

**Reports from the Research Laboratories
of the
Department of Psychiatry
University of Minnesota**

Psychology and the Lie Detector Industry

DAVID T. LYKKEN

University of Minnesota

MSDM
P95
QR311r
74-1

Report Number PR-74-1

January 25, 1974

MSDM

P95

gR31/r

74-1

PSYCHOLOGY AND THE LIE DETECTOR INDUSTRY

David T. Lykken

University of Minnesota

PSYCHOLOGY AND THE LIE DETECTOR INDUSTRY

David T. Lykken
University of Minnesota

Polygraphic Interrogation - the use of the so-called "lie detector" - is already a thriving industry in the United States and it is growing rapidly. Several million polygraphic examinations are conducted annually in this country by more than 3,000 professional polygraphers, most of whom are engaged in the private practice of their art, and some dozen schools, including one operated by the United States Army, are spawning graduates. This great crescendo in what for years had been a rather muted theme resulted, like most such booms, from newly discovered possibilities of profit. Estimates of losses due to employee theft in American business are running as high as \$6 billion annually; thus, a rich and eager market beckons the entrepreneur who claims to have a quick and relatively inexpensive method of detecting peculators. During the 1960's, the Federal Government invested large sums in a search for a covert lie detector, i.e., a method of measuring physiological arousal in a subject who is unaware that his autonomic reactions are being monitored. This hope has now been realized with the development of the "Psychological Stress Evaluator", a device which can detect emotional modulations of the speaking voice and thus can be applied even to tape recordings of telephone conversations and other "bugs". Movements are afoot in most of the 50 states to license polygraphers, legitimating the profession by statutory fiat. Should the hopes of some of the leaders of this profession come to pass, polygraphic interrogation may soon become the most important area of applied psychology, both economically and in terms of social impact.

For it is clear that polygraphic interrogation is an area of applied psychology, since the lie detector clearly is a psychological test. If persons

who administer and evaluate Stanford-Binets or Rorschachs or MMPIs are psychometrists, then the polygrapher is a psychometrist also and one basic science that should underlie his art is the science of psychological assessment. Since the polygraph test involves the study of autonomic (rather than verbal or other operant) responses to psychological stimuli, a second basic science area which is directly relevant to polygraphy is psychophysiology. One might therefore suppose that most polygraphers are trained psychologists - or at least that most polygraphers are trained by psychologists - or at the very least that a sufficient number of psychologists keep in touch with developments in the polygraphy business to provide expert guidance, e.g., to courts or to legislative bodies. In fact, however, the field of polygraphic interrogation is about as well integrated into psychology as is the field of chiropractic into medicine; few professional polygraphers have any significant psychological training and nearly all will have received their training in polygraphy from teachers whose own prior professional background, if any, was in law enforcement. If bills were to be introduced tomorrow simultaneously in the legislatures of the 50 states, limiting, regulating or expanding the use of the lie detector, it is doubtful that more than some half-dozen of those legislatures could find within the borders of the state a certified psychologist competent to offer expert testimony on the matter. This article is motivated by the conviction that such a situation is a kind of scandal, that psychologists ought to know about the lie detector since its use or misuse has social consequences potentially more important than anything most psychologists do know about qua psychologists, and that the "truth about the lie detector" is a matter of considerable intrinsic interest anyway.

A Brief History of Lie Detection

Recalling the ingenuity with which many lower animals, without benefit of

speech, still manage to deceive their pursuers -- or their prey -- it seems a safe assumption that man learned to lie not long after he first learned to talk, and, no doubt, the earliest attempts at lie detection occurred while lying itself was still in the stage of research and development. The history of the methods used by the ancients for determining guilt or detecting deception is extensive and fascinating (Trovillo, 1939).

In Asia Minor, a man suspected of a crime was given a superficial knife wound on his arm before being allowed to relate his alibi. If, after telling his story, the wound had stopped bleeding he was adjudged innocent; if blood still flowed, he was guilty. A similar example from modern Africa is the ordeal of the boiling water. Each suspect in turn is required to dip his arm first in a jar of cold water and then immediately into a pot of water vigorously boiling. It is said that on the following day the guilty person alone among the suspects will show blistering and peeling of the skin of that arm.

A method worth the attention of modern students of divorce was employed in Germany during the middle ages for cases of marital discord in which one partner accused the other of infidelity. The man was required to stand in a tub sunken into the ground with his left arm bound behind him and with a short club in his right hand. The woman was clothed only in a short chemise having one sleeve longer than her arm and in this sleeve was sewn a rock. The wife danced about the tub attempting to hit the husband with the stone while at the same time jumping back before he could reach her with his club. Since it is not recorded how the outcome of this contest was supposed to settle the question of unfaithfulness, it may be supposed that the method was intended rather to discourage litigation altogether.

One of the oldest techniques of which we know was the Hindu rice chewing test employed in ancient India. The suspect was required to chew a mouthful of

rice. If he could then spit out the rice he was declared innocent but if the rice instead stuck to the lining of his mouth and tongue he was considered guilty. The principle of this test has much in common with modern techniques for detection of deception. In most individuals, the act of lying or the experience of being accused of a crime of which they are guilty may produce sufficient emotional reaction and attendant excitement of the sympathetic division of the autonomic nervous system to inhibit salivation. Thus, the guilty person finds the rice inclined to stick to his dry mouth. Unhappily, it must be supposed that many innocent persons were sufficiently frightened to find themselves accused so that they too experienced a dry mouth and failed the test. It was perhaps to obviate any such injustice that, when this test was adapted to the uses of the Inquisition during the Middle Ages as an ordeal reserved for testing the veracity of the clergy themselves, it was modified in the following way. Instead of rice, a piece of barley bread and a piece of cheese were laid upon the altar in front of the suspect priest. After fervent prayer that, should he be guilty of the offense charged to him, God should send His angel Gabriel to stop his throat, the priest then attempted to eat the bread and cheese. According to Mackay ("Memoirs of Extraordinary Popular Delusions," 1852, page 266), "There is no instance upon record of a priest having been choked in this manner."

Apparently the first known instance of direct monitoring of autonomic reaction to a stimulus is attributed to Erasistratus, a Greek physician of 250 BC. Nicator, King of Syria, being concerned over the languishing illness of his son, Antiochus, summoned Erasistratus to the court to treat him. Acting upon a court rumor that the source of Antiochus' malaise might be a consuming and illicit passion for his beautiful young stepmother, Stratonice, Erasistratus undertook

to monitor Antiochus' pulse while conversing with him about various members of the court. Noting its tumultuous rhythm each time the name of Stratonice was brought into the conversation, Erasistratus considered his hypothesis confirmed. The fact that a young princess, also named Stratonice, was subsequently begotten of an intimacy between Antiochus and the queen appears to corroborate this finding.

Thus, the fact that pulse rate, pulse pressure, and other such physiological indicators could involuntarily betray the presence of fear or emotion was apprehended early in man's history. Similarly, the idea of using such indicators for the detection of guilt or of deception cannot be surely attributed to any single historical figure. Therefore, the "lie detector" cannot be said to have been an invention of the modern era. However, its systematic elaboration and development and its widespread application to police work is a twentieth century phenomenon predicated upon technological developments which made possible the relatively accurate measurement and recording of heart rate, pulse volume, blood pressure, respiration, electrodermal phenomena, and so on. Lombroso made occasional use of the plethysmograph and the hydrosphygmomanometer in monitoring pulse patterns and blood pressure changes while questioning criminal suspects. The Harvard psychologist, Munsterberg, explored these and other methods of investigating the credibility of testimony and describes his work in an influential book first published in 1908. One of Munsterberg's students, W.M. Marston, was a vigorous exponent of the use of physiological measures in the detection of deception and has been referred to by some as "the father of the modern lie detector test." Marston was particularly interested in blood pressure variation and claimed to have shown that the act of lying is invariably accompanied by sharp increases in systolic blood pressure, a response which he said is not shown

by the frightened but innocent subject when answering truthfully. These observations have not since been corroborated.

