Polygraph

Volume 12 September 1983 Number 3

Contents

PUBLISHED QUARTERLY

P.O. Box 74, Linthicum Heights, Maryland 21090

SAFEGUARDING NATIONAL SECURITY INFORMATION AND POLYGRAPH EXAMINATIONS OF FEDERAL EMPLOYEES

By

Richard K. Willard

Statement of Richard K. Willard, Deputy Assistant Attorney General, U.S. Department of Justice, before the Legislation and National Security Subcommittee of the Committee on Government Operations, U.S. House of Representatives concerning Presidential Directive on Safeguarding National Security Information and Polygraph Examinations of Federal Employees.

* * * * * *

Mister Chairman, I appreciate the opportunity to appear before this Committee to describe the background and purpose of President Reagan's recent directive on safeguarding national security information. In addition, I will undertake to describe this Administration's views concerning polygraph examinations of federal employees.

Background

Since the days of the Founding Fathers, we have recognized the need to protect military and diplomatic secrets. This need is even more acute today because of the dangerous world in which we live, including the ever-present threat of nuclear war.

Our adversaries employ highly efficient intelligence services, which use overt and covert means to gather information concerning American military capabilities, diplomatic intentions, and our own intelligence efforts. The security of this nation and the peace of the world depend in large part on our ability to keep certain kinds of this information secret.

Our task is complicated by the fact that we have a tradition of free speech and a form of government that depends upon an informed electorate. Unnecessary secrecy is contrary to our most fundamental values.

When information must be kept secret in the interest of national security, its protection is an important constitutional responsibility of the President. Since at least 1940, Presidents have safeguarded such information through Executive orders providing for a system of classification. In a number of civil and criminal statutes, Congress has recognized the President's authority to adopt such Executive orders.

The present Executive Order on classification was issued by President Reagan in 1982. It limits the use of classification to information that "reasonably could be expected to cause damage to the national security" if released without proper authorization. This Executive Order also prohibits the use of classification to conceal violations of law, inefficiency or administrative error, or to prevent embarrassment to a government agency or employee.

The unauthorized disclosure of classified information has been specifically prohibited by each of the Executive orders on this subject. Such disclosures also violate numerous more general standards of conduct for government employees based on statutes and regulations. Moreover, in virtually all cases the unauthorized disclosure of classified information potentially violates one or more federal criminal statutes.

Notwithstanding the clear illegality of this practice, unauthorized disclosures of classified information have appeared in the media regularly and frequently over the last decade. Examples of the harm caused by such disclosures are themselves classified and cannot be provided in a public hearing. However, the intelligence committees of the Congress are well aware of the harm caused by unauthorized disclosures of classified information.

During the Carter Administration, these committees expressed their concern that this problem was not being sufficiently addressed. In 1979, a subcommittee of the House Permanent Select Committee on Intelligence held hearings on the subject "Espionage Laws and Leaks." The chairman's opening remarks included the following observation:

> We begin the process with a concern for the Government's ability to protect against classic espionage, leaks of highly classified material and with what has been called gray-mail ... We will ask why it is that unauthorized disclosures of classified information, particularly leaks of intelligence information, go unprosecuted and unpunished. 1/

The ranking minority member's opening remarks included the following statement:

> We are equally concerned about development of ways by which we can detect leaks and investigate and prosecute those who jeopardize our national interest by leaking classified intelligence information. ... As to the matter of leaks, for varying reasons, no effective action has been taken in order to prevent a problem which is growing at a substantial rate. 2/

A subcommittee of the Senate Select Committee on Intelligence issued a report in 1978 entitled "National Security Secrets and the Administration of Justice." The first conclusion stated in the report was as follows:

> (A) There has been a major failure on the part of the Government to take action in leak cases. -- ... The committee found that leak cases are uniquely difficult to investigate. But, we found cases where no action was taken -- investigation or prosecution -- even where a leak clearly violated an existing statute and caused serious harm to national security. 3/

The problem of unauthorized disclosures of classified information is probably no worse today than it was during the Carter Administration, but it continues to be a matter of serious concern. President Reagan recently took the unprecedented step of sending a memorandum on this subject to federal employees who handled classified information. The President's

memorandum includes the following statements:

> Recent unauthorized disclosures of classified information concerning our diplomatic, military, and intelligence activities threaten our ability to carry our national security policy. ... [These] disclosures are so harmful to our national security that I wish to underscore to each of you the seriousness with which I view them.
>
> The unauthorized disclosure of our Nation's classified information by those entrusted with its protection is improper, unethical, and plain wrong. ...
>
> The American people have placed a special trust and confidence in each of us to protect their property with which we are entrusted, including classified information ...
>
> * * *
>
> ... As servants of the people, we in the Federal Government must understand the duty we have to those who place their trust in us. I ask each of you to join me in redoubling our efforts to protect that trust.

A copy of the complete memorandum is attached to my statement as Exhibit A.

Presidential Directive on Safeguarding National Security Information

In addition to reminding federal employees of their personal responsibilities, the President has issued a directive that requires a number of additional steps to be taken to protect against unlawful disclosures of classified information. A copy of the text of that directive, known as National Security Decision Directive 84, or NSDD-84, is attached as Exhibit B to my statement. In summary, NSDD-84 provides:

-- additional restrictions upon government employees who are entrusted with access to classified information, and upon government agencies that originate or handle classified information;

-- a more efficient approach to investigating unauthorized disclosures, including additional use of polygraph examinations under carefully controlled circumstances; and

-- mandatory administrative sanctions for employees found to have knowingly disclosed classified information without authorization, or who refuse to cooperate with an investigation.

Implementation of NSDD-84 has required a careful review of security regulations and practices throughout the government. A number of changes are being made as a result of this review.

A significant aspect of implementing NSDD-84 has been the development of two new nondisclosure agreement forms for government-wide use. One of

these forms is a classified information nondisclosure agreement, which has been promulgated by the Information Security Oversight Office as Standard Form 189. 4/ This form does not include a provision for prepublication review.

The other form is a nondisclosure agreement to be signed as a condition of access to Sensitive Compartmented Information, or SCI. This agreement was promulgated by the Director of Central Intelligence on August 30, 1983, as a replacement for Form 4193, which had been adopted in 1981. Both versions of this form contain provisions for prepublication review, but we believe the new form will provide the government with an enhanced ability to safeguard classified information. A copy of the new SCI nondisclosure agreement is attached as Exhibit C to my statement.

Secrecy agreements requiring prepublication review have been used at CIA for many years, and in 1980 the Supreme Court approved their use in Snepp v. United States. 5/ The sole purpose of prepublication review is to permit deletion of classified information before it is made public. This program does not permit the government to censor material because it is embarrassing or critical. Earlier this month, the United States Court of Appeals for the District of Columbia Circuit issued an opinion upholding the manner in which CIA conducted its prepublication review of a former employee's magazine article. 6/

The Department of Justice has determined that the two new nondisclosure agreements adopted to implement NSDD-84 would be enforceable in civil litigation initiated by the United States. The Department has also issued regulations for its prepublication review program. Copies of thse documents are attached as Exhibits D and E to my statement.

Use of the Polygraph in Safeguarding National Security Information

NSDD-84 was based upon the recommendations of an interdepartmental group convened by the Attorney General. I served as chairman of this group, which also included representatives designated by the Secretaries of States, the Treasury, Defense, Energy, and the Director of Central Intelligence. Copies of the report of this group, which is unclassified, have been furnished to the Committee.

Our report found a number of deficiencies in the system by which the government investigated unauthorized disclosures of classified information appearing in the media. We concluded that this system was "so ineffectual as to perpetuate the notion that the government can do nothing to stop leaks of classified information."

NSDD-84 includes a number of steps to streamline the reporting and investigation of unauthorized disclosures. Among other things, it clarifies the FBI's authority to investigate cases in which administrative sanctions may be sought instead of criminal prosecution.

Our report considered the question of polygraph use and concluded as follows:

> The polygraph can be a useful tool in leak investigations under certain circumstances. It should be used selectively and its results considered within the context of a complete investigation. The polygraph should not be used for dragnet-type screening of a large number of suspects or as a substitute for logical investigation by conventional means. It is most helpful when conventional investigative approaches have identified a small number of individuals, one of whom is fairly certain to be culpable, but there is no other way to resolve the case. A polygraph examination in this situation can be limited to the unauthorized disclosure that is being investigated and should not include questions about life style that many employees would find offensive. Moreover, polygraph results should not be relied upon to the exclusion of other information obtained during an investigation.

The provision in NSDD-84 that is based upon the foregoing recommendation requires government agencies with employees having access to classified information to revise their regulations as necessary so that employees could be required to submit to polygraph examinations, when appropriate, in administrative investigations of unauthorized disclosures of classified information. The Directive provides further that:

> -- Agency regulations must, as a minimum, permit the agency to decide that appropriate adverse consequences will follow an employee's refusal to cooperate with a polygraph examination that is limited in scope to the circumstances of the unauthorized disclosure under investigation.
>
> -- Agency regulations may provide that only the agency head, or his delegate, is empowered to order an employee to submit to a polygraph examination.
>
> -- Results of polygraph examinations should not be relied upon to the exclusion of other information obtained during investigations.

Government agencies have substantial discretion in deciding how to mplement this aspect of NSDD-84, subject to constitutional and statutory onstraints. The policies of such agencies -- including CIA, NSA, and the epartment of Justice -- already satisfy the requirements of NSDD-84. The ffice of Personnel Management has clarified its policies as being consisent with NSDD-84. The Department of Defense has proposed a comprehensive evision in its polygraph regulation that would, among other things, satsfy the requirements of NSDD-84. Implementing regulations in other agenies are currently being prepared.

Another way in which the polygraph can be used to prevent unauthoized disclosures of classified information was not addressed in NSDD-84 r the study leading up to it. I am referring to its use in the periodic r aperiodic screening of government employees with access to certain inds of sensitive classified information to determine whether they have isclosed such information either to foreign agents or to others not auorized to receive it.

Our major intelligence agencies -- CIA and NSA -- currently use

polygraph examinations to screen candiates for employment and periodically thereafter. The proposed new Department of Defense regulation would provide for some additional uses of the polygraph. This proposal has caused some controversy, and a rider to the current authorization act contains a moratorium on any changes in Department of Defense policy regarding the polygraph until April 15, 1984. The stated purpose of the moratorium contained in the Defense authorization act was to permit congressional hearings upon this subject before new policies were implemented in that Department.

Because these various proposals for polygraph use have been of concern to this committee and others in the Congress, I have endeavored in the balance of my statement to address in a fairly comprehensive manner some general questions regarding the polygraph, with particular emphasis on its use in connection with federal employment.

Polygraph Accuracy

Although referred to as a "lie detector," the polygraph itself does not detect lies. The polygraph is an instrument that measures a variety of physiological responses of an individual undergoing questioning. These measurements assist the polygraph examiner in forming an opinion as to whether the individual has given truthful or deceptive answers to particular questions.

There are two basic categories of polygraph use: investigations of misconduct or other specific issues, and generalized screening. Generalized screening includes pre-employment testing to verify the accuracy and completeness of information provided in connection with the application for employment. Screening can also be used in the post-employment context, on either a periodic or aperiodic basis, in an attempt to determine continuing compliance with particular standards of conduct or other conditions of employment or access to information.

Numerous scientific studies have attempted to quantify the accuracy of polygraph examinations. 7/ It is important to recognize that there are substantial differences in the methodology of these studies. In addition, there are differences in the skills of particular polygraph examiners and in the types of inquiries they are asked to undertake. For these reasons, among others, the results of the studies have varied considerably.

The overwhelming majority of studies show accuracy rates for polygraph examinations within the range of 70 to 95 percent. This is obviously a wide variation, but it reflects a clear scientific consensus on one point: polygraph examinations produce statistically significant indications of deception and nondeception.

Of course, the distinction between 70 and 95 per cent accuracy is likely to make a significant difference in the purposes for which the polygraph technique should be employed. For the reasons that follow, we believe the relevant accuracy rate is closer to 95 percent than 70 percent.

-- First, "field" studies -- most of which involve subjects who are suspected of real crimes -- show a higher accuracy rate than

laboratory studies. We believe this indicates greater accuracy of the polygraph in "real-life" situations, where the subject has a keener appreciation of the consequences of deception and is thus more likely to display physiological symptoms.

-- Second, studies with more highly trained and experienced examiners show a higher accuracy rate than those using students or academics. Since government polygraph examiners generally have a high level of training and experience, we believe they are likely to have a high accuracy rate.

-- Third, studies using field polygraph instruments, which carefully measure three or more physiological reactions, show a higher accuracy rate than those using laboratory equipment and more limited measurements. Government polygraph examinations thus are likely to have the higher accuracy rates.

It is important to recognize that no one can credibly claim a 100 percent accuracy rate for any of the polygraph techniques. The possibility of error is always present. Knowing this, federal polygraph programs generally require a second evaluation of every set of polygraph charts by a senior polygraph examiner, and no report of deception or truthfulness is made without agreement on the charts.

When the examiner cannot reach a conclusion or the reviewer disagrees, the result is reported as "inconclusive." This call of "inconclusive" is a vital aspect of the quality control program, and reduces the risk that the results of any particular examination will be erroneous. When possible, additional examinations are scheduled to resolve the issue. However, there are always a few people who do not test well and a number of topics which are not suitable for resolution by polygraph testing. Thus some inconclusive results are inevitable, as well as necessary to avoid erroneous calls.

Many experimental studies do not take into account the possibility of inconclusive results and requires an evaluation of truthfulness or deception in each case. This is yet another reason that we believe the actual error rate in government-administered polygraph examinations is considerably lower than some laboratory studies would suggest.

For purposes of policy analysis, it is important to recogize that there are two types of possible error in a polygraph examiner's evaluation.

One type of error is the "false negative"; that is, a deceptive subject who is found by the examiner to be truthful. False negatives can be a signficant problem if polygraph examination results are given undue weight or are used to replace other investigatory techniques. However, if the polygraph is used as an additional means of investigation, and its results taken in the context of all available information, then the "false negative" problem does not provide a substantial reason to avoid polygraph use.

The other type of error is the "false positive," in which a truthful person is found to be deceptive. For the reasons explained above, we

believe that the actual "false positive" rate in government-administered polygraph examinations is quite small. Nevertheless, even this small possibility is a basis for serious concern because we have a deep-seated aversion to any procedure that produces false incrimination. In deciding what circumstances warrant polygraph usage in the government, our continuing concern is to eliminate or at least minimize the possibility that an individual will be treated unfairly as a result of a "false positive" polygraph examination.

The fact that polygraph examinations can produce "false positives" is not -- in itself -- a sufficient basis to renounce their use. Virtually all investigative techniques can also produce false positives. For example, the background investigation is a basic screening device for most important government jobs and security clearances. This investigation includes interviews with neighbors, associates, landlords, and references. Obviously, a certain percentage of these interviews or record checks produce "false positives" -- derogatory information that is untrue or distorted. Yet we continue to use these investigative techniques because they generally provide reliable information and we have no better alternatives.

The same kind of analysis applies to eyewitness accounts and even to many scientific tests, such as blood tests, breathanalyzers, voiceprints, fiber analysis, and urinalysis. The degree of accuracy for these techniques may vary, but in each case there is some potential for an erroneous incriminating result, or "false positive." Yet such techniques are regularly used in investigations.

One other question relating to polygraph accuracy is the possibility of countermeasures or attempts to "beat the machine." The techniques commonly used as countermeasures are known, and a well trained examiner is capable of detecting their use. Nevertheless, it may be possible for a subject to use undetected countermeasures, which could lead to a "false negative" result. This is yet another reason that the polygraph technique should not be used as a substitute for conventional means of investigation or its results given undue weight. Yet, as with the case of "false negatives" generally, it provides no basis for avoiding polygraph use altogether.

Most studies of polygraph accuracy deal with its use in investigating particular instances of criminal misconduct (or a laboratory approximation thereof). However, it would appear that similar results would obtain in screening situations, at least where the issues to be resolved are clearly defined, including such matters as prior misconduct or falsified information provided in personal history statements, which would appear to the subject to be equally serious.

The foregoing discussion has related solely to the accuracy of polygraph examinations for purposes of assessing the truthfulness or deception of an individual's response to particular questions. However, this analysis overlooks two other ways in which the polygraph technique has substantial utility.

One major benefit of polygraph use is in enhancing the ability to obtain information from the subject. Experience has shown that subjects

quite often make admissions, or volunteer additional relevant information, in connection with a polygraph examination. This phenomenon may actually produce information that is more useful in terms of the purpose of the examination than the examiner's assessment of the subject's physiological responses.

An additional benefit of polygraph use is its deterrent effect upon certain kinds of misconduct that can be difficult to detect through other means. Employees who know they are subject to polygraph examinations may be more likely to refrain from such misconduct. In addition, use of the polygraph in pre-employment screening serves to deter applicants who do not meet standards of suitability for employment.

Role of the Polygraph Examiner

Notwithstanding the potential accuracy of the polygraph technique, its reliability in particular cases is largely dependent upon the ability and integrity of the particular examiner. The opinion of an incompetent or dishonest polygraph examiner can be worthless or even deliberately misleading. Any polygraph program requires careful attention to the examiners who are used.

In real-life situations, the polygraph examiner knows the identity of the subject and is generally familiar with the subject matter of the inquiry. This background assists the examiner in formulating questions to be used in the examination and in evaluating the subject's responses, thus increasing the accuracy and utility of the technique. However, it also affords an opportunity for the examiner's conscious or subconscious biases to influence the results of the examination.

The possibility of examiner bias is not a sufficient reason to reject use of the polygraph. Eyewitness testimony and information acquired from interviews or physical surveillances are also subject to possible bias. Yet such techniques are the basic methods of conducting any investigation.

Moreover, polygraph programs in the federal government are conducted pursuant to stringent standards. Polygraph examiners employed by the federal government are carefully selected, undergo thorough training, and their work is carefully supervised.

Throughout this process, examiners are scrutinized for indications of bias that could improperly affect their judgment in specific cases. The results of each examination are reviewed by an experienced supervisor as an additional safeguard against improperly drawn conclusions. In most federal polygraph programs, current employees who fail to pass an examination are permitted a second examination by a different examiner. We believe these measures ensure that federal polygraph examiners meet the highest possible standards of competence and objectivity.

Privacy and the Polygraph

Apart from issues relating to its accuracy, there are also questions raised as to whether use of the polygraph involves an unwarranted invasion of personal privacy. Some critics fear that polygraph examinations will

be used to probe the subject's religious and political beliefs or attitudes toward labor unions. Concern is also expressed that employees will be questioned about personal matters in which the government has no legitimate interest.

These privacy concerns are important and must be addressed. Current programs for polygraphing federal applicants or employees include safeguards to ensure that improper and irrelevant questioning dues not occur. We believe the following safeguards are essential:

-- First, certain kinds of questioning must be prohibited in all situations. This prohibition includes such matters as religious and political beliefs protected by the First Amendment.

-- Second, questioning about the subject's personal life must be limited to matters that are clearly relevant to the purpose of the examination.

-- Third, all questions relevant to the purpose of the examination must be reviewed in advance with the subject. This procedure permits the subject to object if he believes an improper question is to be included in the examination.

-- Fourth, technical or control questions, which are necessary to the conduct of the examination, should be constructed to avoid embarrassing, degrading or unnecessarily intrusive matters.

-- Fifth, sufficient records must be maintained to permit monitoring and supervision of polygraph examinations to ensure compliance with these safeguards.,

By observing the foregoing safeguards, we believe that privacy objections to use of the polygraph are minimized. In this regard, it is important to recognize that the polygraph technique is not the only type of investigative technique that could result in unwarranted invasions of personal privacy. Conventional investigative techniques--such as interviews with neighbors, friends and associates--can be used in an abusive manner. For privacy purposes, the most important safeguard is a limitation on the scope of the investigation. Denial of the use of particular investigative techniques, such as the polygraph, does not address the real problem of deciding what topics of inquiry constitute an unwarranted invasion of privacy.

Another privacy-related objection to polygraph use is that the technique relies upon self-incrimination. Apart from the legal issue, which is discussed later, the phenomenon of potential self-incrimination is not a valid basis for objecting to government use of the polygraph. Most government screening and investigation is conducted on the basis of information supplied by the individual. For example, applicants for security clearances fill out extensive background information forms, which are signed under threat of criminal penalties for any false statement. The polygraph examination is simply another way of obtaining, and verifying, the same kind of information.

Legal Issues Regarding the Polygraph

In 1923, the United States Court of Appeals for the District of Columbia Circuit held that the polygraph technique was not sufficiently accepted in the scientific community as to be admissible evidence in a criminal trial. For many years this decision, Frye v. United States, 8/ was the leading case on the admissibility of polygraph examination results. However, the authority of this decision has been undermined in recent years by increasing scientific documentation of the polygraph technique and changes in the law of evidence.

All of the scientific studies referred to previously were performed after 1923, and this body of knowledge was obviously unavailable to the Frye court. Indeed, a number of courts and commentators have been favorably impressed with the currently available scientific documentation of the polygraph technique. Some commentators have also expressed the view that the legal theory of the Frye decision is inconsistent with the new Federal Rules of Evidence. 9/

There has never been a Supreme Court ruling on the admissibility of polygraph evidence in federal court. The Courts of Appeals are divided on the issue. Some adhere to the Frye decision and have a per se rule of excluding polygraph evidence. 10/ In a number of other circuits, the decision whether to admit polygraph evidence is left to the discretion of the trial judge. 11/

The state courts are similarly divided on the question of admitting polygraph results into evidence. Many states do not permit polygraph evidence to be admitted for any purpose, but a number of other states allow polygraph evidence under limited circumstances. 12/ It is fairly common for polygraph results to be admitted pursuant to a stipulation of the parties. 13/ In at least one state, polygraph evidence is generally admissible on the same basis as any other scientific evidence. 14/

The Justice Department has traditionally opposed the use of polygraph examination results in criminal trials as evidence of the guilt or innocence of the accused. Use of polygraph evidence in a criminal trial involves a number of considerations that do not apply to its use as an investigatory technique. Among these considerations are the following:

-- First, a guilty subject may be able to use undetected countermeasures, thus causing the examination to produce false evidence of innocence. As with other "false negatives," this problem is of less concern when the polygraph is used in an investigation and is cumulative to other investigative techniques.

-- Second, there is a substantial likelihood that a jury would give undue weight to polygraph examination results. This phenomenon could ultimately displace the jury's function in cases where polygraph evidence is received on the ultimate issue of guilt or innocence. 15/

-- Third, attempts to introduce polygraph evidence could greatly increase the length of criminal trials in order to accommodate the necessary expert testimony on each side of the issue.

> -- Fourth, polygraph examinations given to defendants by "friendly" examiners may be more likely to produce exonerating results, including false negatives. In this situation there is no risk that detection will have adverse consequences to the subject, since unfavorable results need never be known to anyone except defense counsel. Indeed, a defendant could be examined by a number of polygraphers until he passes a test, and then seek to use the one favorable examination as evidence of innocence.
>
> -- Fifth, because of Fifth Amendment considerations, the prosecution cannot obtain polygraph evidence without the consent of the accused. The Fifth Amendment may also prevent use of polygraph results in cases where the accused chooses not to testify. Therefore, polygraph evidence is likely to be a one-sided weapon in criminal trials, available mostly to defendants and not to prosecutors.

The foregoing concerns apply only to use of polygraph evidence in a trial, primarily a criminal trial. The same considerations do not apply to use of the polygraph as an investigatory technique.

The Justice Department has traditionally supported use of the polygraph as an adjunct to the normal interview and interrogation process in certain kinds of matters within its investigative jurisdiction.

> ... with proper ethics by the polygraph examiner and tight administrative control by the user agency, there is no question but that the polygraph can be a valuable investigative aid to supplement interrogation in selected criminal and national security cases. Interrogation is a basic tool of any investigative agency and the FBI considers the polygraph technique a thorough and specialized interview procedure in which a skillful interrogator is attempting to simply ascertain the truthful facts from a consenting individual regarding a matter in which we have jurisdiction.
>
> In some instances suspects will admit deception and furnish confession and/or signed statements. In most instances valuable new information or investigative direction is developed as a result of the examination and follow-up interrogation. 16/

The fact that polygraph evidence is not admissible in a criminal trial does not preclude its use as an investigative technique. 17/ Indeed, an investigative file will typically contain a large amount of hearsay and other information that is not admissible evidence.

Confessions and other evidence obtained as a result of polygraph examinations are also generally admissible, so long as the examination is not conducted in a manner that violates the subject's right to counsel or privilege against compulsory self incrimination. The Supreme Court has specifically held that statements made during a polygraph examination can be admissible, even if the results of the examination are not. 18/

An important caveat applies to the foregoing discussion of the law pertaining to admissibility of polygraph evidence in court: in every case

the examination must be taken voluntarily. The Supreme Court has indicated that the government cannot require a criminal suspect to submit to a polygraph examination. 19/ This limitation is a consequence of the Fifth Amendment's prohibition of compulsory self-incrimination. A criminal suspect may not be required to answer questions--with or without the polygraph.

Outside the criminal context, however, a different rule applies. A public employee can be required to answer questions or sign affidavits relating to his fitness to perform public duties. So long as the information thereby obtained is not to be used against the employee in a criminal proceeding, he can be fired for refusing to answer questions. 20/ For this reason we believe there is no valid Fifth Amendment objection to requiring a government employee suspected of misconduct to take a polygraph examination, provided that there is a recognition that the results cannot be used against the employee in a criminal proceeding. 21/

Not all uses of the polygraph in administrative investigations of misconduct require that the results be excluded from evidence in a criminal trial. The basic rule is that "statements obtained under threat of removal from office" may not be used in a criminal trial. 22/ However, the examination is not regarded as involuntary for Fifth Amendment purposes, if the refusal is only one of a number of factors to be considered in determining whether administrative sanctions should be imposed. 23/ Thus, an employee's refusal of a request to take a polygraph examination may be considered for a variety of purposes--including as giving rise to an adverse inference in disciplinary proceedings, or as a basis for reassignment or denial of access to classified information. Only if the employee can be discharged or demoted solely for refusal to take a polygraph examination, are the results of such examination inadmissible as evidence in a criminal proceeding.

