Kircher, 2002) and the case against (Iacono & Lykken, 2002) forensic applications of the CQT. The most thorough analysis of polygraph testing undertaken to date was published as a book by the National Research Council (NRC) of the National Academy of Sciences (NRC, 2003). The NRC was composed of a panel of distinguished scientists who, although they had no direct involvement with polygraph testing, had the expertise to evaluate polygraph tests. The NRC examined the world literature on polygraph testing, including unpublished U.S. government data and reports, some of which, because they were classified, have never been available for analysis. Because the CQT has received the most research attention, much of the NRC report is relevant to the evaluation of the CQT. Some of the NRC findings are highlighted below.

THE PRACTITIONER PERSPECTIVE

In my experience, which now spans several decades and includes considerable contact with examiners working for government agencies and in private practice, practitioners are convinced that errors are so minimal that they are virtually nonexistent. Furthermore, although they may acknowledge that errors are occasionally made, they attribute them to incompetent examiners. It is not uncommon to encounter experienced examiners, who have given hundreds or thousands of tests, claiming that they are unaware of ever having made a mistake. Confronted with the fact that the CQT has no compelling theoretical basis and that therefore, claims of high accuracy seem implausible, practitioners respond that the absence of theory is irrelevant because, in practice, the CQT works and that's all that matters.

Few examiners have the scientific training necessary to evaluate the literature on CQT accuracy. Their training consists of practical experience in law enforcement and completion of a course on polygraph techniques completed with one semester or less of class work. Hence, they turn to institutions such as the American Polygraph Association or to PhD-level psychologists who also practice polygraphy for their opinions regarding validity. Both sources marshal evidence suggesting that the CQT has extremely high accuracy. The American Polygraph Association Web site (www.polygraph.org) cites an analysis of validity studies that they commissioned to an entity referred to as Forensic Research Incorporated (Forensic Research Incorporated, 1997). This analysis proclaims a 98% accuracy figure based on a review of a dozen CQT field studies involving over 2,000 real life forensic evaluations. The handful of psychologists directly involved in the polygraph profession have reached similar conclusions. Two of the most visible scientists practicing polygraphy, David Raskin and Charles Honts, have argued that when used as forensic evidence in court, their

analysis of CQT field studies supports an overall accuracy rate for the CQT of 97.5% (Honts et al., 2002).

An interesting question concerns how it is possible that dozens of field studies based on thousands of cases can be so wrong in their appraisal of polygraph accuracy. The answer derives from the inability to appreciate the significance of a research design flaw that runs through field studies. Despite the fact that the existence of this serious defect has been known for decades, the polygraph community has continued to publish and cite these flawed studies to substantiate their claims of near infallibility.

A case in point is a recent CQT validity study carried out by three polygraph examiners (Mangan, Armitage, & Adams, 2008), which claims 100% accuracy in the identification of both guilty and innocent suspects. Like the other studies based on real life applications of the CQT that preceded it, this investigation relied on a confession as the guilt criterion. A confession not only identifies the guilty party, but it also establishes as innocent cosuspects in the same case. A typical confession study determines validity by examining how often charts scored deceptive came from guilty people and how often those scored nondeceptive came from innocent people. In Mangan et al. (2008), when charts scored inconclusive were eliminated, following this verification method, the test outcome matched the criterion perfectly, so CQT accuracy was 100%! Although many questions can be raised about the adequacy of the confession criterion (e.g., are cases that yield confessions representative of cases in general? Aren't some confessions false?), the serious problem introduced by relying on confessions is that they are obtained by the examiner (during the posttest phase of the CQT described earlier) after the examiner judges the CQT deceptive. Under these circumstances, the outcome of the test (deception) and the criterion (confessing) are not independent leading, of course, for the two to be in agreement virtually all of the time. Note that had the examiner erred by passing a guilty person, there would be no posttest interrogation and therefore no confession, and this false negative error would be missed by Mangan et al. Note as well that had the examiner erred by failing an innocent person, despite proceeding with the posttest interrogation, there would be no confession, and this false positive error would be missed by Mangan et al.