Marston's early publications probably provided the major impetus to the application of polygraphic interrogation in professional police work, largely through their impact in 1921 upon a medical student J.A. Larson, who was then in the part time employ of Chief Vollmer of the Berkeley, California, Police Department. Larson put together a portable polygraphic apparatus and proceeded to make ingenious use of it while enjoying the full cooperation of the Berkeley police. Larson described his work extensively in the criminological literature and his ideas and developments and particularly his reports of success in the field were influential. Most of the modern professional polygraphers trace their interest and early training to Larson or Vollmer as a source. Leonard Keeler began as a high school student employee of Larson's at Berkeley from whence he moved to the Scientific Crime Detection Laboratory of Northwestern University School of Law, subsequently a unit of the Chicago Police Department. Keeler contributed a number of improvements in polygraphic instrumentation and interrogation technique and was active in the training of police officers in polygraphic procedures. C.E. Lee, a Captain of Detectives at the Berkeley Police Department, developed a portable polygraph known as the Berkeley Psychograph and authored an influential text book in the area (Lee, 1953). Two of Keeler's associates, Fred Inbau and John Reid, also produced a well known text (Reid and Inbau, 1966).

Methods of Polygraphic Interrogation

There are two fundamentally different methods of polygraphic interrogation, one that is properly called "lie detection" and another which I shall refer to as the "guilty knowledge test" (GKT). Since this basic and important distinction

is for some reason not generally recognized, it will be elaborated with some care here at the outset. There is no great mystery about the first category; any interrogation technique which involves asking the subject one or more relevant questions ("Did you kill Cock Robin?") and then attempting to determine by some method or combination of methods whether his answer to that question is or is not deceptive can be classified as a method of lie detection. There are a number of lie detection methods which differ in the nature of the questions used, the manner in which the questions are put to the subject, the sorts of data included in the analysis, and in the methods of analysis employed. It is probable that more than 90 percent of all polygraphic examinations given in the field employ lie detection methods exclusively. Guilty knowledge detection, as the name implies, is intended to determine whether the subject is aware of certain information; the guilty knowledge test (GKT) might be used whenever the demonstration of such awareness might, e.g. differentiate between a guilty and an innocent suspect.

To illustrate both of these methods in the context of criminal investigation, I shall make use of a notorious case that occurred some years ago in New York City. The bass-player in the band at the Stork Club was arrested on the charge of armed robbery, accused of having held-up a number of Manhattan liquor stores and a loan company. The prosecution's case was based entirely on the fact that each of the victims positively identified the musician as the man who had threatened them with a gun and taken their money. After the usual protracted delays, the case went to trial, but before that ordeal was finished the sudden death of the judge resulted in a mistrial. After further delay and a second trial, the musician was found guilty as charged. Just before he was to be transferred to the state prison, another man, remarkably similar to the bass-player

In physical appearance, was arrested for armed robbery and volunteered a confession which exonerated the musician. By this time, more than two years after his initial arrest, that unfortunate victim of circumstance had lost his job, his home, his life savings and even his wife, who had succumbed to the strain and been committed to a state hospital.

Let us suppose that our protagonist had been given a polygraphic interrogation by the New York City police soon after his arrest. Had this happened, the method employed would almost certainly have been some form of lie detector test; that is, he would have been asked "critical" questions of the form, "Did you rob the Friendly Loan Company?", and the autonomic responses accompanying his "No" answer would have been compared with his responses to other, "irrelevant" or "emotional control" questions. Based on this comparison, and on the examiner's observation of the suspect during the examination, and on whatever the examiner happened to know about the facts of the case, the polygrapher would have arrived at a global assessment: "deception indicated" or "deception not indicated" or "indeterminate". We shall examine these lie detection methods in more detail later, including the assumptions upon which they are based and the evidence that exists concerning their validity. For the moment, let's assume that the polygrapher happened to arrive at the correct diagnosis in this instance, that the musician was telling the truth and was in fact innocent. It is possible that such a finding might have led the investigating officers to re-examine their evidence, perhaps to make a greater effort to corroborate whatever alibies the suspect might have been able to produce for the times in question. But in this particular case, with such an array of positive eye-witness identifications, it seems probable that the bass-player's fate would not have been greatly altered by the lie detector test.

The Guilty Knowledge Test

Suppose, however, that the polygraphic examiner had been more sophisticated than most and had recognized that this case was one which lent itself nicely to the use of the guilty knowledge test. At the cost of some initial effort and preparation, and with the cooperation of the investigating officers, the examiner might have prepared a test along the following lines:

Item 1. "The man we're looking for held-up a loan office in Manhattan. If you're the guilty party, you will recognize the name of that loan company. I'm going to name a few loan companies that have offices in the vicinity; you just sit there quietly and repeat the names after me as you hear them. Was it the Ideal Loan Co.? ... Was it the Continental Loan Co.? ... Was it the Guarantee Loan Co.? ... Was it the Friendly Loan Co.? ... Was it the Fidelity Loan Co.?"

If the subject is in fact guilty of the robbery and if he did notice the name of the company he robbed, then it is a safe assumption that he will show a greater autonomic response to the guilty alternative ("Friendly Loan Co.?") than he would if he did not possess this guilty knowledge. If we can further assume that he cannot deliberately but covertly produce an augmented response to one of the other alternatives, in order to mislead the examiner, and also that he has no other reason to respond selectively to one of these others, then we can expect that his response to the guilty alternative will be stronger at least than his mean response to the array of alternatives and probably that it will be strongest of all. On the other hand, if the subject is in fact innocent and if he has no indirect means of knowing or inferring which alternative is "correct", then we can safely suppose that there is only about one chance in five that his largest autonomic response will be to the guilty alternative.

Item 2. "Before showing his gun, the robber pretended that he wanted to take out a loan for a certain purpose. If you're the guilty man, you will know whether that purpose was to buy a car, to pay doctor bills, to pay for a vacation trip, to buy a color TV, or to get a present for his wife. I'm going to name each of these five possibilities in order and I want you to sit quietly and just repeat what I say. Was it - a car? ..." etc.

If the subject had been guilty, it is very likely that he would have remembered talking to the woman at the loan office about needing money to pay his "doctor bills" and therefore that his autonomic response to this one of the five alternatives would be augmented by this guilty knowledge. On the other hand, the probability that an innocent suspect would give his largest response to this alternative is only about 0.2. Similarly, the probability that an innocent suspect might happen to respond most strongly to the "correct" alternative in both Item 1 and Item 2 would be equal to the product of the individual item probabilities or about 0.04.

Item 3. "If you did rob this loan company, you will recognize the woman you talked to, the one who gave you the money. I have some photographs here of women tellers in five different loan offices, just as they look to a customer. I will show these pictures to you one at a time. Just sit there quietly and look at each photograph."

The probability that an innocent suspect might chance to "hit" on all three items would be approximately $(1/5)^3 = 0.008$. Since subjects tend to respond more strongly to the first alternative in any list, the examiner might make a practice of using an extra "incorrect" alternative at the start of each set and then ignoring the response to that alternative. With only 10 such guilty knowledge

items, each with 5 scorable alternatives, there would be only about 1 chance in 10 million that a subject without guilty knowledge would give his largest response to the 'correct' alternative in all ten items. With a 10 item test, the actual culprit would only have to "hit" on 6 of the 10 items to permit us to say that the chances are less than 1 in 1000 that he is innocent.

In the case of the misidentified musician, it would have been a relatively simple matter to construct several tests of 10 items each, taking advantage of the fact that here one could be sure that the guilty suspect would possess a great deal of guilty knowledge relating to details of the several crimes of which he was accused, details which would not be known to an innocent suspect but which could be determined with a little effort and competent investigation and planning prior to the examination. Moreover, of course, each set of items could be used more than once; the probability that the "correct" alternative would elicit the largest response from an innocent subject on both of two separate testings is doubtless greater than $(1/5)^2$ - since that alternative might have been especially provocative for him on both occasions even without guilty knowledge - but the probability that the first test and the retest will both be false-positive is certainly smaller than 0.2. Therefore, after testing our unfortunate bass-player, the examiner would have been able to inform the prosecution that the chances of this man's being guilty were vanishingly small and that the police should be deployed again to seek some other suspect fitting his general description.