Where the polygraph is used for screening purposes and not in an investigation of suspected misconduct, there is ordinarily no Fifth Amendment problem. In such circumstances, the polygraph examination is a condition of employment or access to classified information. The subject's consent to be examined is regarded as voluntary in the constitutional sense, just as in providing answers on an application for employment or for a security clearance. However, it may be necessary to discontinue an examination if an employee reveals information concerning a federal crime in the course of a screening examination that is required as a condition of continuing employment.

Positive authority to use the polygraph in screening for employment or access to classified information is derived from a variety of sources. At bottom are the President's responsibilities under Article II of the Constitution as Chief Executive, Commander-in-Chief, and the principal instrument of United States foreign policy. In Executive Order 10450, as amended, the President has assigned to the head of each department and agency the responsibility to "insure that the employment and retention in employment of any civilian officer or employee within the department or agency is clearly consistent with the interests of national security." In Executive Order 12356, the President required that a person can be eligible for access to classified information only if a "determination of trustworthiness has been made by agency heads or designated officials."

In addition to these general authorities, there are specific provisions in statutes and executive orders that authorize personnel security measures for intelligence agencies and for the protection of intelligence sources and methods. 24/

Apart from the temporary limitation contained in the current Department of Defense authorization act, there is no statute that expressly limits the power of the federal government to require employees or prospective employees to submit to polygraph examinations. Legislation to this effect has been proposed from time to time but never enacted. 25/

In 1968, the Civil Service Commission issued instructions on polygraph use as a screening device for the competitive service, which remain in effect today under the authority of the Office of Personnel Management. 26/ These instructions generally prohibit use of the polygraph to screen applicants for and appointees to the competitive service. However, agencies with a highly sensitive intelligence or counterintelligence mission directly affecting the national security, are permitted to use the polygraph for personnel screening after complying with certain standards set forth in the instructions. The OPM instructions do not apply to use of the polygraph in investigations of suspected misconduct by agency employees, including unauthorized disclosures of classified information.

General standards of conduct for federal employees permit discharge or discipline for refusal to cooperate with investigations of suspected misconduct. 27/ This authority also supports requiring government employees to submit to polygraph examinations in appropriate circumstances. 28/ As noted above, exercise of this authority can cause the results of the examination to be inadmissible in criminal proceedings.

In our view, an employee who refuses an order to take a polygraph examination in an appropriate case may be subject to a range of administrative sanctions to include removal, as well as lesser forms of discipline, such as a letter of reprimand or suspension without pay. The appropriateness of any sanction for refusal to comply with an order to take a polygraph examination would obviously depend upon the circumstances of the case, including the reason given by the employee for refusing the order.

We are not aware of any litigation in which an employee has challenged the power of a federal agency to require a polygraph examination in connection with an administrative investigation of suspected misconduct. There have been a number of cases in state courts on this issue, many dealing with policemen suspected of misconduct. Most courts to consider the issue have upheld the authority of a governmental agency to discharge an employee who refused to submit to a polygraph examination. 29/

The Merit Systems Protection Board held in the Meier case that polygraph examination results can be admitted into evidence in administrative proceedings to determine whether a federal agency has cause to discharge an employee from the competitive service, provided that a proper foundation is laid. 30/ However, the Board also held that it would not draw an adverse inference of guilt from an accused employee's refusal to volunteer for a polygraph examination, stating that, "[T]his Board will not permit any adverse inferences to be drawn from an individual's refusal to submit to a polygraph examination." 31/

The direct, legally binding effect of the Meier decision is limited by its posture, which involved an evidentiary ruling by a presiding official in a specific factual situation, and by the scope of the Board's authority to review personnel actions taken by federal agencies, which in general extends only to the removal, suspension, or demotion of competitive service and preference eligible employees. The case did not involve any national security considerations, which could well have caused the Board to reach a different result on the adverse inference issue.

Therefore, we do not believe the Meier decision legally precludes a policy of using polygraph examinations in administrative investigations of unauthorized disclosures of classified information. The Meier decision, however, does stand as a precedent for future appeals to the Board challenging adverse personnel actions based, at least in part, on an employee's refusal to submit to a polygraph examination. Any agency that proposed to take an adverse personnel action subject to the Board's jurisdiction must be prepared to demonstrate that the use of polygraphs in the investigation, and the sanctions imposed for refusal to take the polygraph examination, were justified by the reliability of the polygraph and the particular circumstances of the case, and necessary to protect the national security interests of the United States. 32/

Administration Policy on Polygraph Use

This Administration is opposed to indiscriminate use of the polygraph by the federal government. Many employees and potential employees view the prospect of polygraph examinations as offensive and unwelcome. In a small but cognizable percentage of cases, polygraph examination results are inaccurate. For these reasons, we do not believe the polygraph should be used as a screening device for government employment generally or as a routine technique for conducting investigations.

However, there are certain situations where the disadvantages of the polygraph are outweighed by specific and significant governmental interests that are served by the use of this technique. Particularly in the area of national security, we have for many years recognized that polygraph examinations have a proper role.

Specifically, it is the position of this Administration that the polygraph examinations can be properly and lawfully given to federal employees or applicants in the following situations:

-- first, as a condition of initial or continuing employment with or assignment to CIA and NSA, and for positions in other agencies that entail equally sensitive responsibilities directly affecting national security;

-- second, as a condition of access to highly sensitive categories of classified information which are likely to be of extraordinary interest to hostile intelligence services;

-- third, to investigate serious criminal cases, where the employee voluntarily consents to the examination after an opportunity to consult with counsel; and

-- fourth, to investigate serious administrative misconduct cases under limited circumstances, including unauthorized disclosures of classified information.

The following summary of Administration policy is not intended to suggest that the polygraph must be used in any particular situation. Decisions on the extent of polygraph use should be made by the head of the employing agency, taking into consideration a variety of circumstances. In addition, I have not attempted to address the use of polygraph examinations in other contexts, such as for contrators, foreign nationals, informants, or non-employee criminal suspects.

(1) As a Condition of Employment

Certain jobs are so sensitive that the federal government should leave no stone unturned in assuring that only trustworthy candidates are hired. Even if use of the polygraph may unfairly screen out some candidates who are actually qualified, we view it as more important to avoid hiring candidates who may pose a risk to national security. One noted polygraph critic has conceded:

> There is evidence that the polygraph lie test does better at detecting liars than it does in identifying the truthful. This is an important reason for advocating banning of polygraph testing of employees or job applicants in the private sector. However, for special high security situations, where it is clearly more important to screen out undesirable applicants than to give a fair employment opportunity to all applicants, then this bias against the truthful subject may not be regarded as such an important defect. 33/

This analysis supports the longstanding practice of our largest intelligence agencies -- CIA and NSA -- in using polygraph examinations as part of their overall program of screening candidates for employment or assignment. To the extent some examination results are "false positives," we regard this inequity as outweighed by the importance of assuring the suitability of individuals who are hired at these agencies.

A major by-product of this process is that applicants will volunteer additional useful information in connection with the polygraph examination. 34/ This additional information will necessarily improve the overall effectiveness of the employment screening process. Indeed, CIA and NSA have found that the great majority of their decisions not to employ individuals because of security concerns are derived from information obtained in connection with the polygraph examinations. Without the polygraph, most of these individuals would have been hired. Indeed, use of the polygraph probably deters many candidates who do not meet security standards from applying in the first place.

CIA and NSA do not have a monopoly on sensitive jobs affecting national security. We know of no plans to expand the use of pre-employment polygraph examinations beyond those agencies, except to a very limited extent at the Defense Intelligence Agency. However, we believe that the polygraph could also be properly used by other agencies to screen candidates for employment or assignment to jobs that are equally sensitive as those at CIA and NSA.

The decision to use polygraph examinations for pre-employment screening must be made by the agency head, taking into account the overall needs of the agency or component in question. Use of the polygraph to screen candidates for employment in the competitive service should also be consistent with the 1968 Civil Service Commission instructions contained in Appendix D to Chapter 736 of the Federal Personnel Manual.

It is important that use of the polygraph in pre-employment screening not be permitted to substitute for other measures such as the background investigation. In addition, polygraph examination results should not be given undue weight, but considered in the context of all available information. These conditions diminish the risk associated with "false negative" results. The polygraph examiner's conclusion simply provides an additional safeguard to ensure the accuracy of information otherwise provided by the applicant or obtained through background investigation.

Use of the polygraph as a screening device for this small number of highly sensitive jobs is fair to the applicant, who has a choice of whether or not to apply. CIA and NSA advise applicants that polygraph examinations are a condition of being hired and retained in such jobs. We believe similar warnings should be given for all jobs where polygraph examinations are a condition of employment or assignment. Potential employees who object to being polygraphed can avoid its use by not applying for these jobs. In practice, only a minute fraction of government jobs will be foreclosed to such persons.

Agencies that use polygraph examinations for pre-employment screening also use such examinations from time-to-time thereafter. Such post-employment examinations are part of an effort to ensure that employees continue to meet standards for employment and access to classified information. Applicants for employment or assignment to these agencies are advised that submitting to such polygraph examinations will be a condition of continuing employment. If such notice is provided in advance, then we believe use of the polygraph for post-employment screening by these agencies is equally justified as its use for pre-employment screening.

Of course, current employees have a greater stake in keeping their jobs than applicants have in obtaining them in the first place. For this reason, CIA and NSA take great care to assure accuracy in evaluating examinations of current employees. It is extremely rare for an employee to be discharged solely on the basis of the polygraph examiner's assessment of deception. However, this possible outcome is justified by the national interest in assuring the reliability of individuals in these jobs.

(2) As a Condition of Access to Information

The preceding category concerned polygraph examinations given as a condition of employment in certain sensitive jobs, with the scope of the examination potentially covering any matter that is legitimately related to eligibility for that employment. In contrast, this second category concerns polygraph examinations given as a condition of access to certain categories of highly sensitive classified information that is of extraordinary interest to hostile intelligence services. Such examinations would be limited in scope to "counterintelligence-type" questions.

Executive Order 12356, §4.2, authorizes certain agency heads to create Special Access Programs to control access, distribution, and protection of particularly sensitive classified information. Included among these programs is information concerning certain intelligence sources and methods, as well as communications security or cryptographic techniques and highly advanced research and development programs.

While we do not belive that all persons in these programs need be polygraphed as a condition of access, there may be certain Special Access Programs, or elements of such programs, that require the extraordinary precaution of a polygraph examination. In these situations, we believe it would be appropriate for an agency head to consider the initiation of a counterintelligence-type polygraph examination within his agency as a condition of granting or continuing access to classified information within the Special Access Program concerned. The criterion for designating programs for this purpose would be a finding that the information is likely to be of extraordinary interest to hostile intelligence services and that its continued protection is critical to U.S. national security interests.

The decision to require this sort of polygraph examination should be made by the head of the agency in which the individuals subject to this requirement are employed or assigned. The agency head is in the best position to weigh all factors that must be taken into account in imposing such a requirement.

Polygraph examinations administered for this purpose would be limited to such matters as whether the employee is acting on behalf of a foreign power, has had unauthorized contacts with foreign government representatives, or has revealed classified material to unauthorized persons. Questions about the employee's lifestyle would not be permitted. A listing of the kinds of questions that could be asked under such a program is contained in Appendix B to the proposed Department of Defense polygraph program regulation.

Such examinations, given for the purpose of determining continuing access, need not be given to all covered employees but could instead be given on an aperiodic basis to randomly-selected employees within the designated programs.

Additional safeguards in this program would include the following:

-- No adverse personnel action could be taken solely on the basis of a polygraph examiner's assessment of deception.

-- The only consequence of refusal to take a polygraph examination for this purpose would be denial of access to classified information in Special Access Programs.

We believe there is ample justification for such limited use of the polygraph as a condition of initial or continuing access to designated Special Access Programs. Under these circumstances, the polygraph examinations would be narrowly targeted at the specific dangers that justified the establishment of the Special Access Programs in the first place. The consequences of an employee's refusal to be examined for this purpose would be limited to a denial of access to particularly sensitive classified information.

Some transitional arrangements should be made if polygraph examinations are to be required of employees with current access to information in designated Special Access Programs. Most of these employees were not hired for or assigned to their current positions with an understanding that polygraph examinations would be a condition of employment or assignment. Such employees in the competitive service or uniformed services who object to being polygraphed should be transferred or reassigned to other government jobs with no loss in grade or pay. (Of course, this guarantee would not preclude adverse personnel actions for reasons other than the refusal to be polygraphed.)

In the future, no employee in the competitive service would be involuntarily assigned to a job requiring access to information for which the polygraph is a condition of access, unless this requirement was made known at the time the individual entered into employment. With respect to members of the uniformed services, the possibility that such a requirement might be imposed at some time in their careers would not ordinarily be known upon their entry into service. However, as a practical matter, servicemen who object to polygraph examinations can ordinarily avoid being assigned to the specialized positions for which this requirement would be applied.

(3) Criminal Investigations

Polygraph examinations of federal employees should continue to be used in criminal investigations on the same terms as in the past. That is, the agency with criminal investigative jurisdiction may ask the employee to consent to an examination after an opportunity to consult with legal counsel. Use of the polygraph technique should be limited to the small portion of cases where it can make an appropriate contribution. An example would be a situation where there is a direct conflict of credibility among a small number of suspects and there is not other means to resolve the case.

The polygraph must not be used in criminal investigation in a manner that violates the employee's Fifth Amendment privilege against compulsory self-incrimination. Therefore, an employee's refusal to consent to a polygraph examination for these purposes cannot form the basis of an adverse personnel action that would result in loss of employment or any other substantial economic deprivation.

(4) In Administrative Investigations

The polygraph may be used in administrative investigations of suspected misconduct on the same consensual basis as it is used in criminal investigations, as described above. However, the Fifth Amendment does not prevent an agency from requiring an employee to submit to a polygraph examination related to his fitness for continuing employment. In situations where there is no apparent violation of criminal law or where the agency with criminal prosecutive authority has decided for other reasons that no prosecution will be undertaken, an employee can legally be required to submit to a polygraph examination in an appropriate case and disciplined or discharged for refusal.

The extent to which employees will be requested or required to submit

to polygraph examinations in such administrative investigations is a matter for the sound discretion of the agency head. The decision should be made on the basis of the facts and circumstances of each case. However, the following minimum criteria should be met in each case:

-- The suspected misconduct must be a serious offense affecting national security or the integrity of the employee's official conduct. The unauthorized disclosure of properly-classified information can be such an offense.

-- Other information or means of investigation must have produced a substantial objective basis for seeking to examine the employee.

-- There must be no other reasonable means of resolving the matter.

-- The examination must be limited to the circumstances of the suspected misconduct and cannot include irrelevant questions about the employee's personal life or other matters.

-- The polygraph examiner's assessment cannot be conclusive of the matter under investigation but instead must be considered in the context of all other available information.

An employee who is requested to take a polygraph examination in an administrative investigation of suspected misconduct should be advised on the potential consequences of refusal. For employees in the competitive service or uniformed services, refusal of a request to be examined for such purposes should not be considered as evidence to support an adverse personnel action based upon a conclusion that the employee is guilty of the suspected misconduct. However, agency heads may draw an adverse inference of culpability in appropriate cases, for political appointees and other members of the excepted service.

In this regard, it should be noted that the following are not considered to be adverse personnel actions requiring specific evidence of misconduct:

-- denial of access to classified information generally, or to particular categories; or

-- transfer or reassignment to another job at the same grade and pay.

An employee's refusal of a request to be polygraphed in an investigation of suspected misconduct may be taken into account in any of the foregoing determinations. However, such refusals should not themselves be conclusive of any of these determinations.

The minimum requirements of NSDD-84 are satisfied if an agency has the ability to consider an employee's refusal of a request to be polygraphed as a factor in deciding whether the employee should be denied access to classified information or transferred to a less sensitive job.

In certain cases, agency heads or their delegates may also order--not

simply request--an employee to take a polygraph examination in connection with an administrative investigation of misconduct. Refusal of such an order can itself form the basis for administrative sanctions without requiring any assessment of culpability for the misconduct being investigated. In such situations, the employee is discharged or disciplined for insubordination or impeding the conduct of an official investigation.

There is no requirement that the authority to order an employee to submit to a polygraph examination in this situation be exercised in order to satisfy the requirements of NSDD-84. Indeed, this authority should be rarely used. A major concern in this regard is that information acquired in a polygraph examination to which an employee suspected of criminal misconduct is ordered to submit, and the fruits of that information, cannot be used as evidence in a criminal prosecution. Thus, the order to be examined is tantamount to a grant of "use" immunity, which should only be made in consultation with the appropriate criminal prosecutive authority.

In general, we believe the power to order a polygraph examination under these circumstances, and to discharge employees for refusal, should be limited to the following:

-- employees of CIA, NSA, and others who are advised that polygraph examinations are a condition of employment;

-- political appointees;

-- other members of the excepted service.

This authority would not extend to members of the uniformed services or the competitive service, except those who are advised prior to being hired that polygraph examinations are a condition of employment.

We believe that the foregoing policy on use of the polygraph in administrative investigations reflects a reasonable balancing of competing interests. Use of the polygraph is limited to a small number of cases where the government's interest in resolving the matter outweighs the disadvantages of using the technique. The risk of a "false positive" result would be minimized, because the examination result alone could not form the basis for any adverse action. For most federal employees, the consequences of refusing to be examined are limited so as not to involve loss of employment or demotion.

Conclusion

The Administration's policy is to carefully restrict the use of polygraph examinations for federal employees and applicants. Use of this technique should be limited to situations where it clearly serves an overriding national interest. Assuring that employees in our intelligence agencies and similarly sensitive positions meet appropriate security standards is such an overriding national interest.

Our policy on polygraph use takes into account other competing interests and includes numerous safeguards. In particular, federal employees in the competitive service and uniformed services are provided with substantial additional assurances that they cannot be fired or demoted for refusing to take a polygraph examination.

In 1976 this Committee recommended "that the use of polygraphs and similar devices be discontinued by all Government agencies for all purposes." 35/ We believe that such a policy of absolute rejection of the polygraph is unsupported by the technical studies and unwise as a matter of policy. It fails to give sufficient weight to the national interest in taking every reasonable step to assure the security of our intelligence services and sensitive classified information. Therefore, this Administration supports a policy of carefully limited use of the polygraph, as described in my Statement.

Footnotes

1/ Espionage Laws and Leaks, Hearings Before the Subcommittee on Legislation of the House Permanent Select Committee on Intelligence, 96th Cong., 1st Sess. 1 (remarks of Hon. Morgan F. Murphy) (Jan. 24, 1979).

2/ Id. at 2 (remarks of Hon. Robert McClory).

3/ National Security Secrets and the Administration of Justice, Report of the Senate Select Committee on Intelligence, Subcommittee on Secrecy and Disclosure, 95th Cong., 2d Sess. 3 (Comm. Print 1978).

4/ 48 Fed. Reg. 40,849 (Sep. 9, 1983).

5/ 444 U.S. 507 (1980) (per curiam).

6/ McGehee v. Casey, No. 81-2233 (D.C. Cir., Oct. 4, 1983).

7/ Many of these studies have been reviewed by the Office of Technology Assessment in preparing a report at the request of this Committee. A number of Executive agencies provided comments on a draft of this report, but the final version had not been made available at the time this Statement was prepared. The Department of Defense has prepared its own survey of polygraph studies entitled, "The Accuracy and Utility of Polygraph Testing" (Garwood & Ansley, eds.), which is being provided to the Committee.

8/ 293 F. 1013 (D.C. Cir. 1923).

9/ See, e.g., 22 C. Wright & K. Graham, Federal Practice and Procedures §5169 (1978).

10/ See, e.g., United States v. Clark, 598 F.2d 994 (5th Cir. 1979).

11/ See, e.g., United States v. Glover, 596 F.2d 857, 867 (9th Cir. 1979), cert. denied, 444 U.S. 857; United States v. Kampiles, 609 F.2d 1233, 1244 (7th Cir. 1979), cert. denied, 446 U.S. 954.

12/ See State v. Dean, 103 Wis. 2d 228, 264-65, 307 N.W. 2d 628, 646 n. 18 (1981) and cases cited therein.

13/ See Annot., 53 A.L.R. 3d 1005 (1973 & Supp. 1983), and cases cited therein.

14/ State v. Dorsey, 88 N.M. 184, 539 P.2d 204 (1975).

15/ See, e.g., United States v. Stromberg, 179 F.Supp. 278, 280 (S.D.N.Y. 1959).

16/ The Use of Polygraphs and Similar Devices by Federal Agencies, Hearings Before the House Committee on Government Operations, 93d Cong., 2d Sess. (1974), at p. 419.

17/ See, e.g., People v. Lara, 528 P.2d 365, 370, 12 Cal. 3d 903, 909, 117 Cal. Rptr. 549, 554 (1974).

18/ Wyrick v. Fields, 103 S.Ct. 394, 396 (1982) (per curiam).

19/ Schmerber v. California, 384 U.S. 757, 764 (1966) (dictim) (unlike blood test, polygraph examination elicits testimonial response); South Dakota v. Neville, 101 S.Ct. 916, 922 n.12 (1983) (dictum).

20/ See, e.g., Lefkowitz v. Turley, 414 U.S. 70, 84 (1974); Sanitation Men v. Commissioner, 392 U.S. 280, 285 (1968).

21/ Memorandum of Larry A. Hammond, Acting Assistant Attorney General, Office of Legal Counsel, Feb. 22, 1980, at 8-12.

22/ Garrity v. New Jersey, 385 U.S. 493, 500 (1967).

23/ Baxter v. Palmigiano, 425 U.S. 308, 318 (1976); Hoover v. Knight, 678 F.2d 578, 582 (5th Cir. 1982); United States v. Indorato, 628 F.2d 711, 716 (1st Cir.), cert. denied, 449 U.S. 1016 (1980).

24/ See, e.g., 50 U.S.C. §§ 403(c),(d), 403g, 831-33; E.O. 12333, §§ 1.5(g), (h), 1.8(h), 1.12(b)(10).

25/ See, e.g., S. 2156, 92d Cong., 1st Sess. (1971); H.Rep. No. 94-795, 94th Cong., 2d Sess. 46 (1976).

26/ Federal Personnel Manual, Chapter 736, appendix D.

27/ See, e.g., 5 C.F.R. 735.201a(c), 735.201a(f), 735.209.

28/ Use of the polygraph in a particular case would have to be supported by an adequate record establishing: the reason for its use, including the existence of other independent evidence indicating possible involvement of the individual and corroborating the results, if any, of the polygraph; the expected accurary of the technology; and the qualifications of the examiner. See the Memorandum of Larry A. Hammond, supra, note 21, at 10-11; Memorandum of Theodore B. Olson, Assistant Attorney General, Office of Legal Counsel, August 22, 1983, at 8.

29/ Compare Seattle Police Officers Guild v. City of Seattle, 80 Wash. 2d 307, 319-20 494 P.2d 485, 493 (1972) (permitting discharge) with Farmer v. City of Fort Lauderdale, 427 So.2d 187 (Fla. 1983) (not permitting discharge); Annot., 15 A.L.R. 4th 1207 (1982), and cases cited therein. See also Gulden v. McCorkle, 680 F.2d 1070, 1076 (5th Cir. 1982), cert. denied, 103 S.Ct. 1194 (1983) (permitting discharge).

30/ Meier v. Department of the Interior, 3 MSPB 341, 344-46(1980). The

Merit Systems Protection Board made clear in its decision in Meier that its approval of the presiding official's admission of polygraph results was limited to the facts of the case before it, and that the MSPB did not "imply that taking a polygraph examination will be required under any circumstances or that the results of such a test must be accepted into evidence or accorded any specific weight in the final decision." Meier v. Department of the Interior, 3 MSPB at 344; see also Flores v. Department of Labor, 82 FMSR ¶5407 (Sep. 13, 1982).

31/ Meier v. Department of the Interior, 3 MSPB 341, 344 (1980). But cf. South Dakota v. Neville, 101 S.Ct. 916 (1983) (refusal to take mandatory blood alcohol test is admissible evidence in criminal trial).

32/ Memorandum of Theodore B. Olson, supra, note 29.

33/ Statement of David T. Lykken, Hearings before the Subcommittee on Oversight of the Permanent Select Committee of Intelligence, note 1 supra, p. 164.

34/ Even critics recognize the utility of this aspect of using polygraph examinations in pre-employment screening. See Lykken, supra, note 34.

35/ H. Rep. No. 94-795, 94th Cong., 2d Sess. 46 (1976).

* * * * * *

Richard K. Willard

Exhibit A

THE WHITE HOUSE
Washington

August 30, 1983

MEMORANDUM FOR FEDERAL EMPLOYEES

SUBJECT: Unauthorized Disclosure of Classified Information

Recent unauthorized disclosures of classified information concerning our diplomatic, military, and intelligence activities threaten our ability to carry out national security policy. I have issued a directive detailing procedures to curb these disclosures and to streamline procedures for investigating them. However, unauthorized disclosures are so harmful to our national security that I wish to underscore to each of you the seriousness with which I view them.

The unauthorized disclosure of our Nation's classified information by those entrusted with its protection is improper, unethical, and plain wrong. This kind of unauthorized disclosure is more than a so-called "leak"--it is illegal. The Attorney General has been asked to investigate a number of recent disclosures of classified information. Let me make it clear that we intend to take appropriate administrative action against any Federal employee found to have engaged in unauthorized disclosure of classified information, regardless of rank or position. Where circumstances warrant, cases will also be referred for criminal prosecution.

The American people have placed a special trust and confidence in each of us to protect their property with which we are entrusted, including classified information. They expect us to protect fully the national security secrets used to protect them in a dangerous and difficult world. All of us have taken an oath faithfully to discharge our duties as public servants, an oath that is violated when unauthorized disclosures of classified information are made.

Secrecy in national security matters is a necessity in this world. Each of us, as we carry out our individual duties, recognizes that certain matters require confidentiality. We must be able to carry out diplomacy with friends and foes on a confidential basis; peace often quite literally depends on it--and this includes our efforts to reduce the threat of nuclear war.

