*Part I*

# Introduction

# Science on the rise: birth and development of the Concealed Information Test

*Christopher J. Patrick*

> Intrigued by the polygraphic equipment in my laboratory, my two assistants had asked if I did any lie detector work and I had been forced to admit that I knew nothing about the subject. Equipped as we were with time, facilities, and ignorance, we resolved to do an experiment on lie detection.
>
> (Lykken, 1981)

In 1959, University of Minnesota psychology professor David Lykken reported an experimental study of a new type of lie detector test that he termed the "Guilty Knowledge Test" (GKT). As indicated in the foregoing quote from his 1981 book, Lykken was unfamiliar with established methods of lie detection used by police and other field examiners at the time of this study. As a result, he relied upon basic principles of experimental psychology to devise a test that focused on probing for specific relevant knowledge of the incident under investigation rather than on detection of lying per se.

A cornerstone of Lykken's technique was the fundamental concept of experimental control: to ensure that the observed ("dependent") effect is attributable to the experimental ("independent") manipulation, one must establish a comparison condition that mirrors the experimental condition in all respects aside from the manipulation of interest. With this principle in mind, the GKT was composed of items in multiple-choice format, with alternative choices for each item (one of them crime-relevant, the others extraneous) formulated to appear equally plausible to an innocent examinee. As a function of this, consistently enhanced reactions to crime-relevant alternatives could readily be interpreted as indicating the presence of "guilty knowledge." Another cornerstone of the technique was probability theory: to minimize the likelihood of an innocent examinee exhibiting a guilty pattern of responding by chance, one can simply increase the number of items in the test. With this concept in mind, Lykken designed the GKT to include a series of different items, each referencing a distinct salient detail of the targeted incident – such that a specific probabilistic estimate of "guilt" (i.e., possession of

crime-relevant knowledge) could be computed on the basis of an individual's pattern of reactivity to critical items on the test. A further notable feature of Lykken's GKT was the use of a physiological response variable known to be sensitive to the familiarity or meaningfulness of a stimulus event – electrodermal reactivity, or galvanic skin response (GSR) as it was commonly known at the time.

Lykken's inaugural study of the GKT demonstrated the technique to be highly effective in distinguishing between knowledgeable ("guilty") and naïve ("innocent") participants tested regarding their involvement in alternative mock crime scenarios: forty-eight out of forty-eight participants (100 percent) tested regarding a scenario in which they *had not* participated were correctly identified as innocent; forty-four out of fifty (88 percent) tested regarding a scenario in which they *had* participated were correctly identified as guilty. Thus, the mean accuracy of the test across guilty and innocent suspects was 94 percent. In a subsequent study, published the following year, Lykken (1960) demonstrated that individuals pre-instructed in the use of countermeasures (i.e., strategies for defeating the test by inhibiting responses to crime-relevant alternatives, or augmenting reactions to control alternatives; see Chapter 11 of this volume) could nonetheless be successfully detected as guilty on the GKT by systematically comparing their magnitude of reactivity to varying alternatives across items of the test to assess for unexpected non-randomness in response patterns. Examples of non-random response patterns, indicative of deliberate effort to defeat the test, would include instances in which the largest response on most or all items occurred to the second alternative, or in which the crime-relevant alternative reliably yielded the *smallest* response.

Despite the novelty of the approach and the impressive results of these initial studies, the use of physiological measures to detect lies was not a primary investigative focus of Lykken's and his 1960 follow-up report was the last empirical study he conducted in this area. Nonetheless, stimulated by this foray into this domain of applied psychology, Lykken immersed himself in the available literature on the use of polygraph procedures by field examiners and over the years emerged as one of the world's leading scientific experts on lie detection – contributing many influential conceptual and critical review articles along with what would become the authoritative scholarly book on the subject.[1]

[1] A 1991 (auto)biography of Lykken, published in the *American Psychologist* on the occasion of his receipt of the association's lifetime career award for contributions to the field of psychology in the public interest, noted that: "From 1970 until recently, Lykken estimates that he has spent 25% of his professional time in advocacy relating to polygraphic interrogation. Although there is no longer much intellectual content in

Through his initial readings in the area, Lykken learned that a procedure akin to the GKT had been used at times by police polygraph examiners with criminal suspects. This procedure, termed the "peak of tension" test, was introduced by early American lie detector guru (and Stanford psychology dropout) Leonarde Keeler in the 1920s. The test entailed presenting the examinee with a series of alternative descriptors pertaining to one salient aspect of a crime (e.g., possible sums of money that could have been taken in a theft) and watching for a pattern of steadily increasing physiological arousal up to the point of the key descriptor (i.e., the actual amount taken), followed by a decline in arousal thereafter. A variant of this technique, the "searching peak of tension test," was used to probe for some key detail of a crime (e.g., the location of a body within a general target area) that was in fact unknown to investigators. Keeler is also credited with developing the "card" or "numbers" test, a GKT-like demonstration used by examiners to this day to persuade test subjects of the polygraph's effectiveness prior to the actual test.[2]