Under the circumstances, it should not be surprising that Mangan et al. (2008) reported 100% accuracy. What is surprising is that studies like this continue to be published. Iacono and Patrick (1987) pointed out this fatal flaw more than 20 years ago and demonstrated empirically how it leads to inflated estimates of polygraph accuracy (Patrick & Iacono, 1991b). I illustrated how this procedure can be used to prove 100% accuracy for the CQT even if its true accuracy was chance (Iacono, 1991), and a similar point has been made by other scientists reviewing this methodology (NRC, 2003, pp. 113-115). No discussion of the criterion contamination problem caused by relying on confessions can be found in Mangan et al., and citation to these critical sources is missing from their paper. However, their report is not inconsistent with the past practices of polygraph advocates; neither Forensic Research Incorporated (1997) nor Honts et al. (2002) dealt with this issue when advancing their claims of nearly perfect CQT accuracy.

COUNTERMEASURES

Honts and colleagues have published compelling findings showing that guilty individuals can readily defeat a CQT when briefly instructed how to do so (Honts, Devitt, Winbush, & Kircher, 1996; Honts, Hodes, & Raskin, 1985). This can be accomplished by covertly

enhancing the response to control questions (e.g., by engaging in mental arithmetic when these questions are asked or lightly biting the tongue), and these covert maneuvers are undetectable. Moreover, although the CQT is ideally administered in a manner that makes it difficult to readily identify control questions, anyone who wishes to can easily find out how to identify controls by accessing articles such as this one or Web sites such as www.antipolygraph.org. Nevertheless, these practicing scientists have concluded that the possible use of countermeasures is of no consequence in real life CQTs because those undergoing CQTs cannot figure out on their own how to use them to advantage (e.g., Honts & Alloway, 2007). In Honts and Alloway (2007), this conclusion was reached by testing undergraduates who were asked to steal movie-pass vouchers. A total of 10 individuals in the guilty condition were asked to read a 220-page book containing information regarding how to beat a CQT, and if they were successful, they were to receive theatre passes as a bonus. One can reasonably ask how this motivational manipulation might compare to the motivation criminals have to learn to beat a CQT. The former involves working hard for a possible but trivial reward of movie tickets, the latter involves trying to avoid incarceration. Considered from this perspective, it is perhaps not surprising that only seven guilty subjects admitted trying to use countermeasures, and only two beat the CQT.

INDEPENDENT SCIENTIFIC EVALUATION

Scientists with no direct involvement in the polygraph profession have repeatedly reviewed the scientific literature on CQT accuracy. They have concluded that the accuracy claims of the polygraph profession are exaggerated and indefensible and that guilty individuals can use countermeasures to appear truthful on the CQT. These opinions are reflected in reviews spanning several decades (Lykken, 1974; Saxe, Dougherty, & Cross, 1985) as well as in recent evaluations of the literature carried out for texts in polygraphy and forensic psychology (Ben-Shakar, 2002; Iacono & Patrick, 2006; Oksol & O'Donohue, 2003), on behalf of the German legal system (Fiedler, Schmod, & Stahl, 2002), and for the U.S. government (NRC, 2003). That these negative conclusions are based on review of the same literature that proponents rely on indicates that the standards for what constitutes adequate science are considerably relaxed for those whose livelihood depends on the legitimacy of the CQT as a lie detector. Some of the findings of the NRC (2003) are especially noteworthy:

There has not even been any systematic effort to develop theoretical clarity regarding the mechanisms purported to cause differential responses to relevant and comparison question in relevant-irrelevant or comparison question polygraph tests. (pp. 92-93)

In short, the bulk of polygraph research . . . can be accurately characterized as atheoretical. (p. 94)