Assumptions and Limitations of the Guilty Knowledge Test

The guilty knowledge method requires that the examiner can determine a number of facts which only a guilty subject will be able to recognize (these facts can be perfectly trivial matters which would not appear in newspaper accounts) and that he can present these facts in the form of multiple-choice items, embedded in a set

of 3 or 4 or 5 alternatives that would seem equally plausible to an innocent subject without guilty knowledge. The basic assumption of the GKT is that the guilty subject will show stronger autonomic response to what he recognizes as the significant alternative than he would have shown without such guilty knowledge. The amplitude of the involuntary autonomic responses to the significant alternative has little meaning by itself; a hyper-reactive subject might respond strongly to that alternative without knowing that it was the 'correct' one, while a hypo-reactive suspect might give a small response even though he does have guilty knowledge. But the same subject's responses to the other, plausible but incorrect, alternatives of the GKT provide a nearly ideal control against which to evaluate his response to the significant alternative. In the language of psychophysiology, all of the GKT alternatives can be expected to produce orienting reflexes (ORs) which will vary in amplitude from subject to subject for a number of reasons, of which guilt is only one. However, for the guilty subject only, the 'correct' alternative will have a special significance, an added "signal value" (Berlyne, 1960), which will tend to produce a stronger OR than that subject will show to the other alternatives. Whether he is high or low in reactivity, whether he has confidence in the test or not, whether he is frightened and aroused or calm and indifferent, we can still expect that his response to this significant alternative will be stronger than to the other alternatives as long as he recognizes which alternative is 'correct'. Similarly, if he has no guilty knowledge, neither his reactivity, his present emotional state nor his confidence in the validity of the test can act to influence his GKT score. For all subjects who are without guilty knowledge, the probability of the 'correct' alternative producing the largest response on each of the N test items will be $(1/K)^N$, where K is the number of alternatives per item.

Finally, since the guilty knowledge method does not require that the items be in the form of questions to which the subject must give answers - i.e., since he can be asked merely to repeat the alternative or, even, to simply sit and listen - the interesting question arises as to whether requiring a criminal suspect to submit to a GKT would involve a violation of the Fifth Amendment rights, any more than would requiring him to show his face to a witness or to submit to fingerprinting.

On the other hand, the guilty knowledge method simply cannot be used in many situations where the lie detector is now used and it almost always will require much more careful preparation and pre-investigation than does a lie detector test. Moreover, even where it is appropriate, it is obvious that the construction of a good multiple-choice guilty knowledge test requires considerable ingenuity and psychological sensitivity. Where it is applicable - where the requisite guilty knowledge information is available to a competent examiner - there can be little doubt that the guilty knowledge method could be a powerful tool in criminal investigation and that its greatest benefit might be to protect innocent suspects from prolonged police harrassment and even false imprisonment.

But it must be emphasized that the guilty knowledge method is not in fact employed by professional polygraphers, partly because the distinctive character of the method has not yet been understood. Even the two recent reviews written by psychophysicologists (Barland & Raskin, 1973; Orne, Thackray, & Paskewitz, 1972), fail to distinguish between lie detection and guilty knowledge methods with adequate clarity. One problem here is a tendency to regard the guilty knowledge method as just a variant of something known to polygraphers as the "peak of tension test" (POT). As it is commonly described in polygraphy literature

(e.g., Reid & Inbau, 1966), the POT would involve presenting a series of questions, only one of which is expected to elicit a lie from a guilty subject, after first showing the subject what the sequence of questions is to be. The test gets its name from the notion that a guilty suspect will show increasing autonomic arousal in anticipation of the critical question, a "peak of tension" when that question is answered deceptively, and a decline thereafter. But this is nothing more than a standard lie detector test embellished by the opportunity to look for a predictable trend or peak in tonic level of autonomic function in addition to whatever information is provided by the phasic autonomic responses to the individual questions.

As it is described by other authorities (e.g., Barland & Raskin, 1973), the POT resembles a single-item Guilty Knowledge Test. Even there, however, there seems to be no clear awareness that it is a very different thing to use the polygraph to determine whether the subject can identify the significant alternative, than to use autonomic arousal or "tension" as evidence that the subject is lying.

With a very few exceptions (e.g., Lykken, 1959, 1960; Davidson, 1968; Shakhar, Liebllich & Kugelmas, 1970), the literature on polygraphic interrogation deals exclusively with lie detection methods. The professional practice of polygraphic interrogation employs lie detection methods almost without exception and it should be clearly understood that the guilty knowledge method could not be used in the vast majority of situations where the lie detector is now used - many criminal investigations and all employee screening applications - because of a lack of the guilty knowledge information necessary to construct the item set. Therefore, the several virtues of the guilty knowledge method - the fact that it rests upon reasonable assumptions, that it can produce an objective,

quantitative estimate of the probability of guilt, and that in certain situations it is capable of yielding near-perfect validity - cannot be marshalled in support of current practices.

The Lie Detector Test

The usual field polygraph makes continuous recordings of three channels of physiological data. The electrodermal channel displays changes in palmar skin resistance or GSRs, usually by means of a capacity-coupled bridge circuit which results in the loss of any information about tonic resistance levels. The "cardio" channel is actually an arm plethysmograph which records changes in upper-arm volume associated with the cardiac cycle. From this channel, one can determine heartrate and some indication of changes in pulse volume. The third channel is driven pneumatically or electrically from an expansible belt around the subject's chest and records respiration. The field polygraph as manufactured for the lie detector trade is approximately 20 years behind the state of the art as represented by the modern psychophysiological laboratory. While not very important in itself since the problems with lie detection devolve from logical and psychological considerations rather than from questions of psychophysiological technique, this obsolescent character of field polygraph design is symptomatic of the fact that polygraphic interrogation has developed in virtual isolation from the original parent discipline of psychophysiology.

The "Psychological Stress Evaluator", mentioned earlier, adds a fourth channel which records the presence of a low-frequency component present in normal speech sounds but said to diminish under emotional stress. Assuming that these claims can be independently corroborated, the PSE represents an ingenious technological advance having the unique advantage of being usable without the subject's knowledge. But the questions to be raised here about the logic and validity of the lie detector are not concerned with the nature of the response

being monitored and would seem to apply equally to the PSE as to the traditional polygraph.

The standard lie detector test is normally preceded by a pre-test interview in which the list of questions to be used is gone over with the subject, to be sure that he understands the wording and that he feels he can answer them truthfully and unambiguously with a simple "Yes" or "No". Another purpose of the pre-test interview is to indirectly convince the subject that the lie detector really works, that he has nothing to fear if he is innocent and truthful but that any attempt at deception will be easily detected. After the preliminary conversation, the subject is seated in the examining room and the transducers are attached to his hand, arm and chest. The agreed-upon list of questions is presented, usually two or three times, and the examiner may engage in some discussion with the subject between testings in order to insure that the questions are unambiguous and that, e.g., the subject is not responding to a critical question for some unsuspected but irrelevant reason. Although I have not seen it recommended in published manuals, one may suppose that professional examiners frequently suggest to subjects that their attempts to deceive have in fact been detected (whether they have or have not) with the intent of precipitating a confession from a guilty subject. It is a fact that the lie detector functions effectively as a 'painless third degree' and that the confessions of guilt elicited in this situation might themselves provide sufficient economic justification for its use by the police, even if the technique were wholly invalid and the polygraph merely a stage-prop.

The question list used will consist of from 5 to 10 questions that can be answered "Yes" or "No". It will include one or more "critical" or "relevant" questions of the form, "Did you fire the bullet that hit Jones?" and one or more

"irrelevant" questions pertaining to unrelated and unexciting matters, e.g., "Are you sitting down?". Most modern polygraphers also include several "control" questions which are intended to serve as "emotional standards". The control question should be unrelated to the matter under investigation and it is expected that the subject will answer it truthfully; however, the control question is chosen with the intention that it will elicit an emotional response from the subject, preferably a response involving an attitude of guilt; e.g., "Can you remember ever stealing anything before you were 18 years old?". Finally, some examiners try to include a "guilt complex" question, e.g., a question relating to some other real or imaginary crime of which the subject is innocent.

Methods of evaluating lie detector tests to determine whether deception is or is not indicated - i.e., whether the test has been 'passed' or 'failed' - vary from one examiner to another. Some of the early workers maintained the concept of a specific lie response, i.e., the notion that there exists some unique pattern of autonomic response that is manifested by all individuals when they are deliberately lying but not when they are answering truthfully, even though fearfully. Thus, Benussi (1914) claimed that the respiratory ratio of expiration to inspiration increases after lying but decreases after a truthful answer. Marston (1938), believed that an increase in systolic blood pressure was certain evidence of lying. Summers (1937) contended that the arousal produced by the critical questions diminishes (habituates) with repetition with an innocent subject but not if the subject is lying. All that we know about individual differences in both psychological and physiological response tendencies argues against the possibility that there is such a specific lie response and, since no acceptable evidence has been forthcoming in support of these early claims, such methods of analysis have fallen into disuse.