We must also be able to protect our military forces from present or potential adversaries. From the time of the Founding Fathers, we have accepted the need to protect military secrets. Nuclear dangers, terrorism, and aggression similarly demand that we must be able to gather intelligence information about these dangers--and our sources of this information must be protected if we are to continue to receive it. Even in peacetime, lives depend on our ability to keep certain matters secret.

As public servants, we have no legitimate excuse for resorting to these unauthorized disclosures. There are other means available to express ourselves:

-- We make every effort to keep the Congress and the people informed about national security policies and actions. Only a fraction of information concerning national security policy must be classified.

-- We have mechanisms for presenting alternative views and opinions within our government.

-- Established procedures exist for declassifying material and for downgrading information that may be overclassified.

-- Workable procedures also exist for reporting wrongdoing or illegalities, both to the appropriate Executive Branch offices and to the Congress.

Finally, each of us has the right to leave our position of trust and criticize our government and its policies, if that is what our conscience dictates. What we do not have is the right to damage our country by giving away its necessary secrets.

We are as a Nation an open and trusting people, with a proud tradition of free speech, robust debate, and the right to disagree strongly over all national policies. No one would ever want to change that. But we are also a mature and disciplined people who understand the need for responsible action. As servants of the people, we in the Federal Government must understand the duty we have to those who place their trust in us. I ask each of you to join me in redoubling our efforts to protect that trust.

RONALD REAGAN
(signed)

Exhibit B

Safeguarding National Security Information

As stated in Executive Order 12356, only that information whose disclosure would harm the national security interests of the United States may be classified. Every effort should be made to declassify information that no longer requires protection in the interest of national security.

At the same time, however, safeguarding against unlawful disclosures of properly classified information is a matter of grave concern and high priority for this Administration. In addition to the requirements set forth in Executive Order 12356, and based on the recommendations contained in the interdepartmental report forwarded by the Attorney General, I direct the following:

1. Each agency of the Executive Branch that originates or handles classified information shall adopt internal procedures to safeguard against unlawful disclosures of classified information. Such procedures shall at a minimum provide as follows:

a. All persons with authorized access to classified information shall be required to sign a nondisclosure agreement as a condition of access. This requirement may be implemented prospectively by agencies for which the administrative burden of compliance would otherwise be excessive.

b. All persons with authorized access to Sensitive Compartmented Information (SCI) shall be required to sign a nondisclosure agreement as a condition of access to SCI and other classified information. All such agreements must include a provision for prepublication review to assure deletion of SCI and other classified information.

c. All agreements required in paragraphs 1.a. and 1.b. must be in a form determined by the Department of Justice to be enforceable in a civil action brought by the United States. The Director, Information Security Oversight Office (ISOO), shall develop standardized forms that satisfy these requirements.

d. Appropriate policies shall be adopted to govern contacts between media representatives and agency personnel, so as to reduce the opportunity for negligent or deliberate disclosures of classified information. All persons with authorized access to classified information shall be clearly apprised of the agency's policies in this regard.

2. Each agency of the Executive Branch that originates or handles classified information shall adopt internal procedures to govern the reporting and investigation of unauthorized disclosures of such information. Such procedures shall at a minimum provide that:

a. All such disclosures that the agency considers to be seriously damaging to its mission and responsibilities shall be evaluated to ascertain the nature of the information disclosed and the extent to which it had been disseminated.

b. The agency shall conduct a preliminary internal investigation prior to or concurrently with seeking investigative assistance from other agencies.

c. The agency shall maintain records of disclosures so evaluated and investigated.

d. Agencies in the possession of classified information originating with another agency shall cooperate with the originating agency by conducting internal investigations of the unauthorized disclosure of such information.

e. Persons determined by the agency to have knowingly made such disclosures or to have refused cooperation with investigations of such unauthorized disclosures will be denied further access to classified information and subjected to other administrative sanctions as appropriate.

3. Unauthorized disclosures of classified information shall be reported to the Department of Justice and the Information Security Oversight Office, as required by statute and Executive orders. The Department of Justice shall continue to review reported unauthorized disclosures of classified information to determine whether FBI investigation is warranted. Interested departments and agencies shall be consulted in developing criteria for evaluating such matters and in determining which cases should receive investigative priority. The FBI is authorized to investigate such matters as constitute potential violations of federal criminal law, even though administrative sanctions may be sought instead of criminal prosecution.

4. Nothing in this directive is intended to modify or preclude interagency agreements between FBI and other criminal investigative agencies regarding their responsibility for conducting investigations within their own agencies or departments.

5. The Office of Personnel Management and all departments and agencies with employees having access to classified information are directed to revise existing regulations and policies, as necessary, so that employees may be required to submit to polygraph examinations, when appropriate, in the course of investigations of unauthorized disclosures of classified information. As a minimum, such regulations shall permit an agency to decide that appropriate adverse consequences will follow an employee's refusal to cooperate with a polygraph examination that is limited in scope to the circumstances of the unauthorized disclosure under investigation. Agency regulations may provide that only the head of the agency, or his delegate, is empowered to order an employee to submit to a polygraph examination. Results of polygraph examinations should not be relied upon to the exclusion of other information obtained during investigations.

6. The Attorney General, in consultation with the Director, Office of Personnel Management, is requested to establish an interdepartmental group to study the federal personnel security program and recommend appropriate revisions in existing Executive orders, regulations, and guidelines.

Richard K. Willard

Exhibit C

Sensitive Compartmented Information
Nondisclosure Agreement

An Agreement between _____ and the United States
(Name-Printed or Typed)

1. Intending to be legally bound, I hereby accept the obligations contained in this Agreement in consideration of my being granted access to information known as Sensitive Compartmented Information (SCI). I have been advised and am aware that SCI involves or derives from intelligence sources or methods and is classified or classifiable under the stardards of Executive Order 12356 or under other Executive order or statute. I understand and accept that by being granted access to SCI, special confidence and trust shall be placed in my by the United States Government.

2. I hereby acknowledge that I have received a security indoctrination concerning the nature and protection of SCI, including the procedures to be followed in ascertaining whether other persons to whom I contemplate disclosing this information have been approved for access to it, and that I understand these procedures. I understand that I may be required to sign subsequent agreements as a condition of being granted access to different categories of SCI. I further understand that all my obligations under this Agreement continue to exist whether or not I am required to sign such subsequent agreements.

3. I have been advised and am aware that direct or indirect unauthorized disclosure, unauthorized retention, or negligent handling of SCI by me could cause irreparable injury to the United States or could be used to advantage by a foreign nation. I hereby agree that I will never divulge such information unless I have officially verified that the recipient has been properly authorized by the United States Government to receive it or I have been given prior written notice of authorization from the United States Government Department or Agency (hereinafter Department or Agency) last granting me either a security clearance or an SCI access approval that such disclosure is permitted.

. I further understand that I am obligated to comply with laws and regulations that prohibit the unauthorized disclosures of classified information. As used in this Agreement, classified information is information that is classified under the standards of E.O. 12356, or under any other executive order or statute that prohibits the unauthorized disclosure of information in the interest of national security.

. In consideration of being granted access to SCI and of being assigned or retained in a position of special confidence and trust requiring access to SCI and other classified information, I hereby agree to submit for security review by the Department or Agency last granting me either a security clearance or an SCI access approval all materials, including works of fiction, that I contemplate disclosing to any person not authorized to have such information, or that I have prepared for public disclosure, which contain or purport to contain:

(a) any SCI, any description of activities that produce or relate to

SCI, or any information derived from SCI;

(b) any classified information from intelligence reports or estimates; or

(c) any information concerning intelligence activities, sources or methods.

I understand and agree that my obligation to submit such information and materials for review applies during the course of my access to SCI and at all times thereafter. However, I am not required to submit for review any such materials that exclusively contain information lawfully obtained by me at a time when I have no employment, contract or other relationship with the United States Government, and which are to be published at such time.

6. I agree to make the submissions described in paragraph 5 prior to discussing the information or materials with, or showing them to anyone who is not authorized to have access to such information. I further agree that I will not disclose such information or materials unless I have officially verified that the recipient has been properly authorized by the United States Government to receive it or I have been given written authorization from the Department or Agency last granting me either a security clearance or an SCI access approval that such disclosure is permitted.

7. I understand that the purpose of the review described in paragraph 5 is to give the United States a reasonable opportunity to determine whether the information or materials submitted pursuant to paragraph 5 set forth any SCI or other information that is subject to classification under E.O. 12356 or under any other Executive order or statute that prohibits the unauthorized disclosure of information in the interest of national security. I further understand that the Department or Agency to which I have submitted materials will act upon them coordinating with the Intelligence Community or other agencies when appropriate, and substantively respond to me within 30 working days from date of receipt.

8. I have been advised and am aware that any breach of this Agreement may result in the termination of any security clearance and SCI access approvals that I may hold; removal from any position of special confidence and trust requiring such clearances or access approvals; and the termination of my employment or other relationships with the Departments or Agencies that granted my security clearances or SCI access approvals. In addition, I have been advised and am aware that any unauthorized disclosure of SCI or other classified information by me may constitute a violation or violations of United States criminal laws, including the provisions of Sections 641, 793, 794, 798, and 952, Title 18, United States Code, the provisions of Section 783(b), Title 50, United States Code and the provisions of the Intelligence Identities Protection Act of 1982. I recognize that nothing in this Agreement constitutes a waiver by the United States to the right to prosecute me for any statutory violation.

9. I hereby assign to the United States Government all royalties, remunerations, and emolument that have resulted, will result, or may result from any disclosure, publication, or revelation not consistent with with the terms of this Agreement.

10. I understand that the United States Government may seek any remedy available to it to enforce this Agreement including, but not limited to, application for a court order prohibiting disclosure of information in breach of this Agreement.

11. I understand that all information to which I may obtain access by signing this Agreement is now and will forever remain the property of the United States Government. I do not now, nor will I ever, possess any right, interest, title, or claim whatsoever to such information. I agree that I shall return all materials which have or may come into my possession or for which I am responsible because of such access, upon demand by an authorized representative of the United States Government or upon the conclusion of my employment or other relationship with the Department or Agency that last granted me either a security clearance or an SCI access approval. If I do not return such materials upon request, I understand that this may be a violation of Section 793, Title 18, United States Code, a United States criminal law.

12. Unless and until I am released in writing by an authorized representative of the United States Government, I understand that all conditions and obligations imposed upon me by this Agreement apply during the time I am granted access to SCI and at all times thereafter.

13. Each provision of this Agreement is severable. If a court should find any provision of this Agreement to be unenforceable, all other provision of this Agreement shall remain in full force and effect.

14. I have read this Agreement carefully and my questions, if any, have been answered to my satisfaction. I acknowledge that the briefing officer has made available to me Sections 641, 793, 794, 798, and 952 of Title 18, United States Code, Section 783(b) of Title 50, United States Code, the Intelligence Identities Protection Act of 1982, and Executive Order 12356 so that I may read them at this time, if I so choose.

15. I make this Agreement without mental reservation or purpose of evasion.

Signature _____ Date _____

Social Security Number _____ Organization _____
(see notice below)

The execution of this Agreement was witnessed by the undersigned, who, on behalf of the United States Government, agreed to its terms and accepted it as a prior condition of authorizing access to Sensitive Compartmented Information.

WITNESS and ACCEPTANCE:

Signature _____ Date _____

Organization _____

Notice: The Privacy Act, 5 U.S.C. 552a, requires that federal agencies inform individuals, at the time information is solicited from them, whether the disclosure is mandatory or voluntary, by what authority such information is solicited, and what uses will be made of the information. You are hereby advised that authority for soliciting your Social Security

Number (SSN) is Executive Order 9397. Your SSN will be used to identify you precisely when it is necessary to 1) certify that you have access to the information indicated above, 2) determine that your access to the information indicated has terminated, or 3) certify that you have witnessed a briefing or debriefing. Although disclosure of your SSN is not mandatory, your failure to do so may impede the processing of such certifications or determinations.

Richard K. Willard

Exhibit D

U.S. Department of Justice
Civil Division

Deputy Assistant Attorney General Washington, D.C. 20530
July 19, 1983

Mr. Steven Garfinkel
Director, Information Security Oversight Office
General Services Administration
Washington, D.C. 20403

Dear Mr. Garfinkel:

Your letter to the Attorney General dated July 1, 1983, requests that the Department of Justice review two nondisclosure agreements drafted pursuant to National Security Decision Directive 84, entitled "Safeguarding National Security Information) (referred to herein as NSDD-84), which was signed by the President on March 11, 1983.

Paragraph 1.a. of NSDD-84 requires all persons with authorized access to classified information to sign a nondisclosure agreement as a condition of access. Paragraph 1.b. imposes the same requirement on persons with authorized access to Sensitive Compartmented Information (SCI) and requires, in addition, that such nondisclosure agreements "include a provision for prepublication review to assure deletion of SCI and other classified information." Paragraph 1.c. provides that the agreements required in paragraphs 1.a. and 1.b. must be in a form determined by the Department of Justice to be enforceable in a civil action brought by the United States.

We understand that the draft agreements transmitted with your letter were prepared pursuant to the provision in paragraph 1.c. of NSDD-84 that your office develop standardized forms to satisfy the requirements of the directive. We also understand that use of these forms will be mandatory for each agency of the Executive Branch that originates or handles classified information, unless the National Security Council grants permission to use an alternative form of agreement that has been approved by your office and the Justice Department.

Classified Information Nondisclosure Agreement

The essence of the proposed Classified Information Nondisclosure Agreement is an undertaking by the person receiving access to classified information never to disclose such information in an unauthorized manner. This undertaking is consistent with the provisions of Executive Order 12356, as well as various statutes and other regulations that prohibit the unauthorized disclosure of classified information. In addition, government employees and others who are entrusted with classified information have a fiduciary obligation to protect it from unauthorized disclosure. See Snepp v. United States, 444 U.S. 507, 511 n.6, 515 n.11 (1980).

The protection of national security information is a primary and fundamental constitutional responsibility of the President that derives from his responsibilities as Chief Executive, Commander-in-Chief, and the principal instrument of United States foreign policy. Agreements to preserve the secrecy of classified information are an appropriate method for the President to discharge these constitutional responsibilities. United States v. Marchetti, 466 F.2d 1309, 1315-16 (4th Cir.), cert. denied, 409 U.S. 1063 (1972); cf. Snepp v. United States, supra, 444 U.S. at 509 n.3 (agreement serves "compelling interest" of Government in safeguarding national security information.) These same cases also rely upon the statutory authority of the Director of Central Intelligence to protect "intelligence sources and methods from unauthorized disclosure." 50 U.S.C. §403(d)(3). However, the agreements sustained in Marchetti and Snepp were not limited to information concerning intelligence sources and methods but included promises never to disclose any classified information. Therefore, we believe that the President may require the signing of such agreements as a condition of access to classified information.

SCI Nondisclosure Agreement

The proposed SCI Nondisclosure Agreement has the same basic terms as the Classified Information Nondisclosure Agreement discussed above. SCI is a category of classified information that is subject to special access and handling requirements because it involves or derives from particularly sensitive intelligence sources and methods. The power to require signing such an agreement as a condition of access to SCI is thus supported by the statutory authority of the Director of Central Intelligence to protect intelligence sources and methods, 50 U.S.C. §403(d)(3), as well as the more fundamental constitutional responsibilities of the President regarding national security.

The proposed SCI Nondisclosure Agreement includes provisions for the Government to conduct prepublication review of certain writings by persons who have signed the agreement. The prepublication review provisions of the proposed agreement are similar to the agreement found by the Supreme Court to be enforceable in Snepp v. United States, supra. See also Alfred A. Knopf, Inc. v. Colby, 509 F.2d 1362 (4th Cir.), cert. denied, 421 U.S. 992 (1975); United States v. Marchetti, supra: Agee v. CIA, 500 F. Supp. 506 (D.D.C. 1980).

The rationale of the above-cited cases supports the inclusion of prepublication review provisions in agreements that extend beyond CIA to include other persons with authorized access to SCI. Neither the statutory authority of the Director of Central Intelligence nor the constitutional responsibilities of the President are limited to CIA and its employees. Moreover, a high degree of trust, which creates a fiduciary obligation on the part of CIA employees, would also be involved for government officials outside CIA who are entrusted with equally sensitive information such as SCI.

Paragraph 5 of the proposed SCI Nondisclosure Agreement defines the scope of materials required to be submitted for prepublication review. In certain respects, this provision is narrower than the agreement at issue in Snepp. As in the Snepp agreement, however, certain materials must be submitted for review even if they are not thought to contain classified

information. The Supreme Court in Snepp upheld the validity of such a requirement. 444 U.S. at 511-13.

Among the categories of materials required in paragraph 5 to be submitted for prepublication review is "(c) any information concerning intelligence activities, sources or methods." This category is not limited to classified information as such, but includes any information that is required to be considered for classification pursuant to Executive Order 12356, §1.3(a)(4). We believe that agencies using the proposed SCI Nondisclosure Agreement should include in their implementing instructions some definition of the term "intelligence activities," to include at least a reference to the definition contained in Executive Order 12333, §3.4 (e).

Once material is submitted for prepublication review, there is no authority in the proposed agreement for the Government to delete unclassified information. However, any information that is subject to classification may be deleted pursuant to paragraph 7, even if it does not pertain to SCI or other intelligence matters. See Alfred A. Knopf, Inc. v. Colby, supra, 509 F.2d at 1368-69.

Conclusion

We have reviewed the specific provisions of the two draft agreements transmitted with your letter and have concluded that each would be enforceable in accordance with its terms in civil litigation initiated by the United States.

Sincerely,

RICHARD K. WILLARD (signed)

Deputy Assistant Attorney General

Exhibit E

DEPARTMENT OF JUSTICE

ORDER
[DOJ 2620.8]

Subject: EMPLOYEE OBLIGATIONS TO PROTECT CLASSIFIED INFORMATION AND SUBMIT TO PREPUBLICATION REVIEW

---

1. PURPOSE. The purpose of this order is to explain and clarify Department of Justice (DOJ) policies concerning implementation of the prepublication review program.

2. SCOPE. This order applies to all persons granted access to classified information in the course of their employment at the DOJ and DOJ contractors granted such access.

3. AUTHORITY.

   a. Executive Order 12356, "National Security Information."

   b. National Security Decision Directive-84 entitled "Safeguarding National Security Information."

   c. 28 C.F.R. 0.75(p).

4. POLICY. All persons granted access to classified information in the course of their employment at the DOJ are required to safeguard that information from unauthorized disclosure. This nondisclosure obligation is imposed by statutes, regulations, access agreements, and the fiduciary relationships of the persons who are entrusted with classified information in the performance of their duties. The nondisclosure obligation continues after DOJ employment terminates.

   As an additional means of preventing unlawful disclosures of classified information, the President has directed that all persons with authorized access to Sensitive Compartmented Information (SCI) be required to sign nondisclosure agreements containing a provision for prepublication review to assure deletion of SCI and other classified information. SCI is information that not only is classified for national security reasons as Top Secret, Secret, or Confidential, but also is subject to special access and handling requirements because it involves or derives from particularly sensitive intelligence sources and methods.

5. RESPONSIBILITIES.

   a. The prepublication review provision requires that DOJ employees granted access to SCI submit certain material to the Department, whether prepared during or subsequent to DOJ employment, prior to its publication to provide an opportunity for determining whether an

---

Distribution: BUR/H-1
OBD/F-2 OBD/H-1

Initiated by: Security Staff
Justice Management Division

unauthorized disclosure of SCI or other classified information would occur as a consequence of its publication.

The obligations not to disclose classified information and to comply with agreements requiring prepublication review have been held by the Supreme Court to be enforceable in civil litigation. Snepp v. United States, 444 U.S. 507 (1980).

b. It must be recognized at the outset that it is not possible to anticipate each and every question that may arise. The Department will endeavor to respond, however, as quickly as possible to specific inquiries by present and former employees concerning whether specific materials require prepublication review. Present and former employees are invited to discuss their plans for public disclosures of information that may be subject to these obligations with authorized Department representatives at an early stage, or as soon as circumstances indicate these policies must be considered. All questions concerning these obligations should be addressed to the Counsel for Intelligence Policy, Room 6325, U.S. Department of Justice, 10th & Constitution Avenue, N.W., Washington, D.C. 20530. The official views of the Department of whether specific materials require prepublication review may only be expressed by the Counsel for Intelligence Policy and persons should not act in reliance upon the views of other Department personnel.

c. Employees with access to SCI will be required to sign agreements providing for prepublication review. Prepublication review is required only as expressly provided for in an agreement. However, all persons who have had access to classified information have an obligation to avoid unauthorized disclosures of such information and are subject to enforcement actions if they disclose classified information in an unauthorized manner. Therefore, present or former employees are encouraged voluntarily to submit materials for prepublication review if they believe that such material may contain classified information even if such submission is not required by a prepublication review agreement. Where there is any doubt, present and former employees are urged to err on the side of prepublication review to avoid unauthorized disclosures and for their own protection.

d. Present or former employees who have signed agreements providing for prepublication review are required to submit any material prepared for disclosure to others that contains or purports to contain:

(1) any SCI, any description of activities that produce or relate to SCI, or any information derived from SCI;

(2) any classified information from intelligence reports or estimates; or

(3) any information concerning intelligence activities, sources or methods.

The term "intelligence activities" in paragraph 5.d.(3) means all activities that agencies within the Intelligence Community are authorized to conduct pursuant to Executive Order 12333. However,

there is no requirement to submit for review any materials that exclusively contain information lawfully obtained at a time when the author has no employment, contract, or other relationship with the United States Government and which are to be published at such time.

e. A person's obligation to submit material for prepublication review remains identical whether such person actually prepares the material or causes or assists another person, such as a ghost writer, spouse or friend, or editor in preparing the material. Material described in paragraph 5.d must be submitted for prepublication review prior to discussing it with or showing it to a publisher, co-author, or any other person who is not authorized to have access to it. In this regard, it should be noted that a failure to submit such material for prepublication review constitutes a breach of the obligation and exposes the author to remedial action even in cases where the published material does not actually contain SCI or classified information. See Snepp v. United States, supra.

f. The requirement to submit information or materials for prepublication review is not limited to any particular type of material or disclosure. Written materials include not only books but all other forms of written materials intended for public disclosure, such as (but not limited to) newspaper columns, magazine articles, letters to the editor, book reviews, pamphlets, and scholarly papers. Because fictional treatment may convey factual information, fiction is also covered if it is based upon or reflects information described in paragraph 5.d.

g. Oral statements are also included when based upon written materials, such as an outline of the remarks. There is no requirement to prepare such material for prior review, however, unless there is reason to believe in advance that oral statements may contain SCI or other classified information. Thus, a person may participate in an oral presentation of information where there is no opportunity for prior preparation (e.g., news interview, panel discussion) unless there is reason to believe in advance that such oral expression may contain SCI or other classified information. This recognition of the problems with oral representations does not, of course, exempt present or former employees from liability for any unauthorized disclosures of SCI or classified information that may occur in the course of even extemporaneous oral expressions.

h. Material that consists solely of personal views, opinions or judgments and does not contain or imply any statement of fact that would fall within the description in paragraph 5.d is not subject to the prepublication review requirement. For example, public speeches or publication of articles on such topics as proposed legislation or foreign policy do not require prepublication review as long as the material does not directly or implicitly constitute a statement of an informational nature that falls within paragraph 5.d. Of course, in some circumstances the expression of "opinion" may imply facts and thus be of such a character as to require prior review.

i. Obviously, the purposes of prepublication review will be frustrated where the material in question already has been disseminated

to authorized persons. Comparison of the material before and after the review would reveal which items of classified information, if any, had been deleted at the Department's request. Consequently, the Department will consider these obligations to have been breached in any case, whether or not the written material is subsequently submitted to the Department for prepublication review, where it already has been circulated to publishers or reviewers or has otherwise been made available to unauthorized persons. While the Department reserves the right to review such material for purposes of mitigating damage that may result from the disclosure, such action shall not prevent the United States Government and the Department from pursuing all appropriate remedies availble under law as a consequence for the failure to submit the materials for proper review and/or any unauthorized disclosure of SCI or classified information.

j. Materials submitted for prepublication review will be reviewed solely for the purpose of identifying and preventing the disclosure of SCI and other classified information. This review will be conducted in an impartial manner without regard to whether the material is critical or favorable to the Department. No effort will be made to delete embarrassing or critical statements that are unclassified. Materials submitted to the Office of Intelligence Policy for review will be disseminated to other persons or agencies only to the extent necessary to identify classified information.

k. The Counsel for Intelligence Policy will respond substantively to prepublication review requests within 30 working days. Priority shall be given to reviewing speeches, newspaper articles, and other materials that the author seeks to publish on an expedited basis. The Counsel's decisions may be appealed to the Deputy Attorney General, who will process appeals within 15 working days. The Deputy Attorney General's decision is final and not subject to further administrative appeal. Authors who are dissatisfied with the final administrative decision may obtain judicial review either by filing an action for declaratory relief or by giving the Department notice and a reasonable opportunity (30 working days) to file a civil action seeking a court order prohibiting disclosure. Of course, until any civil action is resolved in court, employees remain under an obligation not to disclose or publish information determined by the Government to be classified.

l. Nothing in this order should be construed to alter or waive the Department's authority to seek any remedy available to it to prohibit or punish the unauthorized disclosure of classified information.

m. A former DOJ employee who subsequently receives a security clearance or SCI access approval from another department or agency is permitted to satisfy any obligation regarding prepublication review by making submissions to the department or agency that last granted the individual either a security clearance or an SCI access approval.

n. The obligations described herein as applying to DOJ employees also apply with equal force to contractors who are authorized by the Department to have access to SCI or other classified information.

WILLIAM D. VAN STAVOREN (signed)
Acting Asst. Att. General for Administration

SCIENTIFIC VALIDITY OF POLYGRAPH TESTING

By

John H. Gibbons

Prepared Statement of Dr. John H. Gibbons, Director, Office of Technological Assessment, United States Congress, before the Subcommittee on Legislation and National Security, Committee on Government Operations, U.S. House of Representatives, October 19, 1983.

Mr. Chairman, I am pleased to appear before your Subcommittee this morning to discuss the scientific validity of polygraph testing and to summarize the results of OTA's study which is now complete. With me today are Dr. Fred B. Wood, the OTA Project Director, and Dr. Leonard Saxe, of Boston University. I have included a copy of the summary and conclusions of our report as part of my written statement submitted for the report.