The polygraph studies that met our criteria for consideration do not generally reach the high levels of research quality desired in science . . . relatively few of the studies are of the quality level typically needed for funding by the U.S. National Science Foundation or the U.S. National Institute of Health. (p. 108)

Virtually all of the observational field studies have been focused on specific incidents and have been plagued by measurement biases that favor overestimation of accuracy, such as examiner contamination, as well as biases created by the lack of a clear and independent measure of truth. (p. 214)

In summary, we were unable to find any field experiments, field quasi-experiments, or prospective research-oriented data collection specifically designed to assess polygraph validity and satisfying minimal standards of research quality. (p. 115)

All of the physiological indicators measured on the polygraph can be altered by conscious efforts through cognitive or physical means, and there is enough empirical research to justify concern that successful countermeasures may be learnable. (p. 216)

Research on the polygraph has not progressed over time in the manner of a typical scientific field. It has not accumulated knowledge or strengthened its scientific underpinnings in any significant manner. (p. 213)

What is remarkable, given the large body of relevant research, is that claims about the accuracy of the polygraph made today parallel those made throughout the history of the polygraph: practitioners have always claimed extremely high levels of accuracy, and these claims have rarely been reflected in empirical research. (p. 107)

Almost a century of research in scientific psychology and physiology provides little basis for the expectation that a polygraph test could have extremely high accuracy. (p. 212)

These NRC conclusions are noteworthy in part because they are the types of observations that are often made in fields that are characterized as pseudoscience, especially those where practitioners have an economic stake to defend the legitimacy of their techniques. From the NRC (2003) critique, it is evident that the field (a) is devoid of meaningful theory, (b) has failed to accumulate knowledge, (c) relies on studies of poor quality, (d) ignores evidence that contradicts the likely effectiveness of the technique, (e) continues to make claims that are unsubstantiated, and (f) makes claims that are difficult to believe given what we know about human psychophysiology.

Further confirmation that scientists are doubtful of the claims of professional polygraphy derives from two surveys of scientific opinion carried out with members of the Society for Psychophysiological Research and Fellows of the American Psychological Association's Division of General Psychology (Iacono & Lykken, 1997). Both surveys obtained high return rates and achieved remarkably similar appraisals indicating that the CQT is not theoretically sound, that claims of high validity are unreasonable, and that countermeasures to defeat the CQT can easily be learned. Only the APA Fellows were asked to estimate the accuracy of the CQT; their estimate was 61.5%, far lower than the estimated accuracy advertised by the polygraph community. Against this backdrop of overwhelming skepticism, 93% agreed that strong empirical evidence should be required before the accuracy claims of proponents are believed.

The consensus among scientists at arm's length from the practice of polygraphy is that the CQT cannot possibly be highly accurate. However, these same scientists are not arguing that the CQT has no validity. The accuracy of the CQT is indeterminate. The NRC concluded that for those untrained in countermeasures, accuracy rates are "well above chance, though well below perfection" (p. 214). Although the meaning of this statement is itself ambiguous, David Lykken and I (Iacono & Lykken, 2002) argued elsewhere that although methodological limitations inherent to the polygraph literature make it impossible to arrive at precise estimates of CQT accuracy, the best available field studies support accuracies of 57% and 75% for innocent and guilty suspects, respectively. This conclusion is derived from an analysis of three studies that used real life cases to estimate accuracy in field applications (Horvath, 1977; Kleinmuntz & Szucko, 1984; Patrick & Iacono, 1991b). Unlike other field studies of the validity of the CQT, each of these investigations included criteria for determining ground truth that was at least in part independent of the outcome of the polygraph test. For example, Patrick and Iacono (1991b) used confessions obtained during the course of the police investigation to determine ground truth instead of relying only on confessions that followed a failed polygraph as the guilt criterion. As such, these studies are among the least likely to overestimate CQT accuracy.