Another thing Lykken learned from his readings was that the type of lie detector test used most commonly by police examiners in real-life criminal investigations was very different from the scientifically oriented GKT. The standard method in use for the testing of criminal suspects was the Control Question Test (CQT), a procedure developed in the 1940s by John Reid, an influential figure in the emerging field of forensic polygraphy. Reid was an attorney and interrogative specialist rather than a researcher, and thus his notion of "control" lacked the precision of an experimentalist's. His control questions consisted of decoy items dealing with acts of general wrongdoing, intended to deflect the "psychological set" of innocent examinees away from the relevant questions targeted at the specific incident under investigation. Skillful interactions between the polygrapher and the examinee – directed at influencing the examinee's perceptions of the relative importance of differing test questions – were viewed as critical to the effectiveness of the CQT. In cases where an examinee's behavior and reactions during the test pointed toward guilt, the examination concluded with a post-test interrogation aimed at extracting a confession. Police examiners who used the CQT argued that it yielded confessions in a high proportion of such cases.

the lie detector controversy, this work has social value, political, psychopathological, and even anthropological interest, and can be adversarial enough to serve as the moral equivalent of war."

[2] Early reports documenting use of the card test procedure included Geldreich (1941), Ruckmick (1938), and Van Buskirk and Marcuse (1954).

As a scientist, Lykken was flatly unimpressed with the CQT – characterizing it as a "bloodless third degree." In particular, Lykken pointed out that inherent dissimilarities between the so-called control questions and the crime-relevant test questions constituted a built-in bias against innocent suspects. His concerns about the procedure intensified as real-life cases came to his attention in which innocent individuals had faced prosecution and imprisonment after failing polygraph tests of this type. In a classic paper published in the *American Psychologist*, Lykken (1974) appealed to psychologists in the academic community to become involved in research and debate on polygraphic lie detection in order to challenge existing techniques developed by non-scientists and establish more credible alternatives. A centerpoint of this paper was an argument for the GKT as an alternative, scientifically based approach to detecting deception in specific-incident criminal investigations: "A polygraphic lie detection method known as the Guilty Knowledge Technique appears to have the potential for very high validity in the restricted number of criminal investigations where it is applicable. But the GKT seems to be unknown to professional polygraphers and there have been no studies either of its range of applicability or of its validity in field situations" (p. 738).

Lykken's article proved effective in getting academic researchers interested in lie detection and in utilizing the GKT as an experimental methodology. During the 1960s, several more published reports using the GKT appeared in the literature. These included a number of studies by two research teams, one led by Martin Orne in the United States (e.g., Gustafson and Orne, 1963; Thackray and Orne, 1968) and the other by Sol Kugelmass in Israel (e.g., Kugelmass and Lieblich, 1966; Kugelmass *et al.*, 1967), along with a small number of studies by other investigators (e.g., Davidson, 1968; Kubis, 1962). By the time the first edition of *Tremor in the Blood* was published in 1981, lie detection had developed into an active area of investigation among psychology researchers and the GKT – nowadays called the Concealed Information Test, or CIT – had emerged as its dominant experimental paradigm. However, the test remained largely ignored by field examiners in the United States and it took several more years for the first published report of the field validity of the CIT – by Israeli investigator Eitan Elaad (1990) – to appear. Even now, well into the twenty-first century – at a time where CIT studies routinely incorporate direct measurement of brain response and advanced quantitative approaches to decision-making – the technique is still not used to any significant extent by field examiners in North America and has achieved standard usage only in the nation of Japan (see Chapter 14 of this volume).

What factors account for this curious persisting gap between scientific progress and practical implementation? While a number of factors could be cited (including the broader implementability of the CQT in field cases), the main one is that field examiners who routinely use the CQT in specific-incident investigations believe – despite conceptual arguments (Lykken, 1974, 1981) and scientific evidence (Iacono and Patrick, 2006; Office of Technology Assessment, 1983) to the contrary – that this procedure is virtually infallible (see also Chapter 8 of this volume). The reason for this persisting belief is that the feedback that field examiners receive regarding the accuracy of decisions they make in polygraph test cases is selective – and systematically biased toward affirming the outcome of the test.