Mr. Chairman, I want to make ten points today.

First, the polygraph is not simply a machine or instrument that signals whether a person is being truthful or deceptive. The instrument cannot itself detect deception. A polygraph test is very complex and depends heavily on the interaction between the examiner and the individual being tested, and requires that the examiner infer deception or truthfulness based on a comparison of the person's physiological responses to various questions. The quality of the questions asked depends in part on what information the examiner already has about the person being questioned.

Second, because of the many variables involved and the absence of an established scientific theory, it is very difficult to conduct scientific research on the use of the polygraph test. Most of even the best polygraph studies reviewed by OTA suffered from various problems in research methodology. In laboratory research, it is very difficult to simulate real world conditions. And in field studies, it is hard to know what the objective truth really is.

Third, no overall measure or single, simple judgment of polygraph testing validity can be established based on existing scientific evidence. The amount and quality of the evidence available depends on the particular application. Some applications, primarily the use of the polygraph in criminal investigations, have been fairly heavily researched. Others, such as polygraph use in pre-employment screening, have had very little research attention.

Fourth, it is only in the area of criminal investigations that OTA found meaningful evidence of polygraph validity. For this use, the cumulative research evidence suggests that the polygraph detects deception better than chance, but with significant error rates.

However, even here there is a wide divergence in the results of relevant research. Six prior research reviews showed average validity ranging from a low of 64 percent to a high of 98 percent. OTA's own review of 28 studies meeting minimum acceptable scientific criteria found that, for example, correct guilty detections actually ranged from 17 to 100 percent.

Fifth, the further one gets away from the conditions of a criminal investigation, the weaker the scientific evidence for polygraph validity.

Sixth, there is no scientific evidence to establish the validity of polygraph testing for screening a large number of people in connection with the investigation of unauthorized disclosures. There has been no research on this type of application. And the screening conditions are likely to vary significantly from the typical criminal situation where the polygraph is used only after a full conventional investigation and a prime suspect has been identified.

Seventh, the available research evidence is also inadequate to establish polygraph validity when used in personnel security screening--whether that be pre-employment, pre-clearance, or aperiodic. OTA identified a total of only 4 relevant studies (one by DOD). But none of these studies specifically research the DOD proposals, and all had serious limitations in study design.

Eighth, OTA recognizes that NSA and CIA believe that the polygraph is a useful screening tool, but it is important to note that both use it primarily as part of a broader interrogation or investigation. Nonetheless, OTA concluded that the available research evidence does nots establish the scientific validity of the polygraph for this purpose. In comments to OTA, the CIA agreed that the cumulative unclassified research evidence reviewed by OTA is not directly relevant to their situation.

Ninth, there is a legitimate concern that DOD use of the polygraph for screening purposes may incorrectly identify significant numbers of innocent persons as deceptive. This could result in misdirecting DOD security resources, and, even if errors are caught and corrected, could have an adverse impact on employee morale.

Tenth, as long as the polygraph is used by Federal agencies even at current levels, which are already triple that of 10 years ago, there is a need for further research. Research on polygraph countermeasures is particularly important. If persons can be trained to beat the polygraph, then increased polygraph use could lead to a false sense of security. There is also a need for research on the scientific theory of the polygraph.

Mr. Chairman, this concludes my opening statement. I would be happy to answer any question.

* * * * * *

SCIENTIFIC VALIDITY OF POLYGRAPH TESTING: A REVIEW AND EVALUATION - A TECHNICAL MEMORANDUM

Prepared By
The Office of Technology Assessment
Congress of the United States*

SUMMARY

Introduction

The primary purpose of the OTA study has been to review and evaluate the current scientific evidence about the validity of polygraph testing. This study focus reflects the February 3, 1983, letter of request from the Committee on Government Operations, U.S. House of Representatives, and the need to complete the study in a manner that is timely to congressional consideration of the polygraph aspects of the President's National Security Decision Directive-84 (NSDD-84) and proposed revisions to Department of Defense Directive (DOD) 5210.48.

The OTA study has been limited to a critical review and evaluation of prior research and has not involved the conduct of any new experimental research. The study has also not considered, in detail, polygraph issues such as utility, ethics, privacy, and constitutional rights. The study, instead, focuses on the nature and application of polygraph tests, scientific controversy over polygraph tsting, data from field and simulation studies, and factors that affect test validity.

Federal Polygraph Use

OTA found that Federal Government polygraph use has more than tripled over the last 10 years, with about 23,000 exams conducted in 1982. Current use now exceeds the previous known peak level of use (about 20,000 exams) in 1963. In all Federal agencies except the National Security Agency (NSA) and the Central Intelligence Agency (CIA), more than 90 percent of polygraph testing in 1982 was for criminal investigation. Only the NSA and CIA make significant use of the polygraph for personal security screening--preemployment, preclearance, or aperiodic--in establishing initial and continuing eligibility for access to highly classified information. However, NSA accounted for almost half of all Federal polygraph examinations administered in 1982. Federal agencies at present make only very limited use of the polygraph for investigation of unauthorized disclosure of sensitive or classified information--261 examinations (excluding the NSA and CIA) for this purpose over the 1980-82 period.

---

* Prepared at the request of Representative Jack Brooks (D-Texas), Chairman, House Committee on Government Operations. The principal author is Leonard Saxe of Boston University. He was assisted by Denise Dougherty, Theodore Cross, Jack Langenbrunner, Katherine Locke.

DOD Directive 5210.48 and NSDD-84

The March 1983 draft proposed revisions to the DOD polygraph regulations (5210.48) would authorize the use of the polygraph to determine initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (Sensitive Compartmented Information and/or special access). The use of the polygraph in determining the continuing eligibility would be on an aperiodic (i.e., irregular) basis.

Also, the proposed revisions to DOD 5210.48 provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may, after consideration of other relevant factors, result in adverse consequences for the individual. Adverse consequences are defined to include nonselection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position. Technically, these expanded uses of the polygraph are considered to be part of DOD personnel security investigations.

NSDD-84, issued by the President on March 11, 1983, authorized agencies and departments to require employees to take a polygraph examination in the course of investigations of unauthorized disclosures of classified information. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences such as administrative sanctions and denial of security clearance, even when a person is not subject to a criminal investigation.

Thus, the combined effect of NSDD-84 and the DOD proposals would be to authorize substantially expanded use of the polygraph for purposes of personnel security screening and unauthorized disclosure investigations.

Polygraph Validity

In 1965 and again in 1976, the House Government Operations Committee concluded that there was not adequate evidence to establish the validity of the polygraph. OTA has assessed the research to determine the present state of scientific evidence.

OTA concluded that no overall measure or single, simple judgment of polygraph testing validity can be established Based on available scientific evidence. Validity is the extent to which polygraph testing can accurately detect truthfulness and deception.

There are two major reasons why an overall measure of validity is not possible. First, the polygraph test is, in reality, a very complex process that is much more than the instruments. Although the instrument is essentially the same for all applications, the types of individuals tested, training of the examiner, purpose of the test, and types of questions asked, among other factors, can differ substantially. A polygraph test requires that the examiner infer deception or truthfulness based on a comparison of the person's physiological responses to various questions. For example, there are differences between the testing procedures used in criminal investigations and those used in personnel security screening. Second, the research on polygrpah validity varies widely in terms of not only results, but also in research design and methodology. Thus conclusions about

scientific validity can be made only in the context of specific applications.

Findings

Personnel & Security Screening

OTA concluded that the available research evidence does not establish the scientific validity of the polygraph test for personnel security screening. OTA identified a total of 4 studies relevant to personnel security screening use (one by DOD). But none of these studies specifically assessed validity, and all had serious limitations in study design.

A 1980 survey, conducted by the Director of Central Intelligence Security Committee concluded that the polygraph was the most productive of all background investigation techniques. However, this was a utility study not a validity study, and had many qualifications.

OTA recognizes that NSA and CIA believe that the polygraph is a useful screening tool. However, OTA concluded that the available research evidence does not establish the scientific validity of the polygraph for this purpose.

In comments to OTA, the CIA agreed that the cumulative unclassified research evidence reviewed by OTA is not directly relevant to national security applications. However, the CIA does claim to have classified research to support their use of the polygraph. OTA did not review this research. No other Federal agency, including NSA, has claimed to have relevant research results that were not available for OTA review on an unclassified basis.

Criminal Investigations

OTA found meaningful scientific evidence of polygraph validity only in the area of criminal investigations. However, even here, there is a wide divergence in the results of relevant research. Six prior research reviews showed average validity ranging from a low of 64 percent to a high of 98 percent. OTA's own review of 28 studies meeting minimum acceptable scientific criteria found that, for example, correct guilty detections ranged from 17 to 100 percent. Overall, the cumulative research evidence suggest that when used in criminal investigations, the polygraph test detects deception better than chance, but with significant error rates.

However, in a typical criminal investigation, the polygraph, if used at all, is used only after prior investigation has been completed, and a prime suspect or suspects have been identified. For so-called "dragnet" screening where a large number of people would be given polygraph tests in the investigation of unauthorized disclosures, relevant research evidence falls short of establishing polygraph testing validity. There has been no direct scientific research on this application.

False Negatives/Countermeasures

Theoretically, polygraph testing is open to a large number of countermeasures, including physical movement or pressure, drugs, hypnosis,

biofeedback, and prior experience in passing an exam. The research on countermeasures has been limited and the results conflicting. OTA concluded that this lack of research is particularly significant to the extent that the polygraph is used and relied upon for national security purposes, since even a small false negative rate (guilty person tested as nondeceptive) could have very serious consequences.

False Positives

OTA concluded that the chance of incorrect identification of innocent persons as deceptive (false positives) is probably highest when the polygraph is used for screening purposes. The reason is that, in screening situations, there is usually only a very small percentage of the group being screened that might be guilty. So, in the case of DOD, there may be, perhaps, one person per 1000 engaged in unauthorized activity. Therefore, even if one assumes that the polygraph is 99 percent accurate, the laws of probability indicate that one guilty person would be correctly identified as deceptive but 10 persons would be incorrectly identified (false positives). DOD argues that there are safeguards to protect against this, and that ultimately no one would be penalized if a deceptive result could not be verified through other means. But this potential problem warrants attention.

Voluntary v. Involuntary

NSDD-84 and the DOD proposals authorize adverse consequences for refusal to take a polygraph test. Apart from the ethical and legal implications, which OTA did not address, in effect coercing persons to take a polygraph test could affect validity. It is generally recognized that, for the polygraph test to be accurate, the voluntary cooperation of the individual is important. Thus, OTA concluded that imposing penalities for not taking a test may create a de facto involuntary condition that increases the chances of invalid or inconclusive test results.

Polygraph Theory

The basic theory of polygraph testing is only partially developed. The testing process is complex and not amenable to easy understanding. The most commonly accepted theory at present is that, when the person being examined fears detection, that fear produces a measurable physiological reaction when the person responds deceptively. Thus, in this theory, the polygraph instrument is measuring the fear of detection rather than deception per se. And the examiner infers deception when the physiological response to questions about the crime or unauthorized activity is greater than the response to other questions.

A stronger theoretical base is needed for the entire range of polygraph applications. Basic polygraph research should consider the latest research from the fields of psychology, physiology, psychiatry, neuroscience, and medicine; comparison among question techniques; and measures of physiological response.

Further Research

OTA identified a need for further research on polygraph countermeasures, polygraph theory, and polygraph validity under field conditions

(for both screening and criminal investigative situations). The currently planned Federal research on countermeasures appears to be inadequate. There is no known research planned on polygraph theory. And the Army's current 10 year research program to develop a new, perhaps computerized, state-of-the-art polygraph instrument should be reevaluated to determine if research priorities and direction need adjustment. Finally, the planned FBI-Secret Service polygraph validity study needs an extensive scientific review.

Conclusions

A major reason why scientific debate over polygraph validity yields conflicting conclusions is that the validity of such a complex test is very difficult to assess and may vary widely from one application to another. The accuracy obtained in one situation or research study may not generalize to different situations or to different types of persons being tested. Scientifically acceptable research is hard to design and conduct.

Advocates of the use of polygraph testing argue that thousands of polygraphs have been conducted which substantiate its usefulness in criminal or screening situations. Claims of usefulness, however, are often dependent on information (e.g., confessions and admissions) obtained before or after the actual test, and on its perceived value as a deterrent.

The focus of the OTA study was not whether the polygraph test has been useful, but whether there is a scientific basis for its use. OTA concluded that, while there is some evidence for the validity of polygraph testing as an adjunct to criminal investigations, there is very little research or scientific evidence to establish polygraph test validity in screening situations, whether they be pre-employment, pre-clearance, random, or "dragnet." Substantial research beyond what is currently available or planned would have to be conducted in order to fully assess the scientific validity of the NSDD-84 and DOD polygraph proposals.

VARIETIES OF POLYGRAPH TESTING AND USES

Introduction

Polygraph examinations have been likened to psychological testing (cf. Keeler, 1934; Kleinmuntz & Szucko, 1981; Larson, 1932). As such, polygraph testing is best described not in the singular but, instead, as a series of tests. These tests are designed to assess truthfulness and deception in situations that range from screening job applicants to criminal investigation. Polygraph examiners, employed both within and outside Government agencies, use a variety of polygraph testing techniques, each of which has a somewhat different underlying logic and demonstrated validity.

The choice of polygraph technique depends primarily on the circumstances under which the polygraph is being used. The test of a subject who is suspected of a specific criminal activity typically involves application of a different polygraph technique than the examination of a prospective Government employee. Some variation in technique is also related to examiners' training, but such differences probably affect the way in which a technique is employed rather than which technique is used. A description of the instrument used in polygraph testing and an analysis of the types of

test situations and polygraph techniques are presented below.

Polygraph Instrument

Although there are numerous variations in testing procedures, the polygraph instrument itself is fairly standard. The polygraph measures several, usually three, physiological indicators of arousal. Changes in physiological arousal exhibited in response to a set of questions are taken to indicate deception or truthfulness. The polygraph instrument, it should be noted, is not a "lie detector" per se; that is, it does not indicate directly whether a subject is being deceptive or truthful. There is no known physiological response that is unique to deception (Lykken, 1981; Orne, Thackray & Paskewitz, 1972; Orne, 1975). Instead, a polygraph examiner obtains a subject's responses to a carefully structured set of questions, and based on the pattern of arousal responses, infers the subject's veracity. This assessment has been called the "diagnosis" of truthfulness or deception (Reic & Inbau, 1977).

In actual field testing, subjects' physiological responses are measured by a three or four channel polygraph machine that records responses on a moving chart. Usually, three different types of physiological responses are measured. The rate and depth of respiration is measured by pneumographs strapped around the chest and the abdomen. A blood pressure cuff (sphygmomanometer) placed around the bicep is used to measure cardiovascular activity. In modern polygraph instruments, sphygmomanometer readings are electronically enhanced so as to permit lower pressure in the cuff. The electrodermal response (EDR), a measure for perspiration, requires electrodes attached to the fingertips. This has also been referred to as galvanic skin response (GSR) or skin conductance response (SCR). Each of these physiological assessments has been shown to be related to physiological arousal (Brown, 1967). There is some literature to suggest that one or more of the physiological channels (EDR, in particular) is most sensitive (e.g., Orné et al., 1972). Actual field testing, however, almost always involves measurement of all three types of responses.

Types of Testing Procedures

A polygraph examination normally takes anywhere from 1 to 3 hours, although shorter or longer tests may result in a variety of circumstances. The length of an examination depends on the purpose of the examination, as well as the subject's attitude and a number of other factors. Examinations may be very short because a subject "confesses" or may be lengthy when an examiner seeks to resolve an inconsistent or inconclusive pattern of responses. The examination can be divided into at least three components: pretest interview; question procedures; and post-test interview. In addition, polygraph examiners may learn beforehand information already collected in an investigation (Barland & Raskin, 1973). The actual questioning aspect of the examination, which may be repeated three or four times, lasts no longer than a few minutes for each question set (limited, in some cases, because the blood pressure cuff can be inflated for only 10 to 12 minutes without causing the subject undue discomfort). Each aspect of a polygraph test is described below in detail. Unless specifically noted, generally-used polygraph procedures are described. Federal Government procedures are often different and, where important, such differences are noted.

The Pretest Interview

The pretest interview has been considered an indispensable component of the polygraph examination (Mullenix & Reid, 1980; Reid & Inbau; Waid & Orne, 1981). The importance of the pretest lies not only in its role to provide subjects with information about the examination and to inform them of their legal rights, but also in its ability to generate the psychological climate necessary to produce the most valid polygraph recordings (Barland & Raskin, 1973). An important purpose of the interview is to persuade a subject that the examination is professionally conducted and that any deception attempted "will be very obvious to the examiner" (Barland & Raskin, 1973. p. 424). Such instructions, it is thought, place truthful subjects at ease and increase anxiety in subjects who intend to be deceptive. Persuading subjects about the effectiveness of the examination should sharpen differences between deceptive and nondeceptive subjects in their reactions to questions about a particular incident.

The pretest also allows the examiner to assess the effect of special conditions or circumstances which might affect physiological responsiveness. Thus, for example, subjects are typically queried about medical problems and use of drugs that could influence autonomic responding. Such assessments are usually made without collecting "hard" data, such as blood samples.

Depending on which polygraph method is employed, as well as the subject's attitude and the situation under investigation, pretest interviews may take from 20 to 90 minutes (Barland & Raskin, 1973; Ben-Shakhar, Bar-Hillel, & Lieblich, 1983). One aspect of the pretest interview involves obtaining the subject's consent to be examined. Consent procedures vary depending on the nature of the interview, most importantly between criminal or preemployment polygraph tests. According to Barland and Raskin (1973), a typical polgyraph examination conducted as part of a criminal investigation requires that the examiner, as soon as he has identified and introduced himself, advises the examinee of his or her Miranda rights (or rights under the Uniform Military Code). The subject is also told that the polygraph examination is voluntary. Subjects should also be informed whether or not the examination will be observed from outside the room or recorded. These disclosures are usually included in a written form which the subject is asked to sign. According to Reid and Inbau (1977), criminal suspects may already have been informed of their Miranda rights and been asked to sign a consent form before coming to the examination room.

Applicants for employment need not be advised of their right to speak with an attorney but may, depending on local laws, require advisement about the voluntarism of the examination. In the case of such employment-related tests, along with a provision concerning voluntary consent, subjects will be told how the results of the examination will be used. Thus, for example, they may be told that a copy of the test results will be provided to the sponsor of the exam, that the subject has a right to obtain a copy of the test results, that the subject will not be asked questions concerning such areas as political activities, union affiliations, racial or religious beliefs or sexual activities unless these areas are specifically related to the issue under investigaiton (Buckley, 1983).

Examples of consent forms used in criminal investigations by Federal

agencies are shown in appendix A. The contents of Federal consent forms vary somewhat by agency, although all require that subject "voluntarily" consent to the examination. Some agencies (e.g., Department of the Treasury, 1983) indicate that the subject has the right to stop the examination at any time. Although the National Security Agency (NSA) reports that the full cooperation of the subject "is essential or the results will be inconclusive," NSA also reports (see App. B) that the polygraph examination is part of the Agency's security processing, and that failure to complete processing (which includes polygraph testing) may result in failure to be accepted for employment. As discussed more fully below (see Current Government Use), NSA conducts polygraph examinations primarily in the context of preemployment and periodic security screening; most other agencies conduct polygraph examinations for specific issue criminal investigations.

The remainder of the pretest interview also varies. In the method taught to Federal examiners at the U.S. Army Military Police School (USAMPS),* the interview focuses on questions about the subject's background: employment, family, education, health and any previous legal problems (Barland & Raskin, 1973). The examiner aims to learn enough to assess the subject's readiness for the examination and to prepare anxiety-provoking control questions, if they are to be used. The polygraph examiner then explains the polygraph technique to the subject and queries the subject in detail about the incident being investigated.

Another form of the pretest interview advocated by Reid (founder of the Reid College of Lie Detection) in criminal investigations makes use of a structured series of questions and deemphasizes gathering biographical data (Horvath, 1973; Reid & Inbau, 1977). Questions deal with matters such as the subject's suspicions about who committed the crime and the subject's feelings about the test. Questions are intended to provoke so-called "behavioral symptoms" (Reid & Inbau, 1977, p. 17-24) that are believed to be indicators of deception. These symptoms include evasiveness in answering, or complaints that one's physical disabilities will invalidate the recordings. When an examiner who uses the Reid method later makes an assessment of truthfulness, this information is used to supplement the data gathered from the physiological measures.

Whatever the format of the pretest interview, if control questions are to be used in the test, the last part of the interview will be used to design such questions and review them with the subject. In this phase, biographical and behavioral information collected earlier becomes essential. The information permits the examiner to tailor control questions to the individual subject. The process of designing control questions is complex and is discussed further in the section below which describes the control question technique (CQT).

Testing procedure. Actual testing procedures have been described in detail by Barland and Raskin (1973) and Reid and Inbau (1977). Polygraph measuring devices, including pneumographs, a sphygomomanometer, and electrodes, are placed on the subject either during the pretest interview or at its conclusion. After the end of the pretest interview, the sphygmomanometer is inflated and the recording of responses begins. A short period

*The USAMPS provides polygraph examiner training for all service members and almost all government agencies utilize polygraph examiners.

of about 10 to 15 seconds, is used to observe initial respiratory cycles (baseline) and to allow any initial response to fade; then, the examiner asks the first question. Between each question, the examiner waits about 15 to 20 seconds until the response to the last question is finished and physiological response is closer to baseline. The examiner notes on the chart when the exam begins, when questions are asked, and when it ends. Extraneous behavior that affects the recordings may also be noted. When questions for the first chart end, the examiner deflates the cuff.

The examiner then inspects the chart and asks the subject about his or her reaction to the questions. The usual purpose for obtaining subjects' reactions is to allow refinements in the questions. The questions are reviewed again, and, when necessary, further clarified. The examiner may administer a stimulation test, designed to improve test validity. The examiner will then continue to test and obtain two or three more charts in the same way. The examiner may use other stimulation tests between charts, and different questioning techniques (see below) to record different charts. Different questioning techniques may then be used based on information revealed by the subject. In most techniques, any new questions would be discussed with the subject before being asked. The procedures, in preemployment screening or in other personnel screening tests, may differ.

Stimulation tests. Polygraph examiners typically conduct what is known as a "stimulation" test, designed to further convince subjects of the accuracy of the polygraph examination. Although not actually a part of the pretest, stimulation tests can be given either before the first actual set of test questions or after the first chart has been recorded. Stimulation tests are intended to reassure truthful subjects and provoke anxiety in deceptive subjects (cf. Barland, 1978). Their effect should be to increase differential responsivity of deceptive and nondeceptive subjects to different questions on the examination. Some research suggests stimulation tests increase the validity of polygraph examination (Bradley & Janisse, 1981; Senese, 1976).

The most common "stim" test is a "number" or "card" test. A subject is instructed to select, from a deck, a card that has a number, word or suit on the back, or to write a number within a certain range (Decker, 1978; Fingerhut, 1978). Sometimes, the cards are secretly marked or otherwise arranged so that the examiner is sure to know the correct answer (Reid & Inbau, 1977). Many polygraph examiners claim this is unnecessary, however, because the technique is accurate enough without use of such deception (cf. Orne et al., 1972), and secret markings are not employed by Federal agencies. The examiner then may repeat a range of suits, numbers or a set of words, asking the subject if each is the concealed item. The suit, number or word that is actually the concealed item is supposed to provoke the greatest physiological response. Often, the examiner will show the subject the polygram (i.e., the actual chart recordings) to further convince subjects of the instrument's efficacy.

Types of Questions

The central element of any polygraph examination is the test of subjects' responses to a set of questions or items within questions. How these quesitons are structured represents the principal difference among polygraph techniques. There are four different kinds of questions or items

used in polygraph testing, different combinations or questions (generally referred to as question techniques), and different applications for the various techniques. The range of questions, techniques and applications is described more fully below.

Questions. The kinds of questions that are used for polygraph testing have been labeled relevant questions, control questions, irrelevant questions, and concealed information or guilty knowledge questions. Basically, relevant questions are questions about the topic under investigation (a theft, drug use, contact with foreign agents). Suspects' responses to relevant questions are of greatest interest to polygraph examiners.

Control and irrelevant questions can be grouped together as questions used for purposes of comparison to relevant questions. It is important to note, however, that the name one gives to a question may depend on the specific context in which it is used. Thus, one cannot easily give an example of a relevant question or a control question because in different situations and at different times during an examination relevant questions may be used as control questions. Likewise, irrelevant questions may become relevant, depending on a subject's response (Weir, 1974).

Relevant questions. Functionally, relevant questions are questions directly related to the focus of an investigation. In the investigation of a theft, for example, a relevant question might be "Did you steal that money?", or even more specific, "Did you take $750 from Jones' office?" Relevant questions can be broader, however. In preemployment screening and a periodic or random screening, the area of interest may be the subjects' entire background. Thus, there may be a series of relevant questions, such as "Have you ever been fired from a job?" or, "Have you stolen more than $50 in monies in any one year from any of your employers?" (Matte, 1980). Intelligence agencies may ask broad questions concerning unauthorized contact with foreign intelligence agents or involvement in communist activities. Questions for an intelligence screening may also deal with areas which, potentially, may make an applicant susceptible to blackmail. It is important to note, however, that when several relevant questions relating to different issues are used, subjects are not expected to exhibit physiological responses to all of them; the relevant questions that do not evoke responses are used, after the fact, as a type of control question.

To summarize, relevant questions are questions about the topic under investigation, but topics can be very specific (Did you take $750 from Jones' office?) or cover a long period of time and a variety of acts (Have you ever stolen money from an employer? Have you ever had unauthorized contact with a foreign agent?). It is not clear what effect, if any, the breadth of a relevant question has on polygraph results, nor has there been any research done on this issue. As is discussed further, the preponderance of research evidence concerns the use of relevant questions to evoke reactions to specific acts.