WHY DOES CQT USE CONTINUE?

In the face of such strong conclusions to the contrary, it is worthwhile to consider why CQT use is widespread and, in fact, on the rise. The answer lies in part with the utility of polygraph testing. As noted, the third phase of the CQT interview involves interrogating those who are deemed to be deceptive. Armed with a failed polygraph, the examiner can use this result as an opening to probe the suspect regarding why the test was failed and what additional knowledge the suspect has about the case that is being withheld. In the hands of a skilled examiner, many suspects are led through a conversation that ends with an outright confession or perhaps the revelation of valuable information, such as evidence that incriminates someone else. A confession not only resolves a case (and as noted, provides a flawed criterion of ground truth in a validity study), but it also ensures successful prosecution. From a law enforcement perspective, a cost-benefit analysis supports the use of the CQT: The benefit of occasionally solving important cases, such as murders and sex crimes, clearly outweighs the cost of a typical polygraph program. In addition, because an innocent person who fails a CQT is unlikely to confess and because only confessions, not the failed CQT, are admissible as evidence, the risk to the innocent is perceived as minimal. Finally, although guilty individuals may erroneously pass a CQT, they can nevertheless be prosecuted based on other evidence. Polygraph testing is thus viewed by law enforcement as a valuable tool for solving cases that conserves limited resources that might profitably be applied to other cases.

This confession-eliciting utility feature has also been used to justify the expansion of polygraph testing to sex offender treatment. Public outrage over repeat offending by previously convicted sex offenders has led many states to mandate the use of polygraphs to monitor sex offenders on probation or parole. Under these circumstances, the examinee must pass a CQT confirming that inappropriate sexual behavior remains controlled. In addition, CQTs are given to sex offenders in state-run treatment programs. Sex offenders are known to conceal many of their offenses, but effective treatment is believed to depend on the ability of therapists to understand fully the range of their past offenses. Hence, many states now require inmates to take polygraph tests regarding their truthfulness about past offending and current possible reoffending, primarily because under the pressure of having to take a CQT, offenders are likely to make admissions about their sexual misbehavior that would otherwise remain hidden (Grubin, Madsen, Parsons, Sosnowski, & Warberg, 2004).

In addition, it is important to realize that CQTs given as part of a criminal investigation are typically administered in circumstances where the evidentiary trail is unlikely to result in a prosecution. Many polygraph programs associated with law enforcement agencies are viewed as a limited resource that should only be used after other means of resolving a case have been exhausted. Many forensic applications of the CQT involve sex crimes where the physical evidence is inconclusive. These may involve children who are unreliable witnesses or sexual assaults where the accused alleges the encounter was consensual. Relying on the CQT to assist evidence gathering, including securing confessions, can be an effective tool to resolve such cases, cases that otherwise would end with the perpetrator back on the street, possibly reoffending. It is for this reason that major law enforcement agencies continue to use polygraph tests despite persisting doubts about their validity.

More important, the confession-inducing feature of the CQT also helps explain why examiners are convinced the CQT is nearly infallible. The selective nature of the feedback, which examiners obtain regarding the likelihood that their verdicts are correct, serves to

convince them they do not make mistakes. That is, when someone fails a CQT and confesses, the examiner learns immediately that “deception indicated” was the correct call. If an innocent person fails, no confession is likely, and this error will go undetected (note that if the innocent person falsely confessed, this error would not only go undetected, it would further strengthen the examiners’ conviction that errors seldom occur). If a guilty person passes the CQT, no confession would follow, and this error would also go undetected. Under these circumstances, polygraphers occasionally learn when they are correct, but they do not learn about their errors.

Of course, there are other ways in which guilt or innocence might be established independently of the outcome of a polygraph test. However, because examiners are convinced of the accuracy of their methods, when there is a disagreement between the CQT outcome and some other criterion of guilt, in my experience, it is the guilt criterion that is seen as faulty, not the CQT. For instance, acquittal by a jury does not signify that the defendant was innocent but that guilt was not established beyond reasonable doubt. Conviction of an individual who passed a CQT simply confirms that sometimes justice is denied to the innocent.