The modern polygrapher simply looks for evidence of autonomic disturbance associated with the answer to the critical questions, disturbance that is more intense or persistent than that associated with the irrelevant questions and, especially, the emotional control questions. Most examiners make a global evaluation, without specific measurement or scoring. At least one of the schools now training polygraphic examiners teaches the "Zone of Comparison" technique (Barland & Raskin, 1973) in which a numerical score is derived from specified comparisons of responses to the several types of question. A high score results from a record in which the autonomic perturbation associated with the critical questions is much more persistent and intense than that associated with the control questions and the examiner is advised to classify tests that give intermediate scores as "inconclusive".

A point that must be emphasized, however, is that the professional polygrapher almost never arrives at his final diagnosis on the basis of the polygraphic records alone; the examiner, rather than the polygraph, is the actual "lie detector". Laboratory studies often employ "blind reading" of charts by judges who know nothing else about the subject or the circumstances and a similar practice is occasionally followed in the field. But, in the vast majority of field examinations, the final diagnosis results from a subjective blending in the mind of the examiner of what he has observed in the charts, in the demeanor of the subject during the test and in the pre-examination interview, what he knows of the evidence against the suspect and what he may infer from the suspect's prior history, and even any prejudices he may hold about the subject's race, age, appearance and the like. Thus, it is a fact that evidence which might be collected relating to the validity of diagnoses arrived at in this standard manner can only be interpreted as evidence relating to the performance of the particular

examiner studied and is almost completely uninformative concerning the objective contribution of the polygraphic data.

Assumptions of the (Autonomic) Lie Detector Test

As we have seen, the basic assumption of the guilty knowledge test is straight-forward and reasonable; given an array of stimuli, X, Y and Z, a guilty suspect who recognizes that Z is related to the crime in question will tend to show a different autonomic response - a stronger orienting reflex - to Z than to X or Y. The assumptions underlying the lie detector test are considerably less straightforward and not nearly so easily accepted. First of all, most professionally administered lie tests are influenced to an unknown degree by the examiner's impressions of the subject and the available evidence, interacting with his evaluation of the actual polygraph record. For present purposes, we shall confine our attention to the latter, to what might be called the autonomic lie detector test, in order to evaluate the assumptions involved in a lie test scored by someone unfamiliar with the accused or with the case against him, working from the polygraphic data solely.

It is a characteristic of the lie test that the 'critical question' (e.g., "Did you do it?") is obviously 'relevant' to all subjects, guilty and innocent alike. It will therefore tend to elicit an emotional response, and the attendant autonomic reaction, in guilty and innocent alike. It is perfectly reasonable to suppose that being guilty and attempting to deceive will tend to add an increment to the autonomic response of guilty, as compared to innocent, subjects. Thus, one might readily imagine that the response to the critical question of 1000 guilty suspects will tend, on the average, to be larger than the mean response of 1000 innocent suspects. However, it is also clear that the magnitude of this response will not be determined solely by the fact of guilt or innocence. Other

things being equal, a highly reactive or labile subject will respond more strongly to the same stimulus than will a subject whose lability is low; the critical response magnitude depends partly on individual lability, L, independently of guilt or innocence. Similarly, individuals will differ in their emotional attitude toward the same situation, in their fear of the consequences of being found guilty. Thus, for some people, the prospect of being tried, convicted and punished for the crime in question may seem catastrophic while for others, e.g., those who have less to lose in such an eventuality, finding themselves in the role of a criminal suspect may be much less fear-inducing. Individual differences in fear of the consequences, F, will therefore also help to determine the critical response, independently of guilt or innocence. Finally, another important determinant will be the subject's confidence in the validity of the test procedure. If he is perfectly confident that the test will yield the correct result (i.e. if C = 1.0), then the guilty suspect will tend to show his maximum response because he "knows" that his guilt is about to be demonstrated. An innocent subject with perfect confidence, on the other hand, will tend to show a minimum response; no matter how labile he may be nor how abhorrent conviction and punishment might seem to him to be, he "knows" that he is about to be exonerated and he responds to the critical question calmly and without great apprehension. The pre-examination interview of the standard lie-test is designed in part to inspire such confidence in the subject; the extent to which this works - the extent to which C approaches unity at the time the critical question is presented - will obviously also vary independently of the fact of guilt or innocence.

These three variables, extraneous to guilt or innocence, which however also interact to determine the magnitude of response to the 'critical' lie test question can be conveniently summarized in a formula; for clarity, we shall use separate

formulas to represent the response of Mr. I, who is innocent and telling the truth, and the response of Mr. G, who is guilty and lying.

$$(1) \quad R_i = L_i F_i \left(\frac{1-C}{2} i \right) + M_i$$

$$(2) \quad R_g = L_g F_g \left(\frac{1+C}{2} g \right) + M_g$$

where: R_i is the critical response of an innocent suspect, Mr. I;
 R_g is the critical response of a guilty suspect, Mr. G;
 L is the subject's autonomic reactivity or lability;
 F is the subject's fear of the consequences of being found guilty;
 C is the subject's confidence in the validity of the lie test ($-1 < C < +1$);
and M is the minimum orienting response which any stimulus might elicit in that subject.

(While these formulæ will be helpful in summarizing the ensuing argument, readers who tend to go faint at the sight of an equation will find that they can ignore the algebra and still be able to follow the discussion.) Note in the two equations above that the variables L and F function multiplicatively; if either the subject's lability or his fear of the consequences happens to be zero, then his response to the critical question will be a minimum (M) irrespective of his degree of confidence in the test or of whether he is guilty. Note also that C , the measure of degree of confidence, functions differently for a guilty than for an innocent subject. If $C = 1.0$, an innocent suspect will give a minimum response ($R_i = M_i$), while a guilty suspect, under these conditions, will give his maximum response ($R_g = L_g F_g + M_g$). That is, Formulas (1) and (2) merely summarize the common-sense ideas already advanced and contain no appreciable 'surplus meaning' deriving from the algebra.

Since a subject's response to the critical question in the lie test can be affected by (is a function of) several factors which will vary from subject to subject independently of who is guilty and who is not, it is clearly not possible to say with any confidence that a critical response larger than some specified magnitude is certain evidence of guilt. A truthful, innocent suspect who happens to be unusually labile and who feels that being prosecuted for the crime in question might ruin his career could obviously give a stronger emotional and autonomic response to the question, "Are you guilty?", than a psychopathic guilty suspect whose lability is low and who doesn't really care whether he 'fails' the test or not. The absolute magnitude of the critical response could provide a basis for a high-validity lie test only if one could be sure that all subjects had nearly perfect confidence ($C = 1.0$) that the test was valid (because, e.g., truthful subjects would then all show just the minimum response). On the other hand, when confidence is zero - i.e., when subjects believe that their passing or failing will be determined as by flipping a coin - then the formulas agree with common sense in indicating that the actual validity of the test will then be zero.

To get a feeling for this last point, it may be useful to imagine one's self in the following situation. You have been arrested on a charge of murder and are administered a lie test by computer. The computer asks, "Are you guilty?", you reply, "No!", and the computer then informs the police either that you are guilty, in which case you will be prosecuted, or that you are innocent, in which event you will be at once released. You happen to know that the computer's verdict is determined by random numbers and that you have a 50:50 chance of being diagnosed "guilty" irrespective of how you respond to the critical question. Under such circumstances, where $C = 0$, you will respond strongly to the critical question as you would to any stimulus presented under such conditions of

apprehension and high arousal and there is no reason to suppose that your response would be any different if you are innocent or guilty since the risk and the stakes are the same in both cases. It is a curious property of the lie detector test that the test's validity is directly related to the subject's confidence in the test's validity. Knowing as I do that I am more likely to be diagnosed as "deceptive" if I respond strongly to the critical question, and knowing also, of course, that I might respond strongly even though innocent, then it is very likely that I will respond strongly, even though innocent, and thus become a 'false positive'. The reader can decide for himself whether he would share the conviction of the professional polygrapher that all subjects - including the reader - can be made to have certain faith in the infallibility of the lie test during the course of the pre-examination interview.