Comparison questions. In contrast to relevant questions, which concern issues of direct interest to the examiner, control and irrelevant questions are used for purposes of comparison. As noted above, there is no known physiological response unique to lying. Thus, a polygraph examination could not consist merely of relevant questions. If only relevant times were used, an examiner would not be able to establish the actual

reason for the response. There are a number of reasons, other than fear of detection (or another hypothetically lying related reaction [Barland, 1982] for a subject to become physiologically aroused during the presentation of relevant questions [Davis, 1961; Lykken, 1981; Raskin, 1982; Waid & Orne, 1981].) Even with the addition of nonrelevant comparison items, it is necessary to run several polygraph charts using the same questions (though, perhaps in different order) to be sure that reactions are consistent. If several charts are not run, a subject's responses could be attributed to surprise, physical movement, or some reasons for concern other than a lying-related cause (Weir & Atwood, 1981). On the other hand, the administration of several charts could theoretically just repeat the initial situation leading to the physiological response if the cause were not a random one (e.g., presence at the scene, knowledge of the incident, concern over being falsely identified). Thus, the essence of polygraph testing is the comparison of responses to the relevant questions with responses to nonrelevant questions, which have been labeled either irrelevant or control questions.

Control questions. Control questions, then, are used for purposes of comparison. However, control questions, like relevant questions, vary in breadth and type. One type of control question concerns what is hypothesized to be the same kind of issue that is under investigation at the time of examination. For example, a control question for "Did you take the $750 from Jones' office?" might be "Other than what you have told me [during the pretest interview], have you ever stolen anything in your life?" In an investigation of unauthorized disclosure of classified information, a control question might be "Have you ever betrayed anyone who trusted you?" Subjects innocent of the crime under investigation are presumed to be more concerned about having ever done anything of this sort (and, thus, being the "kind of person" who might have committed the crime under investigation). It is theorized that although guilty subjects will also be concerned about control questions, they will be more concerned and thus exhibit more physiological reactions to relevant questions.

There are a number of views, however, about what distinguishes a control question from a relevant question. One distinction among control questions is whether a question in inclusive or exclusive. Inclusive control questions are questions which include the specific incident under investigation. An example of an inclusive control question in an investigation of an internal theft would be "Have you ever stolen money from an employer?" Exclusive control questions, on the other hand, cover a period of time not including the incident under investigation. An example is, "Before age 18, did you ever take anything of value?" There is some controversy over how far back in time an exclusive control question must be set for the subject to consider it psychologically separate from the incident under investigation and, thus, not a relevant question. Because inconclusive control questions may, from the suspect's perspective, include the act under investigation, some polygraphers contend that they are really relevant questions; that is, they cannot be used for purposes of comparison. The Federal Government, for example, typically uses exclusive control questions because it views inclusive controls as relevant questions. Examiners from the private polygraph firm of John Reid & Associates use both inclusive and exclusive control questions.

Other kinds of nonrelevant questions other than those that cover the

same kind of incident as the one under investigation, or which cover it in a different way, are also considered to be control questions. Thus, for example, "Have you ever fantasized about giving a document to a foreign agent?" is a type of control question used in some intelligence investigations. In some screening examinations, in which contact with a foreign agent is of primary concern (i.e., constitutes the relevant question), "Have you ever done anything for which you are now ashamed?" could be a control question. When a different issue than susceptibility to blackmail is under investigation, "Have you ever done anything for which you could be blackmailed?" can be used as a control question. It is noteworthy that in a different context, such as a broader screening examination, these would be considered relevant questions.

Control questions, then, are questions for which the responses are designed to be compared to responses to relevant questions. In some screening examinations, relevant questions may function as control questions after the fact. That is, if a relevant question produces a relatively mild physiological response, it may be compared to other relevant questions that produce greater response. Most often, control questions are designed to be arousing for innocent subjects (i.e., those who are not being deceptive on the relevant questions), relative at least to relevant questions. This is usually the central point of control questions, and is central to the control question technique (CQT) discussed below.

Irrelevant questions. Another type of question used, in part, for purposes of comparison to responses to relevant questions is the so-called irrelevant question. Examples of irrelevant questions commonly used in investigations are "Are you called [subject's name]?" or "Is today Tuesday?" Irrelevant questions are questions which are believed to have no, or very little, emotional impact on a subject. Thus, such questions can be used as an indicator of a particular subject's normal of baseline level of arousal; no universal standard of physiological arousal can be applied because individuals differ markedly. Irrelevant questions are hypothesized to serve purposes other than providing a physiological baseline (Reid & Inbau, 1977). Perhaps most important, irrelevant questions interspersed among relevant questions are hypothesized to provide a type of rest period for the subject.

Concealed information questions. Questions about concealed information are the fourth type of question used in polygraph testing. Unlike control and relevant questions, which ask subjects whether they have committed a crime, concealed information items aim to detect information about a crime that only a guilty subject would have. Such information might include details about the site of the crime or the means of committing it, such as the type of murder weapon used. It is hypothesized that guilty subjects will exhibit a different physiological response to the correct (relevant) detail than to the incorrect details, but that innocent subjects will respond the same to all the items. Different types of concealed information tests are described below (see "Concealed Information Tests").

The SKY series is a combination of relevant and control questions and concealed information items. On the SKY series, subjects are asked whether they suspect (S) who committed the crime, know (K) who committed the crime or committed the crime themselves (Y for You). The (S) question, (K) question and (Y) question are each compared to the control questions to

ascertain specific types of deception. Such questions are used as the last question in a series of relevant-control questions.

Summary. For any technique, deception is detected by comparison of suspects' physiological responses on critical or "relevant" questions or items with their responses on non-critical (irrelevant or control) items. Greater physiological responses to relevant items than to non-critical (control, irrelevant) items are assumed to be indicative of deception.

Polygraph Question Techniques

Three types of question techniques combining the four question types are described below: the relevant/irrelevant (R/I) technique, the control question technique (CQT) and concealed information techniques. Each of these test types tends to be used for particular purposes; for example, the R/I technique is used in the great majority of preemployment screening interviews, while the CQT is normal in criminal investigations. There have been adaptations of these techniques for other uses, some of which are discussed below. Also, examiners may combine different techniques in an investigation (see, e.g., Reid & Inbau, 1977). In general, the R/I has the broadest potential use while the concealed information techniques are the least applicable. Within each category, particularly the CQT, there is considerable variability and several versions of each technique are employed.

Relevant/irrelevant (R/I) techniques. The R/I technique was the first standard method of polygraph questioning. The method was developed by Marston (1917), a psychologist and the original proponent of polygraph examinations. An adaptation of this traditional technique is used in most of the preemployment screening conducted in the United States.

However, the R/I technique as used by the Federal Government involves very different types of questions than the traditional R/I and it must be explained separately. In fact, the R/I as currently used relies on a type of control question, and would be better designed as a version of the control question technique. An explanation of it is only comprehensible after the usual CQT is explained, and so it is deferred until that point. The versions discussed in this section are: a) the traditional R/I; and b) the R/I as used in typical preemployment screening tests. The R/I as used by the Federal Government is discussed in the section titled "Control Question Technique".

In a traditional R/I examination, the two types of questions used are relevant and irrelevant questions. Deceptive subjects are assumed to have a significantly greater reaction to the relevant questions than to the irrelevant questions. An underlying assumption of this technique is that nondeceptive subjects should have an equal response to all questions, because, being nondeceptive they would not fear questions about the crime or others asked more than irrelevant questions.

There are numerous well-recognized problems with the traditional R/I technique, at least from the perspective of psychologists who have evaluated polygraph test validity (cf. Lykken, 1981; Podlesny & Raskin, 1977; Raskin, 1982). First, the intent of the relevant and irrelevant questions is transparent, which means that the relevant questions are likely to be

more arousing for the truthful as well as the deceptive subjects. Second, questions in the R/I technique are not usually reviewed with the subjects before the test. A larger response to the relevant question may, thus, be due to surprise or misunderstanding, as well as deception. Third, as with any question technique, reactions may be flattened by drugs or by the generally reduced responsivity of certain subjects (Raskin, 1982). These effects are probably more difficult to detect with the R/I than with other question techniques.

Because of these problems, the confidence one can place in the R/I technique is limited (Raskin, 1982). As a consequence, the R/I technique (at least as traditionally constructed) is typically not used in the case of specific incident examinations by either public or private examiners. Its use is almost exclusively with employees in nonspecific investigations. An adaptation of the R/I technique is the principal method of questioning used in preemployment and aperiodic personnel screening. Unlike the questions used with other techniques, R/I questions need not focus on one specific wrongdoing (Barland & Raskin, 1973; Lykken, 1981). The examiner can, thus, assess any number of issues for which the subject's veracity is to be evaluated.

In polygraph examinations used to screen employees, the polygraph examiner usually presents a series of relevant questions, with several irrelevant questions interspersed to provide a baseline. Most relevant questions ask about prior behavior that might disqualify the subject from a job (e.g., employee theft, drug use, fighting on the job, incurring a large debt). Some examinations may include questions about a potential employee's background or intentions regarding a job, for example, "Did you actually graduate from college?" (Weir, 1974, p. 134) or "Are you seeking a job with this company for any reason other than legitimate employment?" (Matte, 1980, appendix). Listed below is an example of questions from a preemployment screening protocol used by a commercial firm (Matte, 1980; also see Ferguson, 1966; Whiteside, 1981).

Relevant Questions:

Did you tell the complete truth on your job application?

Have you deliberately withheld information from your job application?

Have you ever been fired from a job?

Are you seeking a permanent position with this company?

Since the age of () have you committed an undetected crime?

Since the age of () have you been convicted of a crime?

During the past year, have you used marihuana(sic) more than () per ()?

Have you used any other narcotic illegally in the past () years?

Have you sold marihuana(sic) or other narcotics illegally in the past () years?

Have you ever stolen more than ($) worth of merchandise in any one year from any of your employers?

Have you ever stolen more than ($) in monies in any one year from any of your employers?

Have you ever used a system to cheat one of your employers?

Have you ever had your driver's license suspended or revoked?

Have you ever had any traffic citations in the past five (5) years?

Are you seeking a job with this company for any reason other than legitimate employment?

Have you deliberately lied to any of these questions?

The method used by John E. Reid and Associates employs four standard relevant questions:

In the last five years did you steal any merchandise from prevous employers?

In the last five years did you steal any money from previous employers?

In the last ten years did you take part in or commit any serious crime?

Did you falsify any information on your application?

These standard questions may be modified depending on admissions made during the pretest (e.g., a revision may be, "In the last five years did you steal any merchandise other than minor office supplies?"). In addition to the standard question a fifth relevant question (e.g., concerning the illegal purchase or sale of merchandise; use of narcotics) may be added depending on the nature of the job.

The Reid firm also uses what it regards as control questions in preemployment interviews. Control questions include, "Did you ever steal anything in your life?" and "Did you lie to any of the questions you answered during the application process for this job?" The rationale for the use of control questions is discussed below (see control question technique). Essentially, truthful subjects are believed by polygraphers to be more concerned (and, thus, more physiologically aroused) about control than relevant questions. It is not clear, however, how the Reid preemployment control questions differ from the relevant questions. It seems reasonable to suppose that both truthful and nontruthful subjects (in terms of the relevant questions) may be just as concerned with the subject matter of the control questions as they are with the relevant questions. It is not clear as well, why employers would be less concerned with the control than with the relevant questions.

In the R/I questioning technique, a diagnosis of truthfulness or no deception indicated is made by comparison of each relevant question with

the remaining set of relevant (or in the Reid example, control questions) questions. Presumably, an applicant will be deceptive on no more than a few questions. These questions will provoke a greater physiological response than the others, leading to further inquiries and an eventual diagnosis (Ferguson, 1966; Whiteside, 1981).

Other types of questions are used in some screening examinations, such as questions about sexual practices or gambling. Such questions seek information about an applicant's character rather than his or her job performance and are considered by some to be unduly invasive (U.S. Congress, 1979). In response to this practice, ethical standards have been developed for use of the polygraph in preemployment screening (cf. Slowik, 1979), and some States (e.g., Illinois) prohibit their use. Preemployment polygraph examinations fall under the guidelines for employment interviewing of Title VII of the Equal Employment Opportunity Commission, and so are obliged to conduct the examinations in a way that would not discriminate on the basis of sex, race, etc. (cf. Slowik, 1979). One central principle of ethical standards is that relevant questions be related to the job applied for. Whether questions meet this criterion depends on the job, for example, information about one's driving record would be important in hiring a delivery person, but not in hiring a bank teller. Screening applicants for positions involving national security apparently require questions about sexual behavior, drug use, and mental health as well as areas more directly related to national security (e.g., involvement in espionage). The range of topic areas covered in national security preemployment screening examinations by NSA is discussed below under "Current Government Use."

In so-called aperiodic checking for internal security purposes, employees are asked to submit to occasional polygraph examinations. These examinations can assess drug use, subjects' own or others' employee theft, and other matters including job satisfaction and commitment. In this type of examination, almost all of the questions are relevant questions and apparent deception (arousal) in response to any of the items is explored. Examples of the kinds of questions used in aperiodic screening in a supermarket (Whiteside, 1979), for example, include:

"Are you relatively satisfied with this job now?"

"Do you, as far as you know at this time, intend to stay with this employer?"

"Have you ever intentionally underpriced or underweighed merchandise for you?"

Is there a particular person at your store that is responsible for damaging merchandise due to real carelessness, not caring or intentionally?"

The relevant topic areas covered by NSA in aperiodic screening are discussed under "Current Government Use." Because of its use of control questions, the Federal version of the R/I is discussed in the next section.

Control question technique (CQT). The CQT is the most common technique used in investigations of a specific issue. The CQT was developed to deal with some of the inherent problems in the traditional R/I technique

(Reid & Inbau, 1977). Like the R/I technique, it asks relevant questions about the crime like "Did you steal the $750 from Jones' office?" As with the R/I, the deceptive subject is assumed to produce a greater autonomic response to the relevant than to other questions. But the CQT also adds control questions, which, as discussed briefly above, are designed to provoke a greater response in subjects who are innocent and truthful about the crime being investigated.

Control questions are designed to be arousing for nondeceptive subjects. The questions are designed to cause innocent subjects to be doubtful and concerned about whether they have actually told the truth or to be a lie. These questions usually probe for past misdeeds of the same general nature as the crime being investigated but there are transgressions that polygraphers suspect most people have "committed" or considered committing in some form (Reid & Inbau, 1977). An example of a control question might be, "Before the age of 25, did you ever steal anything from a place you worked?" Control questions are designed to cover a long period of time, which may make the subject even more doubtful about the veracity of answers provided.

Considerable attention in the pretest interview is devoted to development of control questions (Reid & Inbau, 1977). The process of developing control questions, reviewing them with the subject, and then refining them is designed to develop the most appropriate questions, and to convince subjects to view control questions as seriously as relevant questions. In addition, the pretest review is designed to get subjects either to be deceptive to control questions or at least to be concerned about the accuracy of their recollections (Barland & Raskin, 1973; Buckley, 1983; Kirchner & Raskin, 1983; Reid & Inbau, 1977). It is considered crucial to produce in the subject the right psychological set in relation to the control questions. This set is then thought to lead subjects to be more concerned about control questions than relevant questions, and so respond more to them. This difference between response to control and relevant questions is then the basis for the diagnosis of deceptive or nondeceptive. Since the subject's psychological set is so crucial when control questions are used, differential responding to relevant or control questions (and ultimately, the validity of the CQT), depends on the nature of the interaction between examiner and subject. This is true regardless of the act in question, the particular CQT method used, or the method of making assessments. Even the validity of an entirely computerized system of scoring and diagnosis would depend on the nature of the interaction between examiner and subject. In this sense, the CQT examinations, as the technology to conduct polygraph tests now stands, always require the examiner to make important judgments about and interventions in his interaction with the subject.

The polygraph examiner does not tell the subject that there is a distinction between the two types of questions (control and relevant). Control questions are described as intending to determine if the subject is the "type of person" who would commit a crime such as the one being investigated (Raskin, 1982). The examiner stresses that the subject must be able to answer the questions completely with a simple "yes" or "no" answer, that the polygraph will record any confusion, misgivings or doubts, and that the subject should discuss any troublesome questions with the examiner (Barland & Raskin, 1973). Thus, the situation is set up such that the

subject is persuaded that the examiner wants the truth. In reality, however, the examiner wants the subject to experience considerable doubt about his or her truthfulness. According to Raskin (Kircher & Raskin, 1983, p. 7), "Control questions are intentionally vague and extremely difficult to answer truthfully with an unqualified 'No'."

To produce the final version of a control question, the examiner begins by asking the subject a broad version of the question used in the pretest interview. Thus, for example, the question might be structured, "Did you ever steal anything in your life?" At this point, different polygraph examiners use slightly different methods to handle the discussion of past wrongdoing in response to the control questions asked during the pretest interview. In the USAMPS method (Raskin & Kircher, 1983), if the subject confesses to a small transgression in the past (such as taking home pencils from work), the examiner will dismiss it as of no consequence. For other misdeeds, the examiner will rephrase the control questions to rule them out (e.g., "Other than what we have discussed, did you ever steal anything in your life?"). The examiner will actively intervene to prevent subjects from unburdening too much of their anxiety over their past wrongs with the intention of keeping them concerned during the actual polygraph testing. Any troublesome past transgressions the subject brings up are excluded (by such phrases as "Other than what we have discussed ...?") so the subject is always brought to the point at which he or she answers "No" to the control question. The control question is then ready to be used in actual testing.

The Reid method varies from the Federal method in some ways (Reid & Inbau, 1977). If the subject does not admit to a past wrongdoing, the examiner may probe until the subject admits to one, even a crime as small as stealing pocket change from a relative during childhood. Such transgressions are then ruled out by adding the kind of exclusionary phrase discussed above (i.e., "Other than what we have discussed, ...?"). However, as in the USAMPS method, it is assumed at this point that the subject is either concealing other misdeeds or is worried that there are others he or she has overlooked (Reid & Inbau, 1977). This worry has been heightened because of the examiner's emphasis on learning the truth to "ascertain" that the subject is not the kind of person that could have committed the crime referred to in the relevant questions. This concern of nondeceptive subjects is assumed to lead to larger responses to control questions than to relevant questions. In addition to relevant and control questions, irrelevant questions are included during the actual interview in order to provide a baseline of physiological responsiveness.

The control question technique assumes that: a) non-deceptive subjects will produce a greater response to the control question than to the relevant question (since the subject is innocent of the crime being investigated, but probably has a past misdeed to hide or is concerned that he or she does); and b) the deceptive subject will produce a greater response to the relevant question (because it is of greater concern to him in spite of the presence of the control questions). Part of the innocent subject's physiological response to the control question may stem from the "information processing burden" it exacts (Raskin, 1982, p. 325).

Several versions of the CQT are regularly employed and adaptations depend both on the training of the examiners and the testing situation.

The Reid version can include relevant questions about several aspects of the crime (Reid & Inbau, 1977). For example, one test could include questions about breaking into an office, stealing a check, and then cashing it. Examiners who use Reid's CQT make a global comparison between the responses to the relevant and the responses to the control questions. They also note the subject's behavior throughout the interview (as discussed above, the Reid technique includes a series of questions in the pretest interview designed to provoke certain "behavioral symptoms in deceptive subjects). The examiner uses the global comparison of polygraph responses supplemented by information about the behavior of the subject to make a judgment of the subject's veracity. An example of a Reid control question sequence, excluding the pretest behavior provoking items, follows (Reid & Inbau, 1977, p. 31):

1. Do they call you "Red"? (where the pretest interview had disclosed he is generally called "Red.")

2. Are you over 21 years of age? (or reference is made to some other age unquestionably but reasonably, and not ridiculously, below that of the subject.

3. Last Saturday night did you shoot John Jones?

4. Are you in Chicago (or other city) now?

5. Did you kill John Jones?

6. Besides what you told about, did you ever steal anything else?

7. Did you ever go to school?

8. Did you steal John Jones' watch last Saturday night?

9. Do you know who shot John Jones?

10. Did you ever steal anything from a place where you worked?

In contrast, Backster's (1962) zone of comparison (ZOC) techniques makes a diagnosis of deceptive or truthful from a standardized numerical scoring of the charts. Each relevant question is paired with a control question. Scores are derived for each relevant question by comparing it only with the previous control question. On each physiological measure, the examiner derives a "plus" (truthful" score if the subject responds more to the control question and a "minus" (deceptive) score if the subject responds more to the relevant question. A positive score above a certain criterion level is diagnosed as truthful, a minus score below a certain level is diagnosed as deceptive, and scores in between are considered inconclusive.

A version of the ZOC is used by Federal polygraph examiners. The Federal version differs from the Backster ZOC in that it may ask about several aspects of the crime in one chart. Relevant questions are asked about primary involvement (e.g., "Did you steal _____?"). secondary involvement (e.g., "Did you help steal _____?"), and so-called evidence connecting (e.g., "Do you know where any of that money is now?"). In addition

to relevant, control and irrelevant questions, the Government ZOC test contains a version of the peak of tension test (see below), and "symptomatic" questions of two types. One type of symptomatic question (e.g., "Do you understand that I'm not going to ask any trick or surprise questions?") is designed to test whether the examinee trusts the examiner's words that no surprise questions will be asked. A large response is symptomatic of distrust. A second type of symptomatic question (e.g., "Is there something else you are afraid I will ask you a question about, even though I have told you I would not?") is to test whether there is some other issue the examinee is concerned about (e.g., another crime) that may be absorbing his or her arousal.

Other versions of the CQT or related techniques are also used by Federal agency examiners. One, the modified general question test (MGQT), resembles the Reid CQT with the following differences: a) Only the polygraph charts are used to make determinations of truth and deception and global evaluations using inferences about behavior are dispensed with; b) Charts are numerically scored; c) Control questions exclusively concern a time and place separate from the time and place of the crime under investigation, with the intention of clearly separating responding related to the crime and the control question; d) The content of control questions is always related to the crime under investigation, i.e., control questions about theft are used to investigate a theft, control questions about assault are used to investigate assault, and so forth. Presumably, when unauthorized disclosures are at issue, control questions would concern some sort of unauthorized disclosures in the past.

As discussed above, another technique employed by Federal Government examiners is called the relevant/irrelevant technique (and is called the general question test [GQT]). Unlike other versions of the R/I, the Federal R/I relies on a type of control question. Like the Reid CQT, it uses inclusive control questions, which pertain to the subject's entire life, such that a complete answer would also include the specific incident being investigated. Thus, with a question like, "Did you ever steal anything from a place where you worked?," the theft being investigated would in actuality be part of the answer. Technically these are seen as "relevant" questions, because they are pertinent to the incident in question. Yet they are functionally control questions, because they are intended to provoke a greater response in innocent subjects than questions about the misdeed provoke.

To summarize, there are a number of control question techniques, the most commonly used being the Reid CQT, the MGQT, and the ZOC. Despite differences among them, they share the same premise and underlying rationale. Use of each of the control question procedures relies on subjects' not knowing when they are being asked the relevant and control questions. If they know which questions are more important for scoring purposes they may be able to make anticipatory responses which could invalidate their charts (see Factors Affecting Polygraph Examination Validity).

Concealed Information Tests. Another polygraph questioning technique works on an entirely different premise than either the CQT or R/I. Instead of detecting deception about having committed a crime per se, concealed information tests aim to detect information about a crime that only a guilty subject would have. Such information might include details about the site

of the crime or the means of committing it (e.g., the type of murder weapon used). Raskin (1982) has aptly described these "concealed information tests." Concealed information tests take two forms: the peak of tension (POT) and the guilty knowledge test (GKT).

The POT was developed by Keeler (cf. Harrelson, 1964) and has long been used in criminal investigations. The POT test uses a set of five to nine nearly identical "yes or no" questions asking if the subject knows about a particular detail related to a crime. The detail may be a type of object used, or the color of an item. One question actually includes the relevant detail, while the others include plausible but false details of a parallel nature. The questions and the sequence in which they are asked are reviewed with the subject in the pretest interview. The subject is usually instructed to answer "no" to each question. The question with the true detail is usually presented in the middle of the sequence, so that the subject's physiological reactions will increase up to the critical question, where they will reach a peak, hence the name, and fall back down again. The card and number stimulation tests discussed above are actually examples of the POT. Barland and Raskin (1973, p. 429) provide a hypothetical example of a POT in a criminal case:

1. Regarding the color of the stolen car, do you know it was yellow?

2. Do you know it was black?

3. Do you know it was green?

4. Do you know it was blue?

5. Do you know it was red?

6. Do you know it was white?

7. Do you know it was brown?

Occasionally, criminal investigators use the POT technique to discover and develop additional information about a case. The examiner asks the suspect about a series of details, but does not know which is actually relevant to the crime. The detail that provokes an exceptional physiological response is used as a clue in the investigation. For example, an examiner might use the POT to determine the exact location where stolen goods were hidden. This kind of examination is called a searching Peak of Tension test (Barland & Raskin, 1973). The searching POT technique has been used, for example, in cases in which employees are suspected of having stolen money, but there is no evidence about the extent of the theft (Lykken, 1981). The examiner asks the employee if he has stolen money ranging from a small amount to the entire amount taken. The amount that provokes the largest response is assumed to be the amount of the total that the employee stole.

The Guilty Knowledge Test (GKT), described initially by Lykken (1959, 1960) works in much the same way as the POT. The GKT, however, often includes a larger set of questions, and the questions may be of the multiple choice type rather than the "yes or no" type. Also, studies investigating the GKT have only used the electrodermal response (EDR), while the POT

tests have employed standard three-channel polygraph recordings. An example of two questions from a GKT used in a laboratory study by Lykken (1959, p. 389) is listed below.

1. If you are the thief, you will know where the desk was located in the office in which the theft occurred. Was it (a) on the left, (b) in front, or (c) on the right?

2. The thief hid what he had stolen. Where did he hide it? Was it (a) in the men's room, (b) on the coat rack, (c) in the office, (d) on the windowsill, or (e) in the locker?