RELYING ON THE CQT

Despite the convictions of polygraphers that the CQT is highly accurate, science does not support the use of polygraph tests. Even the proposition that the CQT has utility as a means of gathering incriminating evidence has not been evaluated empirically. Nevertheless, there is widespread evidence from anecdotal accounts that crimes are frequently solved by obtaining confessions following failed CQTs. In addition, the use of polygraph tests by law enforcement is not abating; if anything, it is expanding to other areas, such as monitoring programs targeting sex offenders on conditional release, and even in the investigation of presumed terrorists in the Iraq and Afghan wars (using a hand-held portable polygraph, Battelle Memorial Institute, 2007). Against this backdrop, what uses of the CQT seem justified?

1. It must be realized that the CQT cannot be used as the ultimate arbiter of truthfulness. In my experience of working with police who rely on polygraph tests, this is the most common mistake officers make when they are given the results of polygraph tests on suspects they are investigating. As Patrick and Iacono (1991b) found in their work with the Canadian police, once someone fails a polygraph test, the presumption is made that the perpetrator is known, and additional investigative work is likely to cease if the person does not confess. Sex offender treatment programs have also made this mistake, assuming that a failed CQT provides prima facie evidence of guilt (see Meijer, Verschuere, Merckelbach, & Crombez, in press).
2. Innocent individuals are more likely to fail than are guilty individuals to pass. This conclusion, which is supported by field studies of CQT validity (Horvath, 1977; Kleinmuntz & Szucko, 1984; Patrick & Iacono, 1991b), reflects the fact that CQT control questions do not provide adequate control for the emotional impact of the threatening accusation contained in the relevant questions. This means that in general, passed CQTs are more credible than failed CQTs.
3. Qualifying this conclusion, guilty individuals can produce truthful CQTs by engaging in countermeasures. It is not known to what degree countermeasures are used in real life cases, but information about how to use them can be easily accessed on the Internet (www.polygraph.com; www.antipolygraph.org). Polygraph examiners do not believe that countermeasures are effective because they sometimes catch individuals who make clumsy attempts to use them, such as by controlling their respiration, something that shows up readily in polygraph tracings.

However, when mental and covert physical manipulations have been found to be effective, even experienced examiners are not able to detect them (Honts et al., 1985, 1996).

4. The utility of the CQT for eliciting confessions makes it a valuable option to use in crimes where the evidentiary trail is weak. Because abuses can occur when the CQT is used this way, steps must be taken to avoid coercion. At a minimum, the entire CQT procedure should be recorded to allow independent review of the methods used to elicit the confession. Although most confessions that follow failed CQTs are probably genuine, accounts of false confessions are known. Whether confessions could be obtained as easily without using the CQT is not known. Voice-stress analysis, a discredited method for identifying liars (NRC, 2003), is also used by the police to obtain confessions, despite its lack of validity. Social psychologists have long known that belief in the efficacy of the technique may be all that is needed to obtain admissions from guilty individuals (Jones & Sigall, 1971).
5. The results of a CQT should best be seen as generating a hypothesis about where to focus resources during criminal investigation. Just as there may be many reasons why a suspect's fingerprints are found or are not found on a murder weapon, so there are many reasons why an individual may pass or fail a CQT.

LOOKING TO THE FUTURE

The CQT has been with us for more than 60 years. During this time, its use has grown, not diminished, despite repeated highly critical evaluations that have not been effectively rebutted. Hence, it is unlikely that forensic application of the CQT will change significantly in the near future.