Assuming that the best one could hope for would be that most subjects could be led to have at least some confidence in the lie test, the remaining problem is to find a way to control for individual differences in the variables L and F which will act to influence the critical response independently of guilt or Innocence. It will be recalled that, in the guilty knowledge test, the response to the 'correct' alternative will also be influenced by these same variables, autonomic lability and fear of the consequences of 'failing' the test. But these same variables also affect the responses to the 'incorrect' alternatives which therefore provide the necessary control; by subtracting from the 'correct' response the mean response to the 'incorrect' alternatives, one can estimate the increment added by the guilty subject's recognition of the 'correct' alternative. In other words, in the case of the GKT, the responses to the 'incorrect' alternatives provide an excellent estimate of the response which that subject would be expected to make to the 'correct' alternative if he is in fact innocent.

How might one devise an equally acceptable control question for the lie

detector situation, a question which could be expected to elicit a response which is a good estimate of the response which that subject could be expected to make to the 'critical' question if he is in fact innocent? An ideal situation could be imagined along the following lines. Suppose that our subject is suspected of being guilty of Crime X. As far as he knows, he might equally well be suspected of Crime Y, which carries penalties of the same degree of severity as Crime X. However, unbeknown to the subject, we happen to have certain evidence that he is not guilty of Crime Y. Under these rather special circumstances, the subject's response to the control question, "Did you commit Crime Y?", provides us with a good estimate of how that subject might be expected to respond to the critical question, "Did you commit Crime X?", if he is innocent of Crime X. In terms of the foregoing formulas, we can express the difference between these two responses as follows:

$$(3) \text{ Expected "lie score" if innocent} = R_x - R_y = [LF(\frac{1-C}{2}) + M] - [LF(\frac{1-C}{2}) + M] = \text{Zero.}$$

$$(4) \text{ Expected "lie score" if guilty} = [LF(\frac{1+C}{2}) + M] - [LF(\frac{1-C}{2}) + M] = CLF.$$

That is, the difference in magnitude of the critical and control responses would be expected to average about zero for an innocent subject but, for a subject who is guilty, this difference would be expected to vary directly with the product of L, F and C, i.e. with his autonomic lability, his fear of being prosecuted, and his confidence in the validity of the lie test. As long as none of these individual difference variables are zero, the guilty might be expected to be well differentiated from the innocent with this form of lie test.

But the lie detector test actually employed in the field does not use the sort of control question we have been supposing here. Instead, the critical response is compared with that subject's response to the so-called "emotional

control" question, of the form, "Did you ever steal anything before you were 18 years old?". The subject is expected to reply truthfully to this question and all that is required of the question is that it should produce for this subject some sort of emotional response. But clearly this response is not an adequate "control" at all. In the terms of our formulæ, although the autonomic lability term, L, should apply equally to the "control" as to the critical response, it would obviously be nonsensical to imagine that the fear variable, F, could be equilibrated for the two questions. How could one undertake to ask Jones a question, to which he will answer truthfully, which would be neither more nor less fearsome or disturbing to Jones than will be the critical question, "Did you rape Miss Fisbee?", so that his response to the former question will serve as a reasonable estimate of how he would respond to the latter if innocent? It is equally difficult to imagine how one might design such an "emotional control" question in response to which the confidence variable, C, would play the same role that it plays in the critical response. In short, the standard lie test in which the "lie score" is a function of the difference between the responses to the critical and "emotional control" questions can only be said to provide some degree of control for individual differences in autonomic lability; the "lie score" will continue to vary, not only with respect to differences among subjects in their confidence in the procedure, but also with individual differences in the relative emotionality or fearfulness of the critical and "control" questions.

As was indicated in the preceding section, some professional polygraphers do make use of what is called a "guilt complex" question as an additional control. The rationale for this procedure in the polygraphy literature is somewhat ambiguous or vague but it does appear that at least some field polygraphers do sometimes try to formulate a question relating to a real or imaginary crime of which the subject is known to be innocent. On its face, the "guilt complex" question

obviously comes closer to the sort of control recommended earlier for the lie detector test but there are several reasons for being dubious about this. For one thing, the "guilt complex" question seems always to be used together with "emotional control" questions as discussed above and treated in the scoring procedure as if the two were equivalent, a practice which would not seem to inspire confidence. Secondly, one does not find in the literature any discussion of the rather formidable problems which must be dealt with if the "guilt complex" question is to serve its control function, i.e. to produce a response which is a reasonable estimate of what the critical response should be if the subject is innocent and answering truthfully. It will be recalled from our earlier example that the control question, concerning Crime Y, will provide an adequate control only if (a) the subject regards Crimes X and Y as similar and prosecution for either as equally serious, and (b) that the subject believes that he is equally and independently suspected of both crimes and that he is therefore truly in jeopardy of being prosecuted for Y even though innocent of that crime. Now it is possible to imagine a field situation in which such a control might be established but it is abundantly clear that this would not be possible in most situations where the lie detector is used (the "emotional control" question is always feasible but has the important defect that it cannot be expected to work very well). Moreover, if it were to become known that the standard lie detector examination had the Crime X versus Crime Y format, criminal suspects could be expected quickly to identify which question is 'critical' and which is the 'control' and the effectiveness of the procedure will be forfeit as soon as that happens.

To summarize this section, we have seen that the autonomic response to the critical question will always be influenced by individual difference variables which are not a function of the subject's guilt or innocence. Therefore, if it is

to have any hope of high validity, the lie detector test must provide a means of comparing the actual critical response to some reasonable estimate of what that subject's critical response would be if he were innocent. Such a control is an integral part of the GKT but the GKT can only be used in a limited proportion of criminal cases and is apparently never used by field polygraphers. We have seen that the 'emotional control' question commonly used in the field does not provide the reasonable estimate required. Under very special circumstances, where the subject can be persuaded that he is also and independently suspected of a second crime, of which he is actually innocent, then his response to a question about this crime might provide the needed estimate. However, this specific method is seldom if ever used by field polygraphers, it could be used properly only in a limited proportion of cases, and it would quickly become invalidated if it were to come into standard use.

The purpose of the foregoing analysis has been to provide a basis for estimating the prior probability of the claim that conventional lie detection methods have extremely high validities, e.g., 92 percent (Bersh, 1969), 99 percent (Arther, 1965) or 100 percent (Kubis, 1950). There can be little doubt that professional polygraphy has a validity significantly - even substantially - better than chance; but for reasons that will become apparent later the important question about the lie detector is not whether its diagnoses are more valid than could be achieved by flipping coins but rather whether they are very nearly infallible, as the professionals claim. If the decision to amputate a leg is to depend entirely on the result of a test for osteosarcoma and if only 10 percent of those tested actually have the disease, then the difference between a test that is 80 percent valid and one that is 92 or 99 percent valid may be all the difference in the world. The true validity of the lie detector is, of course, an empirical question but it would be naive to suppose that such empirical questions are easily

answered. If the a priori likelihood was high that the validity of the lie detector was .90 or better, then one or two studies yielding such an estimate empirically might be considered adequate even though each study might be found to be defective in some way. What I have attempted to demonstrate above is that the prior probability of such high validities for lie detection methods is in fact vanishingly small and therefore, as in the case of ESP, for example, we are entitled to require unusually strong, clear, well-replicated evidence before accepting the fact in the face of the theory. As we shall discover in the next section, such evidence is wholly lacking.

Empirical Evidence of Lie Detector Validity

(1) Estimates given by professional polygraphers on the basis of their own experience are essentially worthless. This dogmatic pronouncement is not intended to disparage the integrity of these professionals, most of whom in my experience are honest and able people. But when one considers what would be required to support an accurate estimate of this sort, one can see at once that these conditions are simply never met in practice. One of my acquaintances was trained by Keeler himself and has given thousands of polygraph examinations in a police setting over a span of more than 30 years. His "conservative" estimate of the cumulative validity of the lie test in his hands is 95 percent. But how could he know which of those thousands of tests produced correct results and which did not? In many instances, a court may eventually adjudicate guilt but a courtroom verdict is hardly an infallible criterion and, in any event, final legal disposition of such cases will typically occur months after the suspect has moved beyond the ken of the polygrapher. I know of no professional who even pretends to have kept up the elaborate bookkeeping system that would be required to follow-up on all subjects interrogated. In a high proportion of cases, there will be no final disposition made at all; the vast majority of employee-screening

interrogations are of this nature.