There is a major difference, however, in the use suggested for the GKT as compared to the use of the POT. The POT is usually used as a supplement to a control question test, or as an aid in investigation. The GKT, however, has been proposed as an alternative to control question techniques (Kleinmuntz & Szucko, 1982; Lykken, 1974, 1981). Proponents argue that the GKT may reduce the number of false positives, becuase it focuses on specific details that would be salient only to the perpetrator of a crime (Lykken, 1979, 1981). Also, they claim, the validity of the GKT can be substantially improved by increasing the number of questions on the test. Critics claim that it is especially susceptible to false negative (Raskin, 1982), and that GKT proponents do not adequately assess the consequences of false negatives.

Concealed information tests have, according to several reviewers (e.g., Lykken, 1981; Raskin, 1982), several important limitations. One problem is that they may not be widely applicable. Knowledge about an incident may not differentiate between a guilty and innocent pereson where, for instance, a suspect is present at the scene of a crime but claims that someone else is responsible (Lykken, 1981; Raskin, 1982). Furthermore, concealed information tests require investigators to gather information that is not always possible to obtain, or must be disclosed to suspects in other parts of the investigation (Raskin, 1982). In some cases, publicity about the details of a crime eliminates the possibility of a concealed information test, since the information is public knowledge (Raskin, 1982). Finally, the false negative rate can create problems.

Post Test Interview

Interspersed among test questioning and measurement of physiological responses are a number of opportunities for examiners to discuss the test with the subject. At each occasion, the examiner reviews the questions, and, depending on the responses, questions subjects about their responses. At the end of the examination, the examiner will make an assessment of whether a subject is being deceptive or nondeceptive. In some methods, e.g., Reid's (Reid & Inbau, 1977), the assessment is a global one employing behavioral as well as polygraph data. But the USAMPS, Backster's ZOC and other methods rely strictly on polygraph chart interpretation (Barland & Raskin, 1973; Backster, 1979). In examinations conducted by the Federal Government, the final official determination is made after supervisory review of polygraph charts. If deceptive, the examiner will attempt to elicit a confession. Usually, this is not done directly but is couched in terms of providing the subject with an opportunity to clarify/explain the responses and differences obtained.

Uses of Polygraph Testing

As has been implied in much of the above discussion, polygraph examinations are used for a variety of purposes. The goal of all such applications of the polygraph is the detection of deception or substantiation of truth. The nature of the test situation, however, leads to important differences in the way a polygraph examination is conducted. Unfortunately, the published research literature deals almost exclusively with the use of the polygraph by police and military examiners for criminal investigations. The research literatures does not focus on a number of important uses of polygraph testing, such as for national security purposes and for employment screening.

Current Use

The majority of uses of polygraph testing appear to be on behalf of private employers, the next greatest number are in the context of local criminal justice investigations, and the remainder are done by the Federal Government. Of greater concern for the present analysis are the numbers and types of examinations currently conducted by agencies of the Federal Government. This section will devote most attention to such uses, although local government and private use are briefly discussed in order to place Federal use in context.

Current Federal Government use. In order to assess the extent of polygraph use among Federal agencies, the Office of Technology Assessment (OTA) conducted a survey of Government use during May 1983. The request for information was sent to the Departments of Defense, State, Justice, Treasury, the U.S. Postal Service, and the Central Intelligence Agency (CIA), all of which were believed to employ polygraph examinations. Information was requested about the number of examinations, purposes and results, as well as about research conducted and/or planned. At the time of this report, all agencies excepting the CIA had provided written responses to the request for information about the number and type of polygraph examinations that have been administered.

The CIA declined to respond because of the classified nature of the information. However, some data about the CIA's use for background investigations were reported in a 1980 study (Director of Central Intelligence, 1980). The number of polygraph examinations are summarized in Table 2.1. Table 2.1 indicates that Federal agencies reported administering a total of 22,597 polygraph examinations in fiscal year 1982, about half of these in the context of criminal investigations. Polygraph examinations are also reported to be used for intelligence and counterintelligence investigations (some [NSA] at aperiodic intervals), and pre-employment screening. The largest single number of polygraph examinations conducted in 1982 were conducted by the National Security Agency (NSA), a component of the Department of Defense (DOD), primarily for preemployment screening. These numbers can be compared to previous surveys conducted in 1963, when Federal agencies, excluding NSA and CIA, conducted 19,796 polygraph examinations, and 1973, when 6,946 examinations (including 3,081 by NSA) were conducted.

NSA reports that it uses primarily the relevant/irrelevant question technique. NSA reports that counterintelligence-type screening examinations, that is, tests given to NSA (or affiliated) personnel who already

Table 2.1
Federal Government Polygraph Examinations
Conducted in Fiscal Year 1982*

Agency	Total
DOD	
Army	3731
Army Intelligence	279
Navy	1337
Air Force	3019
Marines	263
National Security Agency	9672
State	5
DOJ	
FBI	2463
DEA	211
Treasury	
Secret Service	714
Bureau of Alcohol, Tobacco and Firearms	256
U.S. Postal Service	652
CIA	n.a.
TOTALS	22,597

* Notes: Data were also reported for FY 1980, 1981, and, in some cases, YTD 1983.

have access to classified information, would have relevant questions on the topics of involvement in espionage or sabotage against the United States; knowledge of others involved in espionage or sabotage against the United States; involvement in giving or selling classified materials to unauthorized persons; knowledge of others giving or selling classified material to unauthorized persons; and unauthorized contact with representatives of a foreign government (National Security Agency, no date, p. 3). Examinations that are given to applicants for employment and contractors who are applying for access to Sensitive Compartmented Information (SCI) consist of questions about the topics covered in counter-intelligence-type aperiodic screenings (phrased as "Do you plan to commit ...?") as well as questions about a broader range of issues: involvement in communist, fascist or terrorist activity; commission of a serious crime; involvement in adult homosexual activity; involvement with illegal drugs or narcotics; deliberate falsification of security processing forms; treatment for a serious nervous or mental problem (NSA, no date). According to NSA, the scope of specific issue examinations is limited to questions that are relevant to the issue to be resolved. Presumably, specific issue examinations would be conducted using the control question technique. Examples of specific examination questions were not provided by NSA.

Current DOD regulations also allow the use of polygraph examinations to investigate situations in which credible derogatory information about an individual with clearance is provided to officials. The frequency of this type of investigation, however, was not reported. Prior to the President's National Security Decision Directive of March 11, 1983, use of the polygraph in personnel investigations of competitive service applicants and appointees to competitive service positions was limited to executive agencies with highly sensitive intelligence or counterintelligence missions affecting the national security (e.g., a mission approaching the sensitivity of that of the CIA; see Office of Personnel Management, 1973). Approval to use the polygraph could be granted only for 1-year periods. Refusal to consent to a polygraph could not be made a part of an applicant or appointee's personnel file. See next section for a description of proposed changes in Federal use of polygraph testing.

Non-Federal Government use. Outside the Federal Government, polygraph examinations are administered as part of criminal investigations, as well as preemployment screening and periodic screening of employees for purposes of controlling internal crime and recommending promotions. Less frequent uses include examinations in such situations as paternity investigations and workers' compensation cases. It has been estimated that over a million polygraph examinations are given a year (Lykken, 1974), 300,000 of them for employment purposes alone (Privacy Protection Study Commission, 1977).

Both private and police polygraphers use polygraph examinations in the process of criminal investigations (see Raskin, 1982). In some cases (Most typically, rape and kidnapping cases, but see also, Minchew testimony, 1979), witnesses and victims whose veracity is in doubt are asked to take a polygraph examination. Suspects who claim innocence may be asked by their defense attorneys or the prosecution to support their claim by taking a polygraph examination. In such cases, prosecutors and defense attorneys may make informal agreements to drop the charges if the polygraph examination indicates no deception. In such cases, the prosecution and the defense may formally stipulate that if deception is indicated, results of the polygraph examination will be admissible at trial. In some cases New Mexico, Massachusetts, and the 9th Federal Circuit Court of Appeals (Ansley, 1983b; Raskin, 1982; Smith, 1981, 1982, 1983) polygraph evidence has been admitted over objection. Polygraph evidence is also used occasionally in postconviction proceedings such as sentencing and motions for a new trial (Raskin, 1982). In polygraph examinations as part of criminal investigations, some version of the control question technique is typically used.

The use of the polygraph examination by employers is reported to be widespread (Sackett & Decker, 1979). Although it is illegal to ask employees to take an examination in 19 States and the District of Columbia, it is legal to do so in 31 States (Ansley, 1983b; Smith, 1981, 1982, 1983). Polygraph examinations are used most commonly in commercial banking, investment banking and retail operations. In such settings, both risk of theft and fraud are high and, in addition, employee turnover is high. The use of polygraph examinations is also recommended to employers as a check before making promotion decisions (Whiteside, 1981).

Conclusions

What is often referred to as "the polygraph" is actually a set of relatively complex procedures for asking questions and measuring physiological responses in order to detect deception or establish truth. Polygraph testing is employed for a variety of uses, ranging from ascertaining the guilt of a criminal suspect to assessing the honesty of a prospective employee. Because different polygraph procedures are required depending on intended use, it is necessary to consider validity by polygraph technique and situation. In subsequent sections, such a variegated analysis is presented and the scientific-policy context is more fully described.

CONTROVERSY OVER POLYGRAPH TESTING VALIDITY

Introduction

The validity of polygraph examinations to detect deception has long been a controversial issue (cf. Lykken, 1981; Raskin, 1982; Waid & Orne, 1981, 1982). Since development of polygraph techniques almost 80 years ago, their use both within and outside the Government has been the focus of numerous judicial opinions, and, as well, legislative and executive branch debate. Polygraph examinations have been advocated as a way to ascertain guilt of criminal suspects, to exculpate innocent suspects, to protect national security, and to maintain employee honesty. Polygraph examinations have, at the same time, been criticized for providing inaccurate and misleading information, for failing to detect security risks (U.S. Congress, 1965), for interfering with the rights of private citizens (Privacy Protection Study Commission, 1977), and for lowering employees' morale (U.S. Congress, 1983). At the center of controversy over the use of polygraph examinations is the question of its validity: does a polygraph examination actually identify truthful and nontruthful individuals?

Recent interest in polygraph examinations and their validity stems from efforts to broaden Government use. The Department of Defense (DOD), in late 1982, drafted revisions to existing regulations (5210.48). DOD proposed expansion of the use of polygraph tests for preemployment screening and random testing of employees who have access to highly classified information. Currently, only the NSA and CIA are able to use polygraph tests in this way. Expanded use of polygraph testing in all Federal agencies was made explicit in a Presidential national security decision directive (Mar. 1983, NSDD-84). In part, the directive requires agencies and departments which handle classified information to revise existing regulations. The revisions would permit use of polygraph examinations as part of internal investigations of unauthorized disclosure of classified information. Prior to the directive, investigations of unauthorized disclosures had to be referred to the Department of Justice. Employees who refuse to submit to a polygraph examination could, if NSDD-84 is implemented, be subject to adverse consequences.

Proposals to expand use of polygraph examinations to maintain national security have renewed the debate about the appropriateness of various polygraph techniques and their ability to detect deception. In order to provide a context for the present evaluation of scientific evidence on the validity of polygraph testing, previous assessments of accuracy of polygraph testing are reviewed in this section. Legal precedents regarding

polygraph testing and congressional hearings on its use both within and outside of Government are briefly considered. The section also describes scientific criteria for establishing validity and reviews other efforts to evaluate the scientific literature on testing.

Judicial Reviews

When courts have been called upon to resolve disputes concerned with use of polygraph examinations, they have had to consider both the technique's validity and whether its use, however valid, interfaces with other values that the law seeks to protect. The varying decisions reached by State appellate courts and Federal circuits (see summary by Ansley, 1983b) may in large measure reflect varying beliefs about the validity of polygraph examinations. Indeed, for many years, the leading case on the admissibility of novel scientific evidence Frye v. U.S. (1923) was a case about the admissibility of polygraph evidence, and the opinion centered on the question of validity. The issue of how a court is to decide the question of any scientific technique's validity has brought the Frye test into question in recent years and makes salient the problem of establishing judicial standards for assessing validity (Giannelli, 1980).

Polygraph Findings as Evidence

The Frye case involved a 19-year-old defendant convicted of robbery and murder. Prior to his trial, a well-known psychologist and one of the originators of polygraph testing, Dr. William Marston, administered a "systolic blood pressure test" to detect deception (See, e.g., Marston, 1917). Dr. Marston determined, on the basis of this test, that Frye was truthful when he denied involvement in the robbery and murder. The trial judge, however, refused to permit Dr. Marston to either testify about the examination or conduct a reexamination using the blood pressure test in court.

Frye appealed his conviction on the grounds that relevant exculpatory evidence had not been admitted. The appeals court, however, concurred with the initial trial court judgment. The court reasoned that the systolic blood pressure deception test was validated only by "experimental" evidence and was not based on a "well-recognized scientific principle or discovery." The decision stated that, "while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the things from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. Just when a scientific principle crosses the line between experimental and demonstrable is difficult to define" (p. 1014). Ironically, Frye's conviction was later reversed when another man confessed to the crime, thereby providing Frye with more convincing cooroboration of his denials of guilt.

The Frye trial is still used as precedent in most Federal courts. Subsequent opinions (in areas other than the polygraph) have tried to better define that line between "experimental" and "demonstrative" stages of a scientific innovation. For example, the court in U.S. v. Stifel (1970) held that "neither newness nor lack of absolute certainty in a test suffices to render it inadmissible in court" (p. 438). In a second case, U.S. v. Brown (1977), the court seemed also to be concerned with validity: "The fate of a defendant in a criminal prosecution should not stand on his

ability to successfully rebut scientific evidence which bears an 'aura of special reliability and trustworthiness,' although, in reality the witness is testifying on the basis of an unproved hypothesis in an isolated experiment which has yet to gain general acceptance in its field" (p. 556). The Frye test has been held to be too high a hurdle by some trial courts, which have replaced it with the test for admissibility of expert testimony generally: "testimony by a witness as to matters which are beyond the ken of the layman will be admissible if relevant and the witness is qualified to give an opinion as to the specialized area of knowledge" (U.S. v. Stifel 1974 p. 53).

A closely related question for the courts has been who should determine whether some procedure has gained general acceptance in its field. Some have held that the courts must look to the judgment of the scientific community (e.g., U.S. v. Williams, 1978). In other decisions, the court refused to "surrender to scientists the responsibility for determining the reliability of (scientific) evidence" (p. 1198), and that "a determination of reliability cannot rest on a process of 'counting (scientific) noses'" (p. 1198).

Saks and Van Duizend (1983) concluded that whichever set of tests is employed, the courts are in a weak position to assess validity directly or to count scientific noses. The result has been: a) general deference by the courts to the judgments of scientific communities; and b) "numerous incongruities where less reliable scientific and technological information is admitted but the admission of demonstrably more reliable techniques is delayed until the requisite consensus has formed" (Saks and Van Duizend, 1983; see, also, Giannelli, 1980).

When the courts examine polygraph testing, they are faced with a series of dilemmas. To which "particular field" of expertise can the courts turn: physiology, psychology, polygraph? If they look to the data themselves, what are they to make of it? As the present report suggests, validity assessment involves a complex situation and technique specific answer. Even if a final, single accuracy rate could be established, how should a court use it. How accurate must a diagnostic or predictive technique be to be deemed valid for evidentiary purposes? Regularly admitted psychiatric evidence is widely recognized (including by the U.S. Supreme Court, see Addington v. Texas, 1979) as having accuracy rates comparable to flipping coins (e.g., Ennis & Litwack, 1974; Ziskin, 1982). In Barefoot v. Estelle (1983) the Supreme Court acknowledged that psychiatric predictions of dangerousness and violent behavior do not exceed an accuracy level of 33 percent (see Monahan, 1981). Yet, this evidence was held admissible in Barefoot and sufficiently valid to uphold a decision to execute a convicted person.

In summary, then, the courts have found themselves disagreeing on the methods for establishing validity for purposes of admissiblity of evidence, where the critical focus of such judgment should rest. In addition, courts are inconsistent about what decision to make on the basis of judicial findings of fact regarding the validity of a diagnostic or predictive device.

Laws Regulating Polygraphs in Employment Settings

As described in section 2, screening employees and maintaining honest

behavior among employees is the most frequent application of polygraph testing. Many employers argue that use of polygraph testing for preemployment screening, periodic checking, and to resolve actual thefts is necessary. Internal crime has been established to cost private industry up to ten billion dollars annually (see U.S. Congress, 1978) and polygraph testing is regarded as a cost-effective tool. Employers argue that screening applicants, and periodic checking of employees, are the most efficient ways to control pilferage, embezzlement, poaching, and other forms of theft. As noted in section 2, the need for polygraph testing is felt particularly in industries which have high risk of theft and fraud (e.g., commercial banks), high turnover (supermarkets, other retail operations) or both.

According to Ansley (1983b), the use of private polygraph testing is limited by statute in 18 States plus the District of Columbia. Most of these laws seek to protect employees from being requested, required, demanded, or subjected to polygraph examinations by their employers. Employers are reported to be able to find ways around these laws. For example, employers may tell the employee that they suspect them of theft (or whatever) but that if the employee can find a way to demonstrate innocence, the employer will not discharge the employee. In addition to polygraph validity, other polygraph related concerns include issues of voluntariness, invasions of privacy, being compelled to inform on other employees, inhibiting union activity, and the polygraph as a cover for racism and sexism. This list does not exhaust concerns that have been expressed.

A survey of 143 private firms (Belt & Holden, 1978), regarding their use of polygraph testing, yielded a number of interesting findings. Twenty percent of respondents reported using polygraph examinations for preemployment screening, periodic surveys, and investigations of specific onsite crimes. It is interesting that of reasons given for using or not using the method, users ranked moral or ethical considerations last and efficiency first; nonusers, however, ranked validity and reliability second in importance, cost third, and the availability of qualified operators fourth in importance. Belt and Holden found a positive relationship between a State having a licensing requirement and employers' use of polygraph testing. According to Ansley (1983b), 25 States have licensing requirements for polygraphers; licensing is optional in one state.

Although there is testimony that use of polygraph testing reduces employee crime (U.S. Congress, 1978), no formal cost-benefit analyses appear to have been conducted. In addition, there is no research on the predictive validity of polygraph results (Hayden, 1982; Sackett & Decker, 1979). Although employee issues are critical to proposed Government uses of polygraph testing (cf. U.S. Congress, 1983), few data are available on Government employees (see sections 4 and 5).

One additional area of controversy has concerned employee rights and employer-employee relationships. The general matter of invasion of privacy is particularly pertinent in preemployment screening and periodic checking. In preemployment screening, the range of questions that may be asked has been subject to particularly heavy criticism. Questions have been reported to include items concerning union activity, sexual preference, and family problems (U.S. Congress, 1974); and, as well, willingness to make a commitment to the job (Sackett & Decker, 1979), and whether the respondent has ever been tempted to steal (Hayden, 1978). During periodic checking,

respondents are sometimes asked not only about their own possible improper behavior (e.g., underringing in supermarkets), but also about their level of job satisfaction, intention to remain with the employer, and activities of their fellow employees (Whiteside, 1981). There is some concern about whether prejudices of the polygraph examiner based on racial, ethnic, and gender stereotypes bias employees' responses (Sackett & Decker, 1979). These assertions do not appear to have been tested.

One argument against the use of polygraph examinations in the employment situation is that it destroys the trust relationship between employers and employees, and creates employee dissatisfaction. Apparently, five studies have examined whether the use of the polygraph causes private sector employees to be dissatisfied (see Phannenstill, 1983). In one study, Ash and Wheeler found that 96 percent of applicants were willing to take a polygraph examination to get a job; 86 percent of the applicants thought the preemployment examination was fair, and 88 percent were willing to take it routinely as a condition of employment. A problem with the study was that applicants were surveyed immediately after taking the polygraph examination so they may have thought their responses were part of the screening process (Sackett & Decker, 1979). In the one known survey of Federal employees, the Air Force (1983) surveyed individuals who had volunteered to participate in a pilot project on the use of the polygraph for counter intelligence/security examinations. About 99 percent of the respondents felt that the examination was fair, and were willing to take an examination for counterintelligence purposes.

Federal Debate Over Polygraph Validity

Concern about and debate over Federal Government use of the polygraph have emerged at several points during the past twenty years. As shown in figure 3.1, the history is essentially one of legislative concern triggered by some executive branch proposal or action regarding polygraph testing. The questions raised by Congress have included constitutional and ethical as well as validity issues. However, the scientific validity and reliability of polygraph testing has been and is a central congressional concern. This section briefly describes the history of Federal Government involvement with the issue of polygraph validity.

1960s

Congressional interest first intensified in 1963 when controversy developed over an executive branch proposal to use lie detectors to find the source of unauthorized disclosures of sensitive or classified information, sometimes known as "leaks" ("Uproar Erupts," 1963). The then Chairman of the House Committee on Government Operations asked the Foreign Operations and Government Information Subcommittee to study the Federal Government's use of polygraphs. The study found that, excluding the National Security Agency and Central Intelligence Agency (for which information was classified), Federal agencies had conducted 19,796 polygraph examinations in 1963. In 1964, the Subcommittee held hearings and received testimony from private polygraphers, researchers, and Federal officials. In a 1965 report (U.S. Congress, 1965), the House Committee on Government Operations concluded that there was no scientific evidence to support the theory of the polygraph, and that the research evidence as to its accuracy was inadequate. The Committee recommended that further research be conducted and

training for polygraph examiners be upgraded, and that the President establish an interagency committee to study and work out solutions to problems posed by Federal Government use of polygraphs.

Later in 1965, an interagency polygraph committee of representatives from the Department of Defense, CIA, Department of Justice, Bureau of the Budget (now Office of Management and Budget), Office of Science and Technology (now the Office of Science and Technology Policy), and other executive agencies was established. The interagency committee concluded that: 1) there was insufficient scientific evidence concerning the validity and reliability of polygraph tseting; and 2) the use of the polygraph constitute an invasion of privacy of the individual being interrogated. The committee recommended that the "use of the polygraph in the Executive Branch should be generally prohibited, and permitted only in special national security operations and in certain criminal cases" (U.S. Civil Service Commission, 1966, p. 2). The recommendations made at that time concerning personnel screening were promulgated as Civil Service regulations.

Thus, in 1973, the Civil Service Commission (now OPM) issued an appendix (chapter 736, appendix D, of the Federal Personnel Manual) regulating the use of polygraphs in personnel investigations of competitive service applicants and appointees to competitive service positions. According to this appendix, which is still in effect, only executive agencies with highly sensitive intelligence or counterintelligence missions directly affecting the national security such as "a mission approaching the sensitivity of that of the Central Intelligence Agency" are permitted to use the polygraph for employment screening and personnel investigations of applicants for and appointees to competitive service positions. All other uses of a polygraph to screen applicants for and appointees to competitive positions are forbidden.

The appendix also set forth steps for determining whether agencies met the criteria of having a highly sensitive mission, and stipulated that approval to use the polygraph would be granted only for one-year periods. Agencies intending to use the polygraph for personnel screening were required to prepare regulations and directives meeting certain minimum standards. The minimum standards included directives concerning the specific purposes for which the polygraph may be used, and directives that a person to be examined must be informed as far in advance as possible of the intent to use the polygraph and of the fact that refusal to consent to a polygraph examination will not be made a part of the person's personnel file.

Also in response to the House Government Operations Committee's 1965 report, DOD proposed, and in part undertook, an extensive polygraph research program. And in July 1965, DOD issued directive 5210.48 (DOD, 1965) to regulate the conduct of polygraph examinations and improve selection, training, and supervision of its polygraph operators. Some of the results of the DOD research program were later reported in a scientific journal (Bersh, 1969, see section 4), but other reliability and validity studies proposed were never carried out (DOD Joint Services, 1968).

Between 1967 and 1973 a number of bills were introduced which would have either limited the questions that could have been asked or banned altogether polygraph use by Federal agencies (U.S. Congress, 1974). None of these bills was enacted.

1970s

Ten years after the 1964 hearings, this same House Government Operations subcommittee conducted another review of polygraph use by Federal agencies (U.S. Congress, 1974a). In 1974 hearings, the Subcommittee found that the use of polygraphs in the Federal Government had declined substantially since 1963. In fiscal year 1973, a total of 6,946 examinations were conducted, including 3,081 by NSA. This compared to 19,796 in 1963, excluding NSA and CIA. The Subcommittee also found that there was not much additional research on polygraph validity. The only federally funded studies conducted had been those reported by the DOD Joint Services Group (1968), and these studies were considered by DOD to be inadequate for determining the validity and reliability of Federal polygraph testing.

In a 1976 report based partly on the 1974 hearings, the House Government Operations Committee concluded that "the nature of research undertaken, both federally and privately funded, and the results therefrom, have done little to persuade the committee that polygraphs ... have demonstrated either their validity or reliability in differentiating between truth and deception, other than possibly in a laboratory situation" (U.S. Congress, 1976, p. 12). The 1976 report concurred with the 1965 report that "There is no 'lie detector'" (U.S. Congress, 1976, p. 46). Because of the polygraph's "unproven technical validity" and the suggestion that the "inherent chilling effect upon individuals subjected to such examination clearly outweighs any purported benefit to the investigative function of the agency," (p. 46), the Committee recommended a complete ban on the use of polygraphs by all Federal Government agencies for all purposes. However, thirteen committee members dissented, asserting both that the hearings had been held during an entirely different Congress, and participated in by an entirely different group of Members, and that, while testimony at the hearings represented a wide diversity of views, no witness had urged prohibition of the polygraph for all purposes. The dissenters urged adoption of the recommendations originally proposed and voted upon by the members who had participated in the hearings. These recommendations would have, in part, prohibited the use of polygraphs in all cases except "(1) those clearly involving the Nation's security and (2) those in which agencies can demonstrate in compelling terms their need for use of such devices for their law enforcement purposes, and that such uses would not violate the fifth amendment or any other provision of the Constitution."

The concern with scientific validity and its implications for the Federal Government's use of polygraph testing arose again in 1979 at hearings held on pre-employment security clearance procedures by the House Permanent Select Committee on Intelligence, Subcommittee on Oversight (U.S. Congress, 1980). The subcommittee found that there had been insufficient research on the accuracy of the polygraph technique in screening job applicants (p. 15) and that "gaps in the statistics kept by the intelligence services do not make it possible to make the clear judgment that the polygraph is unique and indispensable" (U.S. Congress, 1979b, p. 15). The Director of Central Intelligence (DCI) was urged to conduct a study to validate the accuracy of the polygraph for pre-employment screening. The DCI did conduct a study in 1980, but it was a utility study, not a validity study (DCI, 1980).