How likely is it that new methods of deception detection with proven validity will emerge? Although functional MRI (magnetic resonance imaging) studies have been carried out to help elucidate the neurophysiological basis of lie detection (see Iacono, 2007, for a review), at present, the brain processes underlying deception are not understood. Until they are, it seems unlikely that a human lie detector capable of identifying a unique physiological signature associated with the act of lying will become available any time soon. However, an alternative to lie detection, the detection of guilty knowledge or concealed information, does hold promise as a forensic tool.

The guilty knowledge test (GKT), as introduced by Lykken (1959) 50 years ago, is a recognition memory test that assesses whether an individual has specific knowledge of a crime, knowledge only possessed by the police and the perpetrator. In a typical GKT, a suspect is exposed to guilty knowledge information as well as other types of information with similar thematic content. Imagine pictures of weapons being flashed on a computer screen. The perpetrator would be expected to recognize the weapon used in the crime, and the picture of this weapon, when processed by the brain, produces a recognition response evident in both autonomic nervous system reactivity and cerebral electrical responses. The other weapon pictures, because they are of no special significance to the perpetrator, elicit only weak responses by contrast. Because an innocent person with no crime-related knowledge has no reason to respond differentially to any of the pictures, no recognition response would be evident except by chance.

To increase the likelihood of accurate detection, many item series can be used for the GKT (e.g., pictures of the victim along with victims in other crimes, series of words where one word in the series is clearly associated with the crime—such as a word designating something that was removed from the crime scene). With the inclusion of a sufficient

number of item series, there is little likelihood that an innocent person would show recognition to all or even many of the guilty options by chance, so false positive outcomes are uncommon. This technique is the preferred method of deception detection in Japan (Nakayama, 2002), and it is conceivable that its application in countries that use the CQT will become more commonplace. At present, however, its use outside of Japan is rare.

The application of the GKT has been extended by substituting the recording of cerebral event-related potentials (ERPs) for autonomic reactivity. ERP-based GKTs have included manipulations that ensure a given administration to a specific individual was correctly carried out, allowing for inconclusive outcomes for some individuals. In addition, tests can be structured to lessen the likelihood of success with countermeasures, an area that clearly requires additional research (Mertens & Allen, 2008) before these procedures are advanced for field application. A full exposition of these methods lies beyond the scope of this article, but recent reviews cover them in some detail (Iacono, 2007, 2008; Meegan, 2008). Unlike the CQT, this technique has a strong scientific foundation and is based on current understanding of brain processes underlying memory. As such, it is likely to have other applications as well, including assessing memory loss associated with amnesia (Allen, 2002) and dissociative disorders (Allen & Iacono, 2001) as well as verifying the accuracy of eyewitness identification (Lefebvre, Marchand, Smith, & Connolly, 2007). Because this GKT adaptation has multiple potential applications, the likelihood is high that it will undergo further refinement.

REFERENCES

- Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington Books.
- Allen, J. J. B. (2002). The role of psychophysiology in clinical assessment: ERPs in the evaluation of memory. *Psychophysiology*, *39*, 261-280.
- Allen, J. J. B., & Iacono, W. G. (2001). Assessing the validity of amnesia in dissociative identity disorder: A dilemma for the DSM and the courts. *Psychology, Public Policy, and Law*, *7*, 311-344.
- Battelle Memorial Institute. (2007). *Efficacy of prototype credibility assessment technologies: PCASS final report*. Ft. Jackson, SC: Department of Defense Academy for Credibility Assessment.
- Ben-Shakar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103-126). San Diego, CA: Academic Press.
- Fiedler, K., Schmod, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology*, *24*, 313-324.
- Forensic Research Incorporated. (1997). The validity and reliability of polygraph testing. *Polygraph*, *26*, 215-239.
- Grubin, D., Madsen, L., Parsons, S., Sosnowski, D., & Warberg, B. (2004). A prospective study of the impact of polygraphy on high-risk behaviors in adult sex offenders. *Sex Abuse*, *16*, 209-222.
- Honts, C. R., & Alloway, W. R. (2007). Information does not affect the validity of a comparison question test. *Legal and Criminal Psychology*, *12*, 311-320.
- 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., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, *70*, 177-187.
- Honts, C. R., Raskin, D., & Kircher, J. (2002). 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* (Vol. 2, pp. 446-483). St. Paul, MN: West.
- 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. (1977). The effect of selected variables on the interpretation of polygraph records. *Journal of Applied Psychology*, *62*, 127-136.
- 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* (pp. 201-207). London: Jessica Kingsley.