In a significant proportion of police interrogations, the guilty suspect will confess. When the confession is made after the lie test has been definitely evaluated (but not so long after that the polygrapher never hears about it), then one might generate an adequate estimate of validity for this very limited sub-set of cases - but subjects who will later confess are hardly representative of subjects in general, especially those who have nothing to confess. Oftentimes, the confession will occur during the lie test or even in anticipation of it. Clearly, since no diagnosis has been made, none can be tested for validity in such instances and yet one may be permitted to suspect that each such experience adds a bit to the examiner's subjective certainty that he is dealing with a useful and "valid" technique. One highly experienced polygrapher testified before the 1973 session of the Minnesota State Legislature that he had conducted more than 20,000 polygraphic interrogations and that he had "never once been proven to have made a mistake"; such testimonials should not be taken very seriously.

(2) Laboratory studies cannot provide adequate validity estimates. This may be the only value judgment in the present paper with which the majority of professional polygraphers would agree. Laboratory experiments involving mock crimes or other contrivances of even less ecological validity have commonly achieved a hit-rate of from 70 to 85 percent, usually against a chance expectancy of from 20 to 50 percent. Professional examiners dismiss such studies as irrelevant parlor games and insist, reasonably, that in a police investigation, with real crimes and real punishments, the results might be altogether different and the hit-rates much higher. Since this coin has two sides, one should also point out that moving from the laboratory to the field situation might also serve to lower hit-rates.

While a guilty suspect may indeed be more apprehensive, and hence more reactive, when interrogated in the jail, than will a mock-guilty college sophomore in the laboratory, it is also true that the innocent suspect will be more reactive in the real-life situation and, thus, more likely to become a "false-positive".

(3) Adequate criteria against which to measure lie test validity are next-to-impossible to obtain in the field. Although slightly less dogmatic than the previous two, this proposition is stated strongly enough to emphasize that, while field investigations are the only trustworthy source of estimates of lie detector validity, it is exceedingly difficult in the field to establish an adequate criterion of whether the lie test diagnosis was in fact correct. For this reason, only a single field study has so far been published in which this criterion problem seems to have been well enough handled for the data to be taken seriously. Bersh (1969) obtained records on a fairly large sample of criminal investigations conducted by the military in which a standard lie detector examination had been given to the (serviceman) suspect. Each complete case file (minus only the polygraph findings) was evaluated independently by four attorneys from the office of the Judge Advocate General. These attorneys were instructed to "disregard all legal technicalities and to judge each case solely on the evidence...". Each judge was asked to eliminate cases where he felt the evidence was insufficient and then to arrive at a positive determination of guilt or innocence on the remainder.

Using the unanimous verdict of the 4-judge panel as a criterion, it was found that the polygraphers' diagnosis agreed with the criterion on 92.4 percent of 157 cases. On 59 additional cases for which only three of the four judges were in agreement, the hit-rate was lower, 74.6 percent, possibly because the

majority-agreement criterion was less valid than the unanimous-agreement criterion. As Bersh points out, polygraphic examiners in the military are undoubtedly better and more uniformly trained on the average than polygraphers generally, so that these findings probably represent the state of the art as favorably as possible.

This study is an important contribution to the literature in this field but, as Bersh himself indicates, the findings must be interpreted with caution. First of all, the 157 cases on which the polygraphers achieved the hit-rate of 92.4 percent were a highly selected sample and, presumably, not representative even of the general run of polygraphic investigations done in a military setting. We know that another 59 cases produced a 3:1 vote of the panel (and only a 75 percent hit-rate against a chance expectancy of about 50 percent) but we are not told in how many cases the judges split 2:2 nor, especially, how many of the original cases were discarded because one or more judges found the file data to be inadequate to yield a positive judgment. Most important, however, is the fact that none of these findings tell us anything definite about the validity of the polygraphic lie test itself since, as a matter of routine, the examiners had complete access to the case file - the evidence against the suspect - as well as any information that they could descry in the pre-test interview, the subject's appearance and demeanor and the like, prior to the examination and there is no way of determining the relative weights that were subjectively allotted to this information, as opposed to the polygraph records themselves, as the examiner arrived at his diagnosis. Since the validity criterion was a judgment based on the file records, a jaundiced appraisal of Bersh's findings might be that he has shown merely that when four judges agree that the evidence indicates guilt or innocence, a fifth judge is very likely also to agree on the basis of the same

data. Such an appraisal is probably unduly harsh - Bersh points out that the files were often less complete at the time of the lie test than when evaluated by the attorneys - still, however, it has to be admitted that one cannot say with certainty that the polygraph charts contributed anything at all to the accuracy of the original diagnoses.

Summarizing this examination of the evidence for the validity of the lie detector, after eliminating from consideration the testimonials of experienced polygraphers and after eliminating also the results of laboratory studies which, while obviously useful in extending our understanding of these phenomena, cannot be safely used to estimate the validity of field testing, we find that we are left with exactly one good field study. And that one study, while it indicates that Army-trained examiners can apparently do very well indeed in determining the guilt or innocence of at least a sub-set of criminal suspects, unfortunately sheds no clear light on the validity of the autonomic-response portion of the lie detector test. It was contended in the previous section that the prior probability of the autonomic lie detector as having .90-plus validity was so low that one might reasonably insist on especially clear, well-replicated evidence before accepting such a proposition. We see now that no such evidence exists.

Before continuing, however, it would be only fair to emphasize the important positive conclusion that can be drawn from Bersh's data concerning the validity of the human lie detector. In a criminal investigation situation, where the evidence against a suspect is suggestive but not yet overwhelming and where that suspect still maintains his innocence, a well-trained polygraphic interrogator can in about one hour's time arrive at a diagnosis of guilt or innocence that is likely to be correct 9 times out of 10. It would be interesting

to know whether that hit-rate would be greatly reduced if the polygraph pens were to be activated by random-noise generators rather than by transducers connected to the subject. It would also be interesting to compare the polygrapher's hit-rate against the validity of judgments made by an experienced policeman who had interviewed the suspect at about the same time, without benefit of the polygraphic props and ritual. But, in any case, it should be apparent that a method which can achieve such a degree of accuracy at so little cost could be of great benefit both in increasing the efficiency of investigation and in protecting the innocent. When the polygrapher says, "Guilty", all efforts can be concentrated on finding the physical evidence that will prove the suspect's guilt; when the polygrapher says, "Innocent", those same efforts can be re-deployed in another direction. As long as the polygrapher's diagnosis is regarded as tentative and advisory rather than as conclusive and the investigating officers remember that the polygrapher will be wrong perhaps 10 percent of the time, the use of this (human) lie detector can be of real social benefit.

The Role of the Answer in the Polygraph Examination

The very name, "lie detector test", implies that the subject must be required to answer the critical question, "Yes" or "No", and that it is the function of the test to determine whether that answer is truthful or a lie. We have already seen that the GKT does not require such answers. If a GKT item involves, say, the name of the loan office that was robbed, the alternatives being of the form, "Did you rob the Q Loan Company?", the subject might be required to answer each alternative, "Yes" or "No", or he might be required just to repeat each name, "Q Loan Company", or he might merely be asked to sit quietly and listen. One can see at once that the same options might be available as well for the lie detector test. To the critical question, "Did you rob the Q Loan Company?" - remember that here all subjects, guilty or innocent, will be aware

that the Q-Loan Company was robbed - the subject might be required to answer, "No", or merely to repeat the question, or merely to sit quietly and listen. In both types of test, one would expect the question itself to produce an autonomic response from which the subject's guilt or innocence might be inferred (although with far greater confidence in the GKT format, as we have seen). The subject's verbal response to each question or alternative merely serves to augment the reaction produced by the question itself. In a GKT where, because of my guilty knowledge I recognize that Z is the 'correct' alternative, I will tend to respond more strongly to the question about Z than I will to the alternatives involving X and Y; if I am required to repeat the question or to answer it, it is likely that this response requirement will enhance my autonomic reaction.