As shown in figure 3.1, in addition to interest in Federal use of

polygraphs, Congress has shown interest in the use of polygraph examinations by private employers, in part because of constitutional and privacy issues (see, e.g., U.S. Congress, 1974, 1978, 1979b; the Privacy Protection Study Commission Report [1977] mandated by Public Law 93-579; and several laws introduced since 1967). Various congressional committees have questioned the validity of polygraph testing in a private employment context, in particular as a condition for employment. Nevertheless, attempts to enact Federal legislation regulating the use of polygraph examinations by private employers and/or the Federal Government have not been successful.

1980s

In the recent past, the executive branch has again taken initiatives concerning the Federal use of polygraph testing. In April 1982, a DOD select panel reviewed the DOD personnel security program (DOD, 1982) and expressed dissatisfaction because of inconsistency in polygraph use across component programs (as did U.S. Congress, 1979) and the lack of reinvestigations. The panel observed that military personnel, unlike civilians, were appointed to NSA and allowed access to Sensitive Compartmented Information (SCI) without undergoing a polygraph examination. In addition, personnel could continue to get clearances throughout their careers without ever being subjected to reexmaination. The DOD panel recommended a broadened application of the polygraph for security screening purposes, and selective use of counterintelligence scope polygraph examinations during periodic reinvestigations. The panel noted that the recommended expanded use of the polygraph would require changes in DOD Directive 5210.48.

On August 6, 1982, the Office of the Deputy Secretary of Defense (Carlucci, 1982) issued a memorandum requiring employees with SCI access to agree to submit to polygraph examinations on an aperiodic basis, and revised DOD Directive 5210,48 accordingly. Later in 1982, further revisions to DOD Directive 5210.48 were drafted (DOD, 1982). In 1983, the President issued a National Security Decision Directive (NSDD-84) also authorizing broader use of the polygraph. Congress responded to these developments by conducting several sets of hearings (U.S. Congress, 1982, 1983a, 1983b), by requesting OTA and GAO studies, and by passing an amendment to the DOD appropriations authorization bill (S.675) putting a moratorium until April 15, 1984, on any revisions to DOD Directive 5210.48 retroactive to August 5, 1982. The DOD draft revisions and NSDD-84 are discussed in more detail below.

Draft Revisions to DOD 5210.48

The draft revisions to the DOD polygraph regulations have gone through several iterations. For the purposes of this validity study, a primary proposed revision (as of the March 1983 draft) is to authorize the use of the polygraph for determining initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (SCI and/or special access). The use of the poygraph in determining the continuing eligibility would be on an aperiodic (i.e., random) basis (DOD, 1983).

Also, the proposed revisions provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may, after consideration of all other relevant

factors, result in adverse consequences for the individual. Adverse consequences are defined to include nonselection for assignment or employment, denial or revocation of clearance or reassignment to a nonsensitive position.

Technically, these expanded uses of the polygraph are considered to be part of personnel security investigations. Use of the polygraph within DOD is already authorized under the existing 1975 version of 5210.48 for various criminal, counterintelligence, and intelligence purposes.

A detailed review of the proposed changes is beyond the scope of this study.

NSDD-84

On March 11, 1983, the President issued a National Security Decision Directive intended, according to DOJ officials, to help safeguard against unlawful disclosure of properly classified information. One main provision of NSDD-84 requires that persons with authorized access to classified information sign a nondisclosure agreement, and that persons with access to SCI must also agree to prepublication review. These provisions are outside the scope of this study, as is a full analysis of NSDD-84.

With respect to the polygraph, NSDD-84 in effect authorizes agencies and departments to require employees to take a polygraph examination in the course of investigations of unauthorized disclosures of classified examinations. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences. NSDD-84 permits administrative sanctions, including denial of security clearance, to be applied even when a person is not subject to a criminal investigation (DOJ, 1983).

Scientific Validity and Polygraph Research Reviews

Reviews of scientific literature form the principal means to cumulate research findings and are especially important in order to assess the usefulness of polygraph testing. Single research studies, no matter how well conducted, cannot answer global questions about validity and must be considered in relation to other evidence. Both because research evidence about polygraph testing has rapidly increased, especially within the last 10 years, and because there have been disagreements about the nature of evidence about polygraph testing, there have been a number of such reviews. These reviews are important, because they are frequently cited in both legal and legislative considerations and because they serve to shape future research.

Underlying each of the reviews is the application of a set of criteria, only sometimes made explicit, regarding the validity of individual studies and their implications for overall assessments of polygraph testing accuracy. As introduction to the scientific reviews, the nature of these criteria is described. The reviews, themselves, are then summarized and a preliminary analysis of discrepancies among reviews is presented. More detailed analysis of individual validity studies is provided in sections 4 and 5.

Definitions of Scientific Validity

Validity. The validity of polygraph testing means, in nontechnical terms, accuracy in detecting deception and truthfulness. The problem is assessing polygraph validity is especially difficult, not only because polygraph tests take a number of forms, but also because validity has different dimensions and can be measured in a number of ways. There are, as a result, a number of different forms of validity associated with polygraph examinations depending on the type of polygraph test as well as on its use (e.g., employee screening vs. investigation of a criminal suspect). These difficulties underlie, in part, the failure to have developed agreed-upon assessments of polygraph validity.

In order to make explicit the criteria for validity used in this assessment, below are described several dimensions of validity and how they are assessed. This description is based both on standards for psychological/psychometric tests (cf. American Psychological Association, 1974, 1983) and criteria to evaluate research designs (cf. Cook & Campbell, 1979; Saxe & Fine, 1981). Although criteria for validity can be described objectively, it should be noted that it is essentially a qualitative judgment as to whether (or, to what extent) a given criterion is met. In addition, assessments of the "preponderance" of evidence necessary in order to assess the overall validity of polygraph testing, are similarly subjective. In subsequent sections, a systematic analysis of available research is attempted, although it should be recognized that there are a number of options to conduct such evaluations, each of which may yield a somewhat different outcome.

Reliability. Assessment of any test's validity is based on the assumption that the test consistently measures the same properties. This consistency, known as reliability, is usually the degree to which a test yields repeatable results (i.e., the extent to which the same individual retested is scored similarly).

Reliability also refers to consistency across examiners/scorers. A reliable polygraph test should both yield equivalent outcomes when subjects are retested and, as well, be scored similarly by individuals other than the initial examiner. For example, if a polygraph examiner reviewed a set of charts and concluded that a subject was deceptive, any other polygraph examiner should be able to review the same charts and conclude that deception was indicated. This illustrates interrater-reliability. Such reliability might be lowered depending on the amount and type of training of examiners.

The study focused primarily on validity because if a testing procedure is not measuring what it purports to measure (validity), it matters little that it can measure the same thing again and again. Examiners who consistently agree that they are seeing "deception" may in fact be measuring anxiety or some other form of arousal. Reliability is, however, a necessary condition for validity to be established. A test that is valid will, necessarily, be reliable.

Construct validity. Construct validity refers, in broad terms, to whether a test adequately measures the underlying trait it is designed to assess. A polygraph test is designed to detect deception. It is therefore

important to clearly define the construct of deception, and distinguish it from other variables such as guilt.

To measure construct validity, it is necessary to both describe the construct and show its relation to a conceptual framework. Construct validation, thus, requires that a test be based on some theory or conceptual model. Since different types of polygraph tests have different theoretical bases (see section 2), there are multiple forms of construct validity for the polygraph. Construct validity is established by various means. Most importantly, based on theoretical predictions of how items should interrelate or how other tests should correlate with the measure of interest, actual evidence (e.g., scores from similar tests) is examined. If no such predictions are possible it is impossible to establish construct validity.

Criterion validity. Although from a theoretical point of view, construct validity is most important, from a practical point of view, criterion validity is the central component of a validity analysis. This aspect of validity refers, in the case of polygraph examinations, to the relationship between test outcomes and a criterion of ground truth. In this respect, criterion validity is what is meant by test accuracy. In the absence of construct validity evidence, however, it is difficult to determine to what extent criterion validity data can be generalized. In some situations, it is not clear which aspects of a test are responsible for accuracy, and what factors cause a test to be inaccurate.

Research Design

The above validity criteria are those which are typically assessed in considering evidence about the usefulness of a test. A related set of validity criteria are also used to evaluate the validity of any single study design. These research design criteria include, most importantly, internal and external validity (cf. Cook & Campbell, 1979; Saxe & Fine, 1981).

Internal validity refers to the degree to which a study has controlled for extraneous variables which may be related to the study outcome. External validity refers to the established generalizability of a study to particular subject populations and settings. Internal validity in the case of a polygraph test study is usually enhanced by the presence of control groups. Typically, such conditions of an experiment permit analysis of variables such as different question formats. In most field studies, internal validity is difficult to establish since the investigation cannot control, or in many cases, have definitive knowledge about, whether a subject is guilty or innocent.

External validity is simply the nature of the subjects and settings tested. The broader the population examined and the type of setting investigated, the more the study results can be generalized. In a parallel way, the more similar the research situation to the "real life" situation, the greater a study's external validity. Evidence about external validity is developed both from investigations that test a broad range of subjects and interactions with polygraph test outcomes. External validity is simply the nature of the subjects and settings tested. The broader the population examined and the type of setting or the more similar it is to the situation for which one wants to use a test or a theoretical construct, the greater a study's external validity.

False Positives and Negatives

With any test, the possibility exists of false positives and negatives. False positives are decisions that individuals are being deceptive when they provide truthful responses. Their charts are scored as showing a "deceptive" reaction for some other reason. False negatives are decisions that individuals are not being deceptive when in fact they are being deceptive. There are a number of reasons why such false outcomes might be obtained and, in part, they depend upon the criteria (e.g., amount of physiological change) used to indicate deception or truthfulness.

The rate of false positives or negatives is sometimes difficult to establish because in research studies, a number of criteria for deception/nondeception may be applied. Thus, for example, in studies which employ numerical scoring for polygraph charts, depending on the scoring system (i.e., cut off points), different diagnoses will be made. The rate of false positives and negatives may also depend on the examiner's perception of the "base rate" of guilt/innocence.

In some cases, the examiner will deal mostly with deceptive subjects (e.g., in certain criminal investigation contexts) and, thus may be predisposed to make false positive diagnoses. In other settings (e.g., some personnel screenings), an examiner may test only a small number of deceptive subjects and, then, may be predisposed to false negative decisions. Regardless of rates, assessment of conditions that contribute to either type of error is a focus of the research literature.

Reviews of Polygraph Validity

Since at least 1973, a number of polygraph researchers and psychologists interested in physiological detection of deception have reviewed available scientific literature to assess the validity and reliability of polygraph testing. Most such reviews focus on criterion studies of criterion validity, although a growing number of investigations deal with construct validity. The most important difference among these criterion studies has to do with whether they are conducted in actual field situations or in "analog" situations.

Field studies. For purposes of this report, field studies are those studies or "naturally" occurring polygraph test situations; that is, studies in which the researcher does not exercise experimental control over the situation in which the crime or other event occurred. Not exercising experimental control means that the researcher does not systematically assign people to conditions of, for example, guilt or innocence. We refer here to "field" studies but others (e.g., Ansley, 1983a) use the terminology "real" cases (vs. "laboratory"). Abrams (1973) differentiates between the laboratory and "actual criminal cases."

In polygraph field studies, polygraph examiners' decisions are compared against some post hoc determination of whether subjects are guilty or innocent; that is, "ground truth." These post hoc determinations may, in different studies, consist of confessions by the presumably guilty party, decisions by a panel of attorneys or judges assembled specifically for a particular study who base their decisions on investigative files excluding references to polygraph decisions, judicial outcomes (dismissals,

acquittals, convictions), as well as other criteria. The fact that determinations of guilt or innocence are made post hoc makes drawing conclusions from field studies difficult (Podlesny & Raskin, 1977). In real life situation, truth is seldom available (Ginton et al., 1982).

Attempts to use confessions, panel judgments, judicial outcomes and other criteria as indicators of truth have their own problems. People may confess to crimes which they did not commit (Lykken, 1981). They are sometimes falsely convicted (Borchard, 1932). Panel decisions may be generalizable only to cases in which sufficient investigative information is available to make a decision without the addition of polygraph testing. One can never be certain that the panel decision is indeed correct, and the panel and the polygraph examiner may have been exposed to the same prior information (Ginton et al., 1982). Thus, field studies provide the most direct evidence about polygraph test validity, they have been criticized because they do not adequately meet the standards of "ground truth" for the criterion validity.

Comparison of reviews. A number of independent reviews (listed in table 3.2) of the field evidence on polygraph testing were assessed in order to determine reasons for differences among reviews. The reviews differ in a number of respects. In part reviewers' conclusions differed because they included different kinds of studies and even different studies despite, in several cases, having had the same studies available to them. In addition, some reviews differentiated between accuracy in detecting deceptive versus nondeceptive subjects, emphasizing the problems of false positives and false negatives; other aggregated the overall accuracy rates across both groups of subjects. Finally, there are differences in the way accuracy rates were calculated, in particular, how inconclusives are handled. Each of these differences has important implications for the conclusions developed by the reviews.

Several reviews (Abrams, 1973; Horvath, 1976) conducted 5 to 10 years ago reported relatively positive conclusions based on an evaluation of the scientific literature.

Abrams (1973) reviewed reports of the polygraph's accuracy dating from 1917, including anecdotal as well as experimental data. He calculated approximate estimates of overall accuracy from this data, noting, however, that "it is almost meaningless to total and average these findings because of the great discrepancy in experimental paradigms and the instruments employed" (p. 320). He reported that in studies with complete verification of ground truth, diagnoses were 100% correct. In other field studies prior to 1963 Abrams calculated an accuracy rate of 98 percent. In laboratory experiments prior to 1963, Abrams estimated the average accuracy rate of 81 percent. Averaging the results of the reports between 1963 and 1973, Abrams' estimate of laboratory research accuracy was 83 percent. Horvath's (1976) review used somewhat more stringent criteria in selecting data than did Abrams. His review does not include an overall average accuracy rate calculated across studies.

Those early positive views of the polygraph's worth have recently been challenged by Lykken (1981) and, to some extent, by Ben-Shakhar et al. (1982). Lykken challenged the theoretical assumptions of the most prevalent question technique, the CQT, and asserted that an average 50-percent

false positive rate supported his theoretical challenge. Lykken, howver, continues to believe that particular polygraph techniques are useful (i.e., the detection of guilt by measuring physiological arousal) and offers the use of the guilty knowledge technique as a way to increase overall validity. Adoption of Lykken's suggestion would preclude the use of the polygraph for preemployment testing and periodic checking.

Ben-Shakhar et al.'s (1982) analysis also limited their assessment of the polygraph to the CQT. Their assessment of existing polygraph field research was that it showed that polygraph testing was 83 to 84 percent accurate for guilty suspects and 76 to 81 percent accurate for innocent suspects. As a result, Ben-Shakhar et al. concluded that examiners tend to value detection of guilty suspects highly, even at the risk of falsely classifying innocent suspects their conclusion concurs with Lykken's. Ben-Shakhar et al., in conducting their review, employ a utility theory approach based on Bayes' theorem. They predict dramatically different utility rates based on different base rate assumptions.

Although these recent reviews, by authors who are not professional polygraphers, cause doubt on the validity of at least the most common polygraph technique, a more recent review by Ansley (1983a) comes to the most positive conclusions since those of Abrams. Ansley's review is an important review because it represents the views of the National Security Agency's chief polygraph examiner. (NSA conducts the largest number of polygraph examinations of any Federal agency). As shown in Table 3.2, Ansley concludes that field research shows a 97.2-percent validity rate and laboratory research a 93.2-percent validity rate. Based on these validity calculations as well as separate calculations for reliability and utility, Ansley concludes that the polygraph is "clearly an excellent adjunct to the selection process."

Unfortunately, for the most part, polygraph reviews contained in Table 3.2 do not explicitly state their study selection criteria (see Glass et al., 1982; Light & Pillemer, 1980). The result is that a number of different studies have been included in various reviews, each of which presents different problems for interpretations of validity. The kinds of studies include reports of single criminal investigations in which the actual solution to the crime is the criterion for validity; studies in which "blind" polygraph interpreters compare their polygraph chart evaluations to "ground truth" as established by confession; and studies in which the judgment of legal professionals, actual judicial outcome, or in one case, the judgment of a single psychologist, is used to establish ground truth.

Some reviews do specify criteria for exclusion. Lykken, for example, does not include studies of single criminal investigations. Abrams, on the other hand, includes in his review a number of such studies (e.g., Bitterman & Marcuse; Larson, 1921). Lykken's reasoning was that in single criminal investigations, the examiner has a large chance of being accurate (depending on the number of suspects) merely by calling everyone innocent. The fact that other reviewers do not include Bitterman and Marcuse, and other such reports, implies that they accept Lykken's evaluation of the usefulness of such studies as indicators of validity. It is possible that results of such reports could be useful in assessing polygraph screening of large numbers of individuals in specific incident cases, such as might be

the case in unauthorized disclosure investigations. However, additional factors limit the external validity of Bitterman and Marcuse and other such studies. In Bitterman and Marcuse, for example, the investigators were psychology professors apparently conducting their first polygraph tests, and they did not use accepted polygraph procedures or instruments. There are no recent systematic studies of specific incident investigations involving a large number of suspects.

There is strong disagreement among reviewers about whether another group of studies should be included as indicators of validity. These studies were conducted with records selected from the files of the John E. Reid & Associates polygraph firm. A group of cases was used which the authors considered to be "verified" by confession of the guilty suspect (in most cases they were also verified by some form of corroboration; Buckley, 1983). The polygraph charts in these cases are then reinterpreted by a group of polygraphers who are "blind" (i.e., do not know) to the suspect's guilt or innocence. The degree of agreement of the "blind" evaluators to verify guilt or innocence is the test of validity. Two reviewers (Horvath, Lykken) explicitly excluded the group of studies conducted based on Reid files. Horvath excluded them because they used confessions as a criterion (confessions not being independent of the polygraph examinations), and Lykken because both examiners and "blind" evaluators were Reid polygraphers from the same firm. His claim was that the studies were, thus, "merely demonstrations that Reid's examiners score charts in a similar way" (Lykken, 1981, p. 123) and so were estimates of reliablity rather than validity. However, reviews by Raskin and Podlesny (1979) and Ben-Shakhar et al. (1983) each use all four Reid studies to assess validity.

Conclusions about the validity of the polygraph may depend on whether the reviewer attends to the average accuracy rate or to the accuracy for guilty and innocent subjects separately. The conclusions of all decision statistics contributes to the ability to make an accurate assessment of polygraph testing validity, particularly in view of the concern over both high false positive and high false negative detections. If, for example, the innocent correct rate is 80 percent but the remaining 20 percent consists of inaccurately calling innocent subjects guilty, a different policy conclusion may be drawn than if the remaining 20 percent consists of "inconclusives" or of false negatives. In some cases, for example preemployment screening, inaccurately designating nondeceptive people as deceptive may have worse consequences for the employee than inaccurately deciding that deceptive individuals are nondeceptive. In some cases (e.g., a heinous crime by a potential repeat offender, infiltration by a foreign agent), a false negative may have serious consequences.

In only two reviews (Ben-Shakhar, Lykken) are summary percentages provided in terms of the percent accurately detected for both guilty and innocent; in other reviews, these figures are presented as the average percent of accurate detections. In some cases, the percent inaccurately "detected" as nondeceptive (when they were really deceptive) or deceptive (when they were really nondeceptive) as well as present inconclusives were also reported by reviewers. But for purposes of clarity these have been omitted from Table 3.2.

Another reason reviews differ about the results of the same studies is the fact that they make different decisions about the base rate of

subjects or cases that are included. If, for example, a panel cannot make a decision about 30 percent of the cases (e.g., Barland & Raskin, 1976), some reviewers will omit the number of nonagreements from the number included in the accuracy rate and base accuracy percentages on only the remaining cases. This accounts for the difference between Horvath and Ben-Shakhar et al. analyses of the Barland and Raskin results. In other studies (and reviews of those studies, e.g., Ansley, Abrams) inconclusive polygraph results are excluded from the analysis. This has the effect of inflating the accuracy rates.

Apart from the different base rates on which most of the reviewers calculated accuracy rates (see above), one source of different accuracy rates applies uniquely to Ansley (1983a). In any case in which there is not 100-percent accuracy by dividing the difference between the accuracy rate and 100 percent (the so-called error rate) in half and adds half of the difference to the accuracy rate. Ansley uses this procedure on the grounds that on the basis of chance, errors were probably half in favor of the panel (or other criterion measure) and half in favor of the examiners. For example, in the Bersh study, half of the difference between the typically reported 92.4-percent rate and 100 percent is 7.6 which Ansley divided in half, leaving a validity rate of 96.2 and an error rate of 3.8 percent. The same method is used for the Peters, Elaad and Widacki studies, for which the preadjustment validity rates are 90.2 percent and 96.6 percent and 91.6 percent, respectively. Each of these studies, particularly Elaad (see section 4) have other problems of interpretation as well.

Conclusions

Central to legal, legislative, and scientific assessment of polygraph tests are their validity. Yet, despite many decades of judicial, legislative and scientific discussion, no consensus has emerged about the accuracy of polygraph tests. One explanation is that scientific criteria for validity deal with a number of dimensions and that the criteria vary widely among specific research studies. In order to assess overall polygraph examination validity, it will be necessary to examine details of each of the relevant studies. Such analysis is presented in sections 4 and 5.

Another explanation is that polygraph testing has been viewed as a single technique. Thus, despite testimony (e.g., Raskin, 1978) which urged differential consideration of polygraphs used in, for example, employment screening and criminal investigations, the scientific evidence for particular purposes has not been differentiated. As is demonstrated by the analysis of scientific literature (here and in sections 5 and 6), in assessing validity it is necessary to separate clearly the purposes for which polygraph examinations are conducted and the types of techniques employed.

FIGURE 3.1
DEBATE AT THE FEDERAL LEVEL
1963 - 1983

Legislative Activity on Private Employer Use	Legislative Activity on Federal Agency Use	Executive Branch Activity
	1963-1964	
	Study and hearing by House Government Operations Subcommittee on Foreign Operations and Government Information	Proposal to use polygraph to investigate leak
	1965	
	Report issued by House Government Operations Committee	DOD Directive 5210.48 DOD Research Program Civil Service regulations
	1966	
	1967-1974	
Some bills would ban use of polygraphs by private employers	Bills introduced which would restrict or ban federal government use of polygraph	
	1974	
Report referred to private as well as Federal employer	Report Privacy, Polygraphs and Employment issued by Senate Judiciary Subcommittee on Constitutional Rights Study and hearings by House Subcommittee on Foreign Operations and Government Information	
	1975	
		DOD issued revisions to Directive 5210.48

FIGURE 3.1
DEBATE AT THE FEDERAL LEVEL
1963 - 1983

Legislative Activity on Private Employer Use	Legislative Activity on Federal Agency Use	Executive Branch Activity
	1976	
	Report issued by House Government Operations Committee	
	1977-1978	
S.1845 introduced	S.1845 introduced; would bar Federal agencies and private employers from requiring or requesting	
Hearings	Hearings held on S.1845 by Senate Judiciary	
Privacy Protection Study Commission issues report		
	1979	
Hearings held by House Education and Labor Sub-Committee on Labor-Management Relations	Hearings and Report by House Permanent Select Committee on Intelligence Subcommittee on Oversight	
	1980	
		Director of Central Intelligence study Investigative Scope and Adjudicative Procedures Among Intelligence; Personnel Security Survey

FIGURE 3.1
DEBATE AT THE FEDERAL LEVEL
1963 - 1983

Legislative Activity on Private Employer Use	Legislative Activity on Federal Agency Use	Executive Branch Activity
	1982	
	Hearings by House Judiciary Subcommittee on Civil and Constitutional Rights on DOD proposed revisions	DOD Select Panel Review of the DOD Personnel Security Program Deputy Secretary of Defense issues memos instituting aperiodic polygraph exams for those with SCI clearance DOD prepares proposed revisions to Directive 5210.48
	1983	
HR2403 introduced; would ban use of polygraph by private employers	Hearings by House Post Office and Civil Service Subcommittee on Civil Service and House Judiciary Subcommittee on Civil and Constitional Rights on NSDD-84 Congress passes amendment to DOD appropriations bill prohibiting changes in DOD polygraph regulations until April 15, 1984 Hearings by Senate Committee on Government Affairs on NSDD-84	President issues NSDD-84 DOD prepares further draft of revised regulations

Table 3.2 Reviews of Field Studies of Polygraph Validity

Study[2] /N[5]	Year	Q[3]	D[4]	Abrams (1973)	Horvath (1976)	Raskin & Podlesny (1979)[1]	Lykken (1981)	Ben-Shakhar (1982)	Ansley (1983)[9]
				Reported "hit rate" (%)[6]					
Bitterman & Marcuse (N=2Ex, 81S)	1947	R/I		[100]	NR	NR	EX[7]	NR	[100]
Ben-Ishai (N=1E, 100C)	1962 unpub	C	A	NR[8]	NR	NR	NR	NR	[97.0][9]
Ben-Ishai (N=1E, 100C)	1962 unpub	C	A	NR	NR	NR	NR	NR	[100]
Bersh (N=4E, 216C)	1969	C	G I A	NR	[74.6-92.4][10]	NR	[75-92][10]	85.6 89.3	[96.2][9]
Horvath & Reid	1971	C	G I	[79.1-91.4][11]	EX[12]	89-94[11]	EX[12]	85.0	NR
Hunter & Ash	1973	C	G	-	EX[12]	88 86	EX[12]	87.1 85.7	NR
Wicklander & Hunter	1975	C	G I	-	EX[12]	95 93	EX[12]	90.0 86.7	NR
Slowick & Buckley	1975	C	G I		EX[12]	85 93	EX[12]	83.8 90.5	NR
Raskin (3 studies)	1976[14]	C	G I	-	-	93-100[13] 69-95[13]	NR	NR	NR

Barland & Raskin (N=102C)	1976	C	G I A	–	[78.7–89.7][15]		98 45* [72]	83.0 29.4	NR
Horvath	1977	C	G I	–	Ex[12]	NR	77 51	77.1 51.1	NR
Davidson	1979	C		–	–	–	NR	NR	Ex[12]
Edwards (N=80E, 959C)	1981 unpub		A	–	–	–	NR	NR	[98.3][16]
Elaad, (N=184C)	1976 unpub		A	–	–	NR	NR	NR	[98.3][17]
Peters, (N=172C)	1982 unpub		A	–	–	–	–	–	[95.1][9]
Widacki, (N=38C)	1982	Z	A	–	–	–	–	–	[95.8][9]
Author's conclusion (if any)			G I A	98**	NR	90 89	64–72[19]	83–84[18] 76–81[18]	97.6

Notes: (1) Not technically a review. Original review (Podlesny & Raskin, 1977) relied on lab evidence. When criticized by Lykken, authors responded in part by citing field studies.