This state of affairs was demonstrated in a field study of the CQT by Patrick and Iacono (1991). These investigators followed up all cases tested by the police polygraph unit in a major Canadian city over a five-year period; file records from referring detachments were reviewed to identify instances in which CQT polygraph subjects were verified as either innocent or guilty based on evidence that emerged after the polygraph examination was conducted. A major finding of the study was that virtually all of the feedback police polygraph examiners received regarding their decisions consisted of information stemming from post-test confessions of test subjects – that is, some portion of examinees were verified as guilty by their own confession (either immediately after the polygraph test, or later in the investigation), and others were cleared as innocent based on the confession of some other suspect in the case.

A further striking finding of the study was that the outcome of the polygraph test itself systematically influenced the nature of feedback that examiners received: suspects identified as "deceptive" on the polygraph test were considered guilty whether they confessed or not, leading case investigators to abandon alternative avenues of investigation – and thus opportunities for the outcome of the polygraph to be refuted; subjects identified as "innocent" on the test were normally dismissed as suspects, leading investigators to shift their efforts toward other suspects – again precluding opportunities for the polygraph to be "proven wrong." As a function of these systematic biases, police polygraph examiners almost never received post-examination feedback that disconfirmed their test decisions – perpetuating their view of the CQT as infallible.

Due to systematic biases of this sort, and the barrier they pose to establishing credible estimates of polygraph accuracy in real-life cases, the validity of the CQT remains a matter of ongoing debate – and the CQT continues to be the most common type of test used in

specific-incident investigations by police and other agencies around the world. Nonetheless, alongside the persistant use of the CQT in the field, the CIT has continued to grow and develop as a scientific procedure. During the 1970s and 1980s, substantial research was devoted to examining the impact of various factors on detectability in the CIT, including motivation to appear truthful, innocuous exposure to crime-relevant details, feedback regarding physiological responses to test items, dispositional factors such as anxiousness and extraversion, and drugs of various types including alcohol, anxiolytics, and stimulants. The CIT was well-suited to parametric studies of this kind because of its classic experimental-control format and amenability to standardization. As a function of these attributes, the CIT also gained popularity as an experimental paradigm for studying basic processes contributing to phasic electrodermal response.

An important development in the mid-1980s was the incorporation of brain event-related potential (ERP) measures into the CIT (Farwell and Donchin, 1986; Rosenfeld *et al*., 1987; see Chapter 4 of this volume). Up to this point, most studies of the CIT had utilized electrodermal response as the main dependent measure, with a smaller number of studies including respiratory or cardiovascular indices. The use of ERP measures to detect guilty knowledge was a natural progression because it had long been known that the P300 component of the ERP in particular is sensitive to the familiarity or meaningfulness of stimulus events. The use of P300-based CIT procedures including multiple iterative stimulus presentations opened the door to sophisticated quantitative approaches to classifying individuals as truthful or deceptive – including use of "bootstrap" resampling (Farwell and Donchin, 1991) and Bayesian statistical methods (Allen *et al*., 1992). Brain-based CIT procedures have also been used to elucidate specific cognitive processes underlying deception; studies of this sort (e.g., Johnson *et al*., 2003) have utilized differing ERP components to index putatively distinct operations associated with efforts to deceive. A related exciting development is the recent use of functional neuroimaging methods to investigate deception-related processes in the CIT detection context (see Chapter 5 of this volume). The first published study of this kind was by Spence *et al*. (2001).

Another important development over the past two decades has been the increasing documented use of CIT procedures in field settings (see Chapter 9 of this volume). Although the CIT was used before this time in Japan on a routine basis in police investigations (Fukumoto, 1980; Nakayama, 2002; Yamamura and Miyata, 1990), as noted earlier, the first published study of the field validity of the CIT – by Elaad – did

not appear until 1990. A second study of this type was published by Elaad and his colleagues in 1992, and this was followed a year later by a report of the field validity of a P300-based CIT procedure by researchers in Japan (Miyake *et al.*, 1993). Although these reports raised some concerns about the accuracy of the CIT with guilty suspects – perhaps owing to the lesser saliency or certitude of crime-relevant details in real-life cases – they nonetheless encouraged further implementation and evaluation of the procedure as an alternative to the CQT in field settings.