Similarly, in an idealized lie detector test, where the examiner knows that the suspect is innocent of Crimes X and Y, but the subject believes that he is truly suspected of these as well as of Crime Z, of which he is guilty, then we can expect him to respond more strongly to the question about Z than to the questions about X and Y even if he is not required to answer any of them, i.e., even if he never lies. Alternatively, we might imagine a situation in which the subject voluntarily asserts three propositions, without being questioned, one of which, Z, is a deliberate lie. Here again, we might expect that the autonomic response associated with the statement Z will be stronger than that associated with X or Y, because telling a lie, ceteris paribus, is more arousing than is telling the truth. Both the GKT and the lie detector test could be given without requiring the subject to actually answer the questions. In the case of the GKT, there is evidence (Lykken, 1960) that very high validities (100 percent 'hits') can be achieved when the subjects merely echo the question alternatives. The poorer results achieved by Shakhar, Liebllich and Kugelmass (1970) - 77 percent 'hits' - might have been due in part to the fact that these authors did not require

their subjects to respond to the questions at all and thus did not make use of the response enhancement which echoing provides. As we have seen, there is sufficient doubt about the alleged high validities of lie detector methods, even with maximum "response enhancement" through requiring subjects to answer all questions, to make this issue seem rather academic in connection with the lie detector format.

Should Polygraphers Be Allowed to Testify in Court?

The answer to this question on present evidence, obviously, is "No". But suppose that a number of future studies like Bersh's make it possible to conclude that polygraphers in general, working in a variety of civilian settings, can routinely achieve validities on the order of 90 percent. (Even given my own strong doubts about the role of the autonomic response data, I am prepared to believe that experienced human lie detectors might do this well.) Very few psychometric or diagnostic procedures are as accurate as this and yet psychiatrists and psychologists commonly give expert testimony as to whether an individual is schizophrenic or psychopathic or mentally defective and the like; why not then the professional polygrapher?

One relevant consideration is that the testimony of these other psychometric experts is seldom if ever directed to the central facts at issue in the case - who did what to whom? - but is rather addressed to questions of motivation, degree of damages, responsibility and appropriate disposition. The polygrapher, on the other hand, offers to answer directly the central questions which are traditionally left in the hands of the jury - is the defendent lying? - is he innocent or guilty? Were such testimony to be permitted to be used against a defendent, then one would expect it to be relied upon primarily in those instances in which the prosecution had been unable to assemble sufficient evidence of the

usual sort to obtain a conviction - or even that less than the usual degree of investigative effort might be made since the polygrapher's testimony was to be available. If, as I believe to be the case, the accuracy of the professional polygrapher's judgment derives more from his experience and "clinical" judgment than it does from the objective information-content of the polygraph charts, then to allow the polygrapher to give testimony in court would amount to a fundamental change in our system of justice, substituting the judgment of a professional examiner for the judgment of the lay jury.

Some polygraphers argue in favor of admitting such testimony only on the side of the defense. The defendant would submit to a polygraphic examination on his own initiative and, if the results were in his favor, the examiner would be allowed to testify to that effect in his behalf. The difficulty with this scheme is that only a highly selected sub-set of examination results would then ever reach the courtroom; even if solid evidence existed that polygraphers were correct 9 times out of 10 in general, one could not suppose that these same accuracy figures appertained to this selected sample. If every guilty defendant, having no other defense and nothing to lose, took the lie test in these circumstances, then the only fraction of these tests that would result in testimony would be precisely those that were invalid - the 10 percent false-negatives. Even if the rules required that the test results be reported however they came out, one could not evaluate the testimony without separate validity studies of this specific situation, in which the population studied was restricted to those who volunteer for the examination under these peculiar circumstances.

In those special situations where an adequate guilty knowledge test can be conducted, the question of admissibility might be rather differently regarded. Suppose the defendant is accused of 10 separate armed robberies but could not be positively identified by any of the victims because the robber wore a mask. It

would be easy to construct a 10-item GKT in which each item consisted of a photograph of one of the victims as he looked at the time of the robbery, presented in random order along with pictures of four other people whom an innocent suspect might equally well suppose to be the actual victim. If the polygrapher expert testifies in court that the defendant showed a stronger autonomic response to the picture of the actual victim, rather than to any of the four 'incorrect' alternatives, in 8 out of the 10 instances, and if he explained that the chances of that happening, were the defendant unacquainted with any of the victims as he claimed to be, are less than 2 in 1,000,000, then it would seem to me that a lay jury should be able to properly evaluate those findings without being bemused by technicalities nor overly dependent upon the expertise of the witness.

The Lie Detector in Employment Screening

We turn now to a consideration of an area in which the guilty knowledge method cannot be used but where the use of the lie detector has grown so rapidly in recent years as to threaten to become a major industry - and a major influence on our society. This is the area of employee screening; pre-employment screening of job applicants to determine who should not be hired, and routine post-employment screening to find out who has been pilfering from the company during the preceding year. Just as the guilty knowledge method is fundamentally different from the lie detector, so too is the use of the lie detector in employee screening different in certain basic ways from the use of that same technique in criminal investigation by the police.

For purposes of comparison, consider the police investigation situation first. A crime has been committed and one or more likely suspects are to be examined. In the police situation, it might be reasonable for us to assume that the base rates

for lying may be fairly high, perhaps approaching 50 percent, because the police will not bother to examine anyone who is not a "likely" suspect. In what follows, I am also going to make the unproven assumption that the overall validity of the conventional lie detector method is as high as 90 percent. Now if, over many police examinations of the type we are considering, 50 percent of the suspects do in fact try to lie, and if the lie detector is 90 percent valid, then we can expect to identify 45 out of every 50 liars while only 5 of every 50 truth-tellers will be incorrectly classified as liars, i.e., will be 'false-positives'. Such a result would obviously be very useful to the police and it could be achieved at relatively minor cost, assuming that lie detector results are never tolerated as evidence in court, because the only harm done to the 5 false-positives will be that they will continue for a time to be the subject of intensive police interest and investigation. If we can assume textbook police work, with no beatings and no manufacture of evidence, then we can suppose that "intensive police interest" will eventually flag after it has been unable to produce admissible physical evidence that these suspects are in fact guilty.

The first basic difference between police lie detection and the employment screening situation has to do with this question of the consequences of being a false-positive. The great impetus behind the growing use of the lie detector with employees is the enormous annual losses attributed to employee theft, so there is an understandable attraction in a relatively quick and cheap procedure which claims to be able to weed out potential thieves, drug addicts, and the like before they are hired or to identify light-fingered employees in the annual screening. Since the main point of using the lie detector in this application is to save money, to save both the losses that would occur with no screening or the considerable costs of conventional private detective work or individual background

Investigations, one can see that these savings will be realized only if the lie detector is permitted to make the final decisions; if the examinee fails the lie test then he isn't hired. Although one cannot send a man to prison for flunking a lie detector test, there are no Constitutional guarantees which prevent him from being deprived of a job for the same reason.

Let me illustrate this particular problem with a recent real-life occurrence. An advertising agency in Minneapolis kept a \$6000 movie camera in a locked cabinet. The camera was stolen without forcible entry and a local private detective agency urged that the four employees who had keys to this cabinet should be asked to take a polygraph test. The employees, who felt of course that they could hardly claim to be innocent and yet refuse to take the test, agreed to these plans; one of them clearly "failed" the lie test. This apparent culprit happened to be a young, black account executive who was doing so well with the company that they were reluctant to accept this verdict. The young man was sent back for a second test by a different examiner; again he "failed". A third test was conducted by the most experienced of the polygraphers and this one too resulted in a verdict of "probable deception". Just at the moment when the company president had regretfully determined to fire this apparent thief, a fortuitous set of circumstances led a fifth individual, who had never taken the lie test, to confess that he had stolen the camera, completely exonerating the other suspect. Except for this lucky accident, that first young man would not now be holding a well-paying job, making good use of his considerable talents, but instead he would be out on the street, saddled with a history of having been fired from his previous employment for having failed a lie detector test, and with negligible hope of ever again finding work in his chosen profession.

But I want to emphasize a second respect in which the employee-screening

situation differs fundamentally from that of criminal investigation. If 1000 employees are given a routine annual polygraph examination and asked whether they have stolen anything from the company during the preceding year, it would clearly be unreasonable to expect that as many as half of them will in fact lie. Nobody knows what the base-rates for lying are in such a situation but I should think that 5 percent would be a reasonable figure. Now let us continue to assume that the lie detector can be as good as 90 percent valid on the average, even in this application. Does that then mean that we can expect to detect 90 percent of the liars and also that only 1 in 10 of those who "fail" the test are actually telling the truth? - it doesn't mean anything of the kind, not with these base rates.