- Iacono, W. G. (2007). Detection of deception. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 688-703). New York: Cambridge University Press.
- Iacono, W. G. (2008). The forensic application of "Brain Fingerprinting": Why scientists should encourage the use of P300 memory detection methods. *American Journal of Bioethics*, 8, 30-32.
- Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology*, 82, 426-433.
- Iacono, W. G., & Lykken, D. T. (2002). The scientific status of research on polygraph techniques: The case against polygraph tests. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Vol. 2, pp. 483-538). St. Paul, MN: West.
- Iacono, W. G., & Patrick, C. J. (1987). What psychologists should know about lie detection. In I. B. Weiner & A. K. Hess (Eds.), *Handbook of forensic psychology* (pp. 460-489). New York: John Wiley.
- Iacono, W. G., & Patrick, C. J. (2006). Polygraph ("lie detector") testing: Current status and emerging trends. In I. B. Weiner & A. K. Hess (Eds.), *The handbook of forensic psychology* (pp. 552-588). New York: John Wiley.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349-364.
- Kleinmuntz, B., & Szucko, J. (1984). A field study of the fallibility of polygraphic lie detection. *Nature*, 308, 449-450.
- Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, 44, 894-904.
- 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 detector industry. *American Psychologist*, 29, 725-739.
- Mangan, D. J., Armitage, T. E., & Adams, G. C. (2008). A field study on the validity of quadri-track zone comparison technique. *Physiology & Behavior*. Preprint version available from Scencedirect Website, <http://www.sciencedirect.com>
- Matte, J. A. (1996). *Forensic psychophysiology using the polygraph: Scientific truth verification-lie detection*. Williamsville, NY: J.A.M.
- Meegan, D. V. (2008). Neuroimaging techniques for memory detection: Scientific, ethical, and legal issues. *American Journal of Bioethics*, 8, 9-20.
- Meijer, E. H., Verschuere, B., Merckelbach, H. L. G. J., & Crombez, G. (in press). Sex offender management using the polygraph: A critical review. *International Journal of Law and Psychiatry*.
- Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, 45, 286-298.
- Nakayama, M. (2002). Practical use of the concealed information test from criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 49-86). San Diego: Academic Press.
- National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academy Press.
- Oksol, E. M., & O'Donohue, W. T. (2003). A critical analysis of the polygraph. In W. T. O'Donohue & E. R. Levensky (Eds.), *Handbook of forensic psychology: Resource for mental health and legal professionals* (pp. 601-634). San Diego, CA: Academic Press.
- Patrick, C. J., & Iacono, W. G. (1991a). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology*, 28, 632-638.
- Patrick, C. J., & Iacono, W. G. (1991b). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology*, 76, 229-238.
- Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist*, 40, 55-366.

William G. Iacono is a distinguished McKnight University professor at the University of Minnesota. His research currently includes studies of genetic liability for substance abuse and antisocial behavior as well as lie detection. His work on lie detection was cited by the United States Supreme Court in *U.S. v. Scheffer*, the 1998 case that eliminated polygraph tests from military courts martial. He has served as a consultant regarding lie detection to various U.S. government agencies, including the Department of Defense Polygraph Institute (renamed the Defense Academy of Credibility Assessment). He has testified in state and federal courts regarding the scientific status of lie detection, voice stress analysis, and "brain fingerprinting" on more than 40 occasions.