The current volume, the first to be devoted entirely to the CIT, is a testament to the ongoing growth and development of this scientifically oriented approach to detecting deception. The current volume highlights the impressive body of literature that has emerged using this paradigm in the half-century since Lykken's (1959, 1960) classic studies – including experimental research addressing basic topics such as orienting and habituation, memory, factors contributing to physiological reactivity, and processes underlying deception, as well as applied research evaluating the general validity of the technique and various factors affecting detectability. Following an initial chapter (by Iacono) that presents a compelling case for widespread adoption of the CIT by law enforcement agencies, Chapters 2 through 6 address varying response parameters that have been utilized in studies of the CIT (from electrodermal activity to functional brain response), with Chapter 7 providing a detailed theoretical analysis of processes underlying physiological reactivity to items on the test. Chapters 8–12 address issues pertaining to implementation of the CIT in field settings, including limits to its range of applicability, constraints on generalizability of laboratory findings to real-life contexts, the problem of information leakage prior to testing, and potential moderators of test accuracy in real-life cases (use of countermeasures; psychopathic tendencies). Chapters 13–14 focus on applications of the CIT to the assessment of memory function in clinical cases as well as to the evaluation of criminal suspects in police investigations. Chapter 15 addresses legal issues pertaining to field use of the CIT, and Chapter 16 describes recommended procedures for constructing and administering the CIT in field settings.

Victor Hugo (loosely translated) observed that "there is no army so powerful as an idea whose time has come." The CIT is an idea whose time has come. It has proven highly generative as an experimental paradigm, its practical potential has been demonstrated through a growing list of documented uses in real-life settings, and longstanding barriers to its widespread implementation in the field (including allegiance to the CQT and unawareness of or unwillingness to try alternatives)

appear to be crumbling. In this regard, the current volume represents an important and welcome addition to the literature – one that seems likely to promote increasing reliance on scientific concepts and methods in efforts to distinguish between truth and deception in practical contexts. This is a book David Lykken would be pleased to see side by side on the shelf with his classic *Tremor in the Blood*.

## REFERENCES

Allen, J. J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology*, 29, 504–522.

Davidson, P. O. (1968). Validity of the guilty knowledge technique: the effect of motivation. *Journal of Applied Psychology*, 52, 62–65.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, 75, 521–529.

Farwell, L. A., and Donchin, E. (1986). The "brain detector:" P300 in the detection of deception. *Psychophysiology*, 24, 434.

(1991). The truth will out: interrogative polygraphy ("lie detection") with event related brain potentials. *Psychophysiology*, 28, 531–547.

Fukomoto, J. (1980). A case in which the polygraph was the sole evidence for conviction. *Polygraph*, 9, 42–44.

Geldreich, E. W. (1941). Studies of the galvanic skin response as a deception indicator. *Transactions of the Kansas Academy of Science*, 44, 346–351.

Gustafson, L. A., and Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology*, 47, 408–411.

Iacono, W. G., and Patrick, C. J. (2006). Polygraph ("lie detector") testing: current status and emerging trends. In I. B. Weiner and A. Hess (eds.), *Handbook of Forensic Psychology*, 3rd edn. (pp. 552–588). New York: Wiley.

Johnson, R., Jr., Barnhardt, J., and Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology*, 64, 217–253.

Kubis, J. F. (1962). *Studies in Lie Detection: Computer Feasibility Considerations*. RADC-TR 62–205, Contract AF 30(602)-2270. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Kugelmass, S., and Lieblich, I. (1966). The effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, 50, 211–216.

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

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

(1960). The validity of the guilty knowledge test: the effects of faking. *Journal of Applied Psychology*, 44, 258–262.

(1974). Psychology and the lie detector industry. *American Psychologist*, 29, 725–739.
(1981). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, 1st edn. (2nd edn., 1998). New York: McGraw-Hill.
Miyake, Y., Mizutani, M., and Yamahura, T. (1993). Event related potentials as an indicator of detecting information in field polygraph examinations. *Polygraph*, 22, 131–149.
Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (ed.), *Handbook of Polygraph Testing* (pp. 49–86). San Diego: Academic Press.
Office of Technology Assessment (1983). Scientific validity of polygraph testing: a research review and evaluation. Washington, DC: Office of Technology Assessment.
Patrick, C. J., and Iacono, W. G. (1991). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology*, 28, 632–638.
Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., and Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: a new method of lie detection. *Polygraph*, 16, 258–263.
Ruckmick, C. A. (1938). The truth about the lie detector. *Journal of Applied Psychology*, 22, 50–58.
Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, 12, 2433–2438.
Thackray, R. I., and Orne, M. T. (1968). A comparison of physiological indices in detection of deception. *Psychophysiology*, 4, 329–339.
Van Buskirk, D., and Marcuse, F. L. (1954). The nature of errors in experimental lie detection. *Journal of Experimental Psychology*, 47, 187–190.
Yamamura, T., and Miyata, Y. (1990). Development of the polygraph technique in Japan for detection of deception. *Forensic Science International*, 44, 257–271.