In what follows, I shall make use of the analysis presented in a classic paper by Meehl and Rosen (1955) which illuminates the relationship between the base rates of a condition (such as lying) in a population of persons to be tested, with the psychometric efficiency of the test that will be used to detect that condition (such as a lie detector). If the lie detector could produce an objective, numerical score, like a standardized aptitude or personality scale, then we might establish by preliminary research an optimum cutting score such that, if we called everyone scoring above the cut "liars", we would maximize the overall hit-rate. Under such conditions, the best we could expect to do would be to identify 45 of the 50 liars in our group of 1000 examinees. But, at the same time, 95 of the innocent truth-tellers would also be expected to "fail", which means that $(95)/(45+95) = \underline{68 \text{ percent}}$ of the people who "fail" the test will be actually innocent!

However, even this rather forlorn result is better than we can actually hope to achieve in practice because in fact the lie detector does not yield an

objective, numerical score. As in the Rorschach or the TAT, the examiner's brain is an integral part of the machinery of the test and it would obviously be very difficult to train an examiner to set his subjective cutting score in such a way as to produce an optimum validity for a given base-rate situation. For the case under consideration, if he set his cutting score too low, "failing" 40 percent of the cases for example, then the best he could do would be to "fail" all of the liars but at the cost of "failing" more than a third of the truth-tellers, in which case 88 percent of those identified as "liars" would in fact be innocent. To properly train operators for this situation we would need the same nearly unobtainable facilities that we would need to make meaningful estimates of the actual validity of the lie detector in the employment screening situation, namely an independent criterion which would tell us which few subjects were in fact lying. Since this is almost certainly impossible to achieve, since validity estimates obtained from a very different situation with very different base rates are a most uncertain guide when one is working in this new situation, and since whatever validity one can achieve here will be at the cost of an extremely high proportion of false positives - innocent subjects who "fail" the test and therefore lose their jobs or suffer other unwarranted consequences as a result - it seems plain to me that the use of the lie detector in this lucrative but highly dubious application should not be condoned.

Any professional polygrapher who may read these words will be gnashing his teeth at this point, since these statistical arguments will seem to conflict sharply with his actual experience in this lie detector application. He will be thinking of all those shame-faced individuals who have been led to confess assorted speculations under the pressure of the lie detector examination. This undeniable ability of the polygraph to function as a sort of "painless third-degree" was mentioned briefly earlier. If it were possible to make use only

these elicited confessions, discarding from further consideration all tests which did not produce such a result, then these questions of validity would be dealt with very differently. But, of course, once the population to be tested gets wind of the fact that no one is going to "fail" unless he confesses, there are unlikely to be any more valid confessions. A reasonable proportion of subjects will have to be "failed" on the basis of the test alone and, under the circumstances I have been assuming, we must expect that most of those who "fail" in this way will actually be innocent.

The conclusion from all of this seems to me to be obvious and compelling. It is most unlikely that we shall ever have an objective lie detector procedure which can routinely claim as high as 90 percent validity; there is no reason whatever to imagine that any known lie detector, human or polygraphic, can achieve validities close to 100 percent. Lacking near-perfect validity, both of the previous considerations militate against the use of the lie detector at all in employment screening - i.e., both the fact that decisions here tend to be based on the lie test alone and the fact that, with low base rates for lying, the majority of those who fail the test are going to be innocent. For this reason, I testified before the 1973 session of the Minnesota Legislature in support of a bill (Senate File 612) which specifically prohibits any employer to require or even to request any employee or prospective employee to take a polygraph examination. The single exception authorized by this act is the use of the lie detector in the investigation by a police chief of possible malfeasance on the part of a police officer. I am happy to report that this bill was passed and is now the law in Minnesota.

SUMMARY AND CONCLUSIONS

The polygraphic interrogation industry is rapidly expanding and is already having considerable social impact. Polygraphic interrogation is a form of psychological testing; psychometric theory and psychophysiology are - or should be - its basic sciences. In fact, however, only a handful of psychologists know enough about the lie detection business to evaluate its claims and to provide expert testimony to courts and legislative bodies.

A polygraphic 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. It is certain that the GKT could not be usefully adapted to employee screening, the application which is currently providing the strong economic impetus to growth of the polygraphy industry.

The techniques used by professional polygraphers can be appropriately characterized as methods of lie detection. Analysis of the assumptions underlying lie detection indicates that, while these methods might be expected to have significant validity (better, indeed, than many other commonly used psychological tests), the a priori probability is exceedingly small that the lie detector could be nearly as dependable as the professionals claim. Since these claims of 95, 98 and even 100 percent validity are so implausible, they should be taken seriously only if accompanied by unusually clear, well-replicated empirical evidence. Such evidence is wholly lacking.

In the pre-employment screening situation, where the base-rate for lying can be expected to be low, a high proportion - probably the majority - of those

who "fail" the lie detector test will be false-positives. For example, if 5 percent of those tested are liars, then even with a test having 90 percent validity fully 68 percent of those who "fail" will be innocent truth-tellers. For this reason, and because in the employee screening application the lie test typically determines disposition of the case - if you fail the lie test, you do not get the job - widespread use of these methods in the private sector must inevitably work an unjust hardship on many innocent but autonomically reactive individuals. In certain sensitive occupations such as police work, where social considerations require attaching greater importance to false-negative than to false-positive predictions, the use of the lie detector as a selection device can be justified. The general use of the lie detector in employee screening cannot be justified, however, and psychologists have a professional responsibility to oppose this growing practice.

The use of polygraphic interrogation in criminal investigation is fundamentally different. Here the base-rate for lying is undoubtedly higher and, most important, the penalty imposed upon the unlucky false-positive is less severe - one cannot be sent to prison for "failing" the lie test. Judicious use of the polygraph in the criminal investigation context can not only improve the efficiency of police work but could also serve as a bulwark to protect the innocent from false prosecution. It is argued that the apparent potential of the Guilty Knowledge Method should begin to be exploited in this application.

References

- Arther, R.O. The scientific investigator. Springfield, Illinois: Thomas, 1965.
- Barland, G.H. & Raskin, D.C. Detection of deception. In W.F. Prokasy & D.C. Raskin (Eds) Electrodermal activity in psychological research. New York: Academic Press, 1973.
- Berlyne, D.E. Conflict, arousal and curiosity. New York: McGraw-Hill, 1960.
- Ben Shankar, G., Liebllich, I. & Kugelmas, S. Guilty knowledge technique: Application of signal detection measures. Journal of Applied Psychology, 1970, 54, 409-413.
- Benussi, V. Die Atmungssymptome der Lüge. Archiv für die gesamte Psychologie, 1914, 31, 244-273.
- Bersh, P.J. A validation of polygraph examiner judgments. Journal of Applied Psychology, 1969, 53, 399-403.
- Davidson, P.O. The validity of the guilty-knowledge technique: The effects of motivation. Journal of Applied Psychology, 1968, 52, 62-65.
- Kubis, J.F. Experimental and statistical factors in the diagnosis of consciously suppressed affective experience. Journal of Clinical Psychology, 1950, 6, 12-16.
- Lee, C.D. The instrumental detection of deception. Springfield, Illinois: Thomas, 1953.
- Lykken, D.T. The GSR in the detection of guilt. Journal of Applied Psychology, 1959, 43, 385-388.
- Lykken, D.T. The validity of the guilty knowledge technique: The effects of faking. Journal of Applied Psychology, 1960, 44, 258-262.
- Marston, W.M. The lie detector test. New York: Smith, 1938.
- Meehl, P.E. & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin, 1955, 52, 194-216.
- Orne, M.T., Thackray, R.I. & Paskewitz, D.A. On the detection of deception: A model for the study of the physiological effects of psychological stimuli. In N.S. Greenfield & R.A. Sternbach (Eds) Handbook of Psychophysiology. New York: Holt, Rinehart and Winston, 1972.
- Reid, J.E. & Inbau, F.E. Truth and deception: The polygraph ("lie-detector") technique. Baltimore, Maryland: Williams & Wilkins, 1966.
- Summers, W.G. Science can get the confession. Fordham Law Review, 1939, 5, 334-354.
- Trovillo, P.V. A history of lie detection. Journal of Criminal Law, Criminology and Police Science, 1939, 29, 848-881: 30, 104-119.