(2) Presented in chronological order.

(3) Q = Question type (C = Control Question; Z = Zone of Comparison).

(4) D = Correct decision or rate of agreement (i.e., hit rate) for guilty (deceptive) (G) or innocent (nondeceptive) (I) subjects. A = Average hit rate (in cases where G and I are not reported separately.

(5) N = Number. Ex = Number of examiners; E = Number of evaluators; C = Number of cases evaluated; S = Number of subjects when actual subjects are tested in this study.

(6) Different studies use different criteria for validity. Thus, the hit rate can be either polygraphers' agreeement with ground truth (e.g., Bitterman & Marcuse), blind evaluators' agreement with polygraph examinations verified by confessions (e.g., Horvath & Reid, Hunter & Ash,

Wicklander & Hunter, Slowick & Buckley, Horvath), or polygraph examinations validated by a panel of judges using other criteria (e.g., Bersh, Barland & Raskin) or actual judicial outcome (Widacki). Researchers have not agreed on what constitutes an appropriate criterion for validity in the field.

(7) Excluded by Lykken as not useful evidence because of base rate problem. The examiners in the study assumed only one person was guilty. Therefore, they would have had a 99 percent accuracy rate even if they had called everyone innocent.

(8) NR = Not reported by the reviewer indicated.

(9) It is important to note Ansley's method of calculating overall accuracy. When there are inconclusives (or disagreements between the original examiners and the evaluator), Ansley divides the percent inconclusive by 2 and adds one-half the percentage to the accuracy percent reported by the original researcher.

(10) 74.6 percent figure is for majority decision; 92.4 percent for unanimous decision.

(11) The lower rate is the accuracy rate for inexperienced polygraphers; the higher rate is the accuracy rate for experienced polygraphers.

(12) Excluded because they are seen as reliability studies. Horvath reasons that they used confessions as a criterion and because confessions are not independent of polygraphers' examinations "their usefulness as a criterion measure for estimating validity is limited" (p. 145). Lykken excludes them because both examiners and evaluators were Reid polygraphers.

(13) Lower rate is for non-numerical evaluation; higher rate for numerical evaluations.

(14) Focus of study was to test the "friendly polygrapher" hypothesis (see Raskin, Barland & Podlesny, 1978).

(15) 78.7 percent = agreement between simple majority on panel and polygraph decision; 86.7 percent = agreement between 4/5 majority on panel and polygraph; 89.7 percent = agreement between judicial outcome and polygraph (where judiciary not aware of polygraph results).

(16) Of 2,433 polygraph examinations in the original pool, only 959 (39.4 percent) could be verified as to truthfulness or not; of these 959, 943 were found to be correct. The 98.3 percent figure reported by Ansley represents 943/959 and does not take into account the entire pool of examinations.

(17) Actually, of 1984 cases that were confirmed, 10 were inconclusive and 6 were errors. The percent accuracy given in Ansley's review excludes the inconclusive results and adds half the error rate to the accuracy rate. If one appropriately excludes both the inconclusives and the error rate, the accuracy rate is at most 87 percent.

(18) The lower rate is for studies using confessions criteria; higher rate is for studies using panel of experts criteria. Figures are rounded.

(19) Lie control test only.

* our calculations based on Lykken's conclusion that 6 out of 11 innocent subjects were judged deceptive.

** Abrams uses many other studies to come to this conclusion. Many are anecdotal. Other researchers (e.g., Horvath) say they should not be relied on.

REVIEW AND ANALYSIS OF POLYGRAPH FIELD STUDIES

Introduction

As noted in the discussion of previous scientific reviews of polygraph validity, considerable disagreement exists among reviewers as to which field studies and what kinds of evidence constitute acceptable tests of validity. This section presents the results of a systematic analysis of existing field studies of polygraph testing in order to make an independent assessment of validity. Field studies investigate actual polygraph examinations and constitute the most direct evidence for polygraph test validity (Ben-Shakhar, et al., 1983). Both quantitative and qualitative techniques are utilized in order to make an overall assessment of existing evidence (Glass, McGaw & Smith, 1981; Pillemer & Light, 1980; Rosenthal & Rubin, 1982).

The goal of this analysis is to synthesize available research. Almost all of the available field evidence comes from cases involving specific-incident criminal investigations using the control question technique (CQT). This is an important limitation. Because a systematic review helps to identify this kind of problem, researchers and policy makers have a better basis on which to determine what, if any, additional techniques, test purposes, question designs, and scoring techniques have been studied and which may require further research. The analysis is designed to address many of the problems associated with qualitative or "literary" reviews of the research literature previously discussed. In particular, the analysis makes explicit the criteria used for both study selection and data analysis (cf. Glass et al., 1981; Pillemar & Light, 1980; Rosenthal & Rubin, 1982).

Study Selection

Studies were considered field studies of validity if their sample consisted of actual instances of polygraph examinations conducted by professional polygraph examiners, used field-tested polygraph techniques, and used some independent criterion to assess guilt or innocence. Although ground truth can probably never be known in an absolute sense, studies can be considered studies of validity only if they included some adequately described and systematically determined criterion of "truth" (e.g., panel decision, judicial outcome, confession). Studies in which judgments of one set of polygraphers are correlated with anothers' with no independent criterion of guilt or innocence are, in effect, reliability studies. Such studies have been excluded from the primary analysis reported here. Reports of unsystematically collected cases from police agencies and other organizations, in which the criteria for verification are unclear or unsystematic, have also been excluded.

The population of field studies considered for the present analysis was, in general, taken from those studies referred to in existing reviews of the scientific literature. In addition, researchers active in the field of polygraph research were contacted and asked to supply the names and publication information of any additional recent studies. A bibliography provided by the American Polygraph Association (Ansley, Horvath & Barland, 1983) was also searched for references to field studies of validity. The 10 studies finally included (and listed in Table 4.1) in the analysis are:

Barland and Raskin (1976), Bersh (1969), Davidson (1979), Horvath (1977), Horvath and Reid (1971), Hunter and Ash (1973), Kleinmuntz and Szucko (1982), Raskin (1976), Slowik and Buckley (1975), and Wicklander and Hunter (1975). The following sections briefly describe the studies excluded from the analysis and the kinds of studies included in the analysis.

Studies Excluded

Not all studies referred to as field studies or actual criminal investigations by other reviewers are included in the present analysis. A comparison of studies shown in Table 3.2 and the ten studies included in the present analysis indicates that eight studies included by one or another of the reviewers are not included. The excluded studies are Bitterman and Marcuse (1947), Ben-Ishai (1962), two analyses reported in Raskin (1976), Edwards (1981), Elaad and Schahar (1976), Peters (1982) and Widacki (1982). One study, Kleinmutz and Szucko (1982), not included by various reviewers (because of its recent publication) has been included here. In addition, a number of studies included by Abrams (1973), not shown in Table 3.2, are also excluded from the present analysis. Many of the studies Abrams cited are excluded by later reviewers (e.g., Horvath, 1976) because they are not actual validity studies (and did not use external criteria of "guilt/innocence," e.g., MacNitt, 1942), they did not use appropriate polygraphic instrumentation (e.g., Summers, 1936), or did not use testing procedures common today (e.g., Lyon, 1936). Other studies used by Abrams, but excluded from the present analysis, were unverified self-reports published in popular magazines (e.g., McEvoy, 1941), or surveys of attitudes towards validity of the polygraph (e.g., Cureton, 1953).

The Bitterman and Marcuse (1947) study was excluded because, as pointed out by Lykken (1981) and others (e.g., Horvath, 1976), studies of single crimes for which there is only one possible guilty person raises the probability of accurate deception, regardless of method used, to a level too high for the study to provide valid information. To give an extreme example, if there is one guilty suspect among 100 examined, making an a prior decision to call them all innocent yields a 99 percent accuracy rate. In addition, Bitterman and Marcuse did not meet present criteria for field studies because the polygraphers were not professional examiners (they were psychology professors who had read books and articles about the polygraph technique), and they did not use field-tested measures of physiological response.

Ben-Ishai's (1962) paper reports on two studies, both of which were excluded. One consisted of blind evaluations by Ben-Ishai of 10 polygraph charts. It is more accurately described as a study of reliability. The other used a single psychologist's (Ben-Ishai's) judgments of guilt or innocence based on investigative files as the criterion by which to judge polygraph accuracy. It is difficult to justify use of the judgment of a single psychologist as an adequate criterion of ground truth. Likewise, the information used to establish ground truth for the Elaad, Peters and Widacki reports is not systematically collected, inadequately described, and are more accurately described as a set of anecdotal reports. They use samples of cases collected from police files which are described as having been verified, sometimes by judicial outcome (Widacki), in others by confession (Elaad), and in the Edwards study, by "independent means."

A final set of studies excluded are two of the three studies by Raskin (1976). One analysis was directed primarily at an assessment of whether polygraph examinations are more favorable to defendants when conducted by polygraph examiners chosen by defense attorneys than when they are conducted by examiners chosen by prosecutors (the so-called "friendly polygrapher" hypothesis). The purpose of the second analysis was to discover the source of decision errors; these findings are discussed in section 6. The Raskin study included in the present analysis (Raskin, 1976) was conducted with the only 16 cases from Barland and Raskin's (1976) sample able to be verified by confession.

Studies Included

The field studies included are listed in Table 4.1 in terms of the criterion used, the type of initial examiner decision, and the types of case selected. These characteristics of studies relate to criterion, construct, and external validity, respectively.

The criterion dimension refers to the operationalization of ground truth used in a study. In one type of validity study, polygraphers' original decisions are compared against a criterion of ground truth established by a panel of experts (e.g., lawyers and judges). The panel makes their judgment on the basis of information in an investigative file, from which polygraph results are excluded. In another type of field study, the criterion consists of a second set of examiners' evaluations of charts taken from a file. In most cases, the evaluation is "blind"; that is, the examiner/evaluator does not know the original examiner's decision, the disposition of the case, nor any other information about the subject. In this situation, the original decisions have been verified by confession of the guilty party. Verification by confession is used as the ground truth criterion. In the third, and the least common type of field study, original examiners' decisions (the construct validity component) are judged against guilt or innocence established by judicial outcome, which is the ground truth criterion.

Whether one uses examiner decisions or physiological recordings (above), depends on whether one is testing examiner decisionmaking or physiological arousal in response to certain questions. Blind evaluations of charts are probably less useful because, in the typical examination situation, the decision as to suspects' deception is made by the original examiner and not by a blind evaluator. Even when examinations are subject to review (e.g., quality control procedures used in DOD), final decisions are still based on review of all information. Although a blind analysis is the first task of the quality control office, such quality control reviews do not fully control for the impact of a variety of factors, such as interpersonal expectancy effects which would still be reflected in the original polygraph charts. Interpersonal expectancy effects (Rosenthal, 1976) refer to the possibility that an examiner's pre-examination decision concerning guilt or innocence affects construction of examination questions or the psychological state of the suspect. Either of these could affect a suspect's physiological responses. Therefore, in studies for which results of both original examinations and blind evaluations were included, as in Barland and Raskin (1976), the present analysis uses results of the original examinations instead of those for blind evaluations. It should be noted, however, that in these cases it is difficult to determine to what extent

the decisions are based on the charts and to what extent they are based on interaction with the suspect (see Ben-Shakhar et al., 1983; Kleinmuntz & Szucko, 1982).

Operationalizations of ground truth (the criterion component of validity are as problematic as those for construct validity. Studies using panel decisions have been referred to as the only valid field research (Horvath, 1976), yet there is no way to know whether panel decisions based on investigative files are, in fact, correct. Raskin (1982) notes some of the problems with using judicial outcomes and other criminal justice system resolutions (dismissals, guilty pleas) as criteria for validity. Cases may be dismissed for lack of sufficient evidence rather than actual innocence. Accordingly, if a jury acquits a defendant, it is not possible to determine the extent to which the jury felt that the defendant was actually innocent or whether they felt that there was not enough evidence to meet the standard of "guilty beyond a reasonable doubt." Many guilty pleas are actually confessions of guilty to (lesser) crimes; as Raskin notes, it is difficult to interpret the meaning of such pleadings in regard to guilt on the original charge. The result is that, using criminal justice system outcomes, polygraph examinations may appear to have a high number of false positives (in the case of acquittals), or false negatives (in the case of dismissals).

The use of confessions, the most frequently used criterion of ground truth, is problematic in three ways: 1) confessions, themselves, are not always valid; 2) if the confession occurs prior to or during a polygraph examination, it can not be considered an independent measure of guilt; and 3) those who confess may be a select sample of subjects, as discussed further below.

In addition to the above problems with the criteria, the construct, and types of subjects, studies differ in the adequacy of their research design. The most serious problems concern sampling. In most reported studies, neither cases, examiners nor evaluators were selected randomly. In some studies (e.g., Horvath & Reid, 1971; Barland & Raskin, 1976), the cases of only one examiner are sampled. Non-random selection leaves open the possibility that the studies are not investigating "polygraph testing," in general, only a subgroup of practitioners or testing techniques. When random sampling is used (as in Bersh, 1969), high rejection rates of cases selected for analysis create other sample bias problems.

Some sample selectivity of unknown magnitude and importance occurs when confessions are used as a criterion. Studies using confessions may be using only a select sample of examinations. The magnitude of this problem is illustrated by the fact that in the sample of 92 cases obtained by Barland (Barland & Raskin, 1976; Raskin, 1976) only 16 were able to be verified by confession (Raskin, 1983).

To summarize, because of problems in operationalizing important components of validity, none of the field studies of validity can be taken by itself as an indication of polygraph testing validity. In addition, because of the different operationalizations of construct and criterion validity and variations in research design, the studies are not strictly comparable with each other. These studies, however, constitute the most direct evidence for validity currently available and are analyzed as a group in order to assess the current state of knowledge about polygraph testing.

Coding

In order to conduct the present analysis, each field study was coded for a number of variables which had either been referred to as important factors in previous reviews of the literature, or which were deemed relevant to the various components of validity described in section 3. If the needed information was not available from the studies as published, the study author(s) were contacted and asked to supply the information. Appendix C lists the coding categories and relevant validity components (panel decision or judicial outcomes; confession), as well as design information (sample selection, attrition rate, examiner/evaluators' knowledge of base rate of guilt). All codings were made by two reviewers and each instance of disagreement over coding was resolved before analysis.

Data were coded directly from information provided within the study report or from information directly provided by the authors, with the exception of one variable. The exception was the coding category "objectivity of ratings," which required that the coder make a judgment from high objectivity to low objectivity. Scoring was judged high if some actual standardized measurement (e.g., using a ruler) was taken of the physiological recordings on the polygraph charts. A rating of medium was given if numerical scores were assigned to subjective assessments of suspects' guilt or innocence (see, e.g., Barland & Raskin, 1976; Kleinmuntz & Szucko, 1982), low if ratings of deceptive or nondeceptive were based on global assessments of charts only, and very low if decisions were based on charts plus other available information (in particular, observation and interaction with the subject). Objectivity ratings were made both for the original examiners' judgments and the blind evaluators or judges.

Finally, six categories of outcome data from each study were recorded: 1) guilty/deceptive subjects judged correctly; 2) guilty/deceptive subjects judged incorrectly (i.e., judged nondeceptive); 3) guilty/deceptive suspects judged inconclusive; 4) innocent/nondeceptive subjects judged correctly; 5) innocent/nondeceptive subjects judged incorrectly (i.e., deceptive); and 6) innocent/nondeceptive subjects judged inconclusive. Categories (2) and (5) are the false negative and false positive rates, respectively.

Findings and Discussion

Three questions are of particular importance to an assessment of polygraph validity useful to policymakers. One is the global question, are polygraph examinations valid? The second is, given the wide range of outcomes reported across studies, what accounts for their variability? The third is, how generalizable are the results of studies to the current and proposed uses for national security purposes?

To answer the first question, data from the available field studies were analyzed to ascertain whether polygraph examination accurately differentiate deceptive suspects from nondeceptive subjects. For this analysis, the outcome frequencies for each category were converted to percentages and average percentages within each category were calculated. A measure of predictive association (lambda$_b$ see Goodman & Kruskal, 1954; Hays, 1981) was also calculated.

The $lambda_b$ index shows the proportional reduction in the probability of error in predicting one category (in this case, deception) when a second category (in this case, polygraph examination results) is known. If the information about the second category does not reduce the probability of error in predicting the first category at all, the index is zero, and one can say that there is no predictive association. On the other hand, if the index is 1.00, no error is made in predicting one category from another, and there is complete predictive association. Essentially, lambda provides an index that translates to the percent improvement over the base rate and indicates the percent improvement in prediction when the polygraph examinations are considered versus no further information. There is almost no direct research on the percent improvement of the polygraph over other forms of investigation (cf. Widacki & Horvath, 1978). The results of this analysis of predictive association are shown in Table 4.2. The average $lambda_b$ across studies is .65, which means that, on the average in these field studies, the polygraph diagnosis reduced 65 percent of the error of chance prediction.

To summarize, the analysis of the 10 field studies included in the analysis indicates that while polygraph examinations using the CQT in criminal investigations detect deceptiveness and nondeceptiveness better than chance, there is a high error rate, particularly for nondeceptive subjects. The one study which tested the validity of the R/I technique (the GQT portions of the Bersh study) also detected deceptiveness and nondeceptiveness better than chance.

Variation Among Studies

As implied in the introduction to this section, the use of a single statistic or summary to describe the results of field tests of validity may be misleading. As shown in Table 4.1, although the field studies of polygraph validity are similar in that almost all of them tested control question techniques in criminal investigations, they differ in operationalizations of ground truth and type of examiner decision. The result is that there is a great deal of variability in the results of studies. Correct guilty detections range from 70.6 percent in one condition of the Bersh study to 98.6 percent in a condition of the Wicklander and Hunter study. Correct innocent detections are even more variable, ranging from a low of 12.5 percent in the Barland and Raskin judicial outcome study to a high of 94.1 percent in one condition of the Bersh study. Table 4.3 indicates the range of incorrect judgments and inconclusives among studies. False negatives range from 29.4 percent of the Bersh study to zero percent. False positives range from 75 percent in Barland and Raskin (1976) to zero percent in two studies. Inconclusives range from zero to 25 percent. This section compares studies that used comparable operationalizations of construct and criterion validity in an attempt to discover reasons for the range of results. However, even using this method results in considerable variability. Ideally, lambdas (see above) could be calculated for each study or for conceptually related groups of studies (i.e., using examiners from the same schools, different types of scoring techniques). The main point, however, is that no field studies exist to directly test the situations for which DOD and the President propose to expand polygraph use.

Studies Using Panel Criterion and Examiners' Decisions

Both Bersh (1969) and Barland and Raskin (1976) used a panel to establish the criterion for validity in their studies. The makeup of the panels and the polygraph scoring systems were similar in each study. In the Bersh study, which validated polygraph examinations conducted by military examiners, the panel consisted of four Judge Advocate General (JAG) Attorneys; Barland and Raskin's panel consisted of two criminal defense attorneys, two criminal prosecuting attorneys, and a judge. The examiners in the Bersh study used either the GQT (a version of the R/I) or the ZOC technique; for all but one subject in Barland's study, the Federal ZOC control question technique was used and results evaluated using the Army scoring procedure. Assuming the accuracy of the panel's decisions, the two studies' results are strikingly different. Barland and Raskin attained accuracy rates of 91.5 percent for guilty and 29.4 percent for innocent subjects; comparable figures in Bersh's study are 70.6 percent guilty correct and 80 percent innocent correct. It is not clear why there should be this variation, although differences in the nature of the cases, the completeness of the case files, and sample selection may account for some of the differences.

In the Bersh study, cases were initially drawn at random from a pool of criminal investigations conducted by the three military services over a period of 3 years (1963-66); then, any cases which had been judged "indeterminate" by the original polygraph examiner were eliminated. In addition, after polygraph charts were removed from the investigative files, a preliminary panel of judges eliminated from the sample all files containing insufficient evidence to warrant a positive determination of guilt or innocence. Only those cases which resulted in a unanimous decision by the initial JAG panel were retained in the validation sample. Altogether, one-quarter of the cases (80 cases out of 323) were eliminated because of insufficient evidence. This figure does not include the number initially eliminated on the basis of inconclusive polygraph examinations.

In Barland and Raskin's (1976) study, the initial pool of subjects consisted of 102 (nonmilitary) criminal suspects referred to Barland by police, defense or prosecuting attorneys. These cases represented the entire population of Barland's cases at that time. Then, 92 of these 102 cases were retained for further analysis on the bases of independence (a case was considered independent where two or more subjects had not been examined regarding the same crime). In one respect (the fact that there was only one examiner), Barland and Raskin's sample was less variable than Bersh's. However, Barland and Raskin did not eliminate from consideration indeterminate examinations. Neither, and perhaps more importantly, did Barland and Raskin eliminate cases in which investigative files without the polygraph were inadequate. As Barland (1982) points out, many of the investigative files that were given to the panel were incomplete. The files had been compiled by inexperienced student assistants who often did not know where to obtain necessary information. The officials responsible for providing the information were, more often than not, unavailable or, when they were available, unable to recall the details of a crime. In many cases, few details were available. As a result, one-third of the 92 cases were judged inconclusive by the panel merely on the basis of the investigative files. The figures reported in Table 4.3 are for 64 of the original 92 cases.

It is not clear why there should be an inverse relationship between accurate detection of guilty and innocent suspects in the two studies. It may be that both the panel and the examiner in the Barland and Raskin study consistently tended to presume guilt in the absence of any a priori base rate (see Ben-Shakhar et al., 1982; Szucko & Kleinmuntz, 1981). The cases in the Bersh study, on the other hand, were intially selected to be equally distributed among deceptive and nondeceptive cases. It is not reported whether the panel was aware of the base rate in the Bersh study.

Studies Using Confession as a Criterion and Blind Evaluations

The remainder of the field studies analyzed tested the validity of polgyraph testing by comparing the blind evaluations of polygraph examiners against a criterion of verification by confession. Two exceptions are Barland and Raskin's judicial outcome analysis and one condition in the Wicklander and Hunter study. The confession studies vary somewhat as to source of verified files. The Horvath and Reid, Hunter and Ash, Slowik and Buckley, Wicklander and Hunter, and Kleinmuntz and Szucko all used files from polygraph testing firms. Horvath's cases came from police files, Davidson's from military cases, and Raskin's from the Barland cases reported in Barland and Raskin (1976; discussed above). The first four studies used files from the firm of John E. Reid & Associates and involved various criminal offenses. The firm used by Kleinmuntz and Szucko is not identified; all of their cases involved theft.

In the first four studies, blind examiner evaluators also came from John E. Reid & Associates. The Reid studies did vary with respect to case selection. Only one study (Slowik & Buckley) reports random selection of cases; in other studies, the cases of only one or two examiners were used. Horvath's (1977) blind evaluators were field-trained examiners with a median of 3 years experience, all of whom specialized in conducting polygraph examinations for police agencies. The 25 evaluators in the Raskin (1976) study were volunteers who had trained in a variety of places.

The results of the Reid studies do not vary substantially. The greatest deviation from the mean occurred in one condition of the Wicklander and Hunter study in which examiner/evaluators were given additional information about the suspects (verbal and nonverbal behavioral indicators, demographic information) and the cases. This difference however, was not statistically significant. Even so, it may be reasonable to consider it separately from the other Reid studies, because of the extra information available to evaluators. In the Reid studies, guilty correct identification rates ranged from 84 to 87.1 percent, with an average of 86.5 percent (excluding the 98.6-percent Wicklander result; 88.9 percent including it). The innocent correct rates in the Reid studies range from 85.7 to 90.7 percent with an average of 88 percent. There is no difference when the Wicklander and Hunter condition is included.

An additional difference of note among the Reid studies concerns the false negative rate, which is highest in the studies which either used random selection of cases (Slowik & Buckely) or eliminated the most clearcut charts from their original selection (Horvath & Reid). There is no apparent explanation for the variation in false positive rates in the Reid studies, which ranged from 5 to 14.3 percent.

The Davidson study results are basically similar to those of the Reid studies, except for the absence of false positives. However, the study should be interpreted with caution as one-third of originally (randomly) selected sample was not able to be used.

The Horvath (1977) and Kleinmuntz and Szucko (1983) studies have the lowest accuracy rates. As with the Barland and Raskin (1976) study, the low accuracy rate may be related to the fact that Horvath selected his sample from police files. Perhaps, police records of verification are not reliable, or have greater variability than those of polygraph firms.

Barland (1982) has suggested a number of reasons why Horvath's results are lower than the Reid studies. One reason is that the blind reviewers did not have access to "special charts" administered in 32 percent of the cases, primarily to subjects the original examiner considered deceptive; these charts were removed from the files before being reviewed by blind examiners. According to Barland, Horvath's original exmainers had been 100 percent correct in their judgments. A second reason is that, as noted above, police examiners were used instead of private examiners; the differences between the two kinds of examiners is not explained further. Yet a third reason, which Barland (1982) believes may be the most important in terms of false positives, is that a number of victims and witnesses were included in the sample (i.e., were subjects). According to Barland (1982), one theory of detection of deception predicts that innocent victims or witnesses may react emotionally during a polygraph examination when asked about the events they experienced or witnessed precisely because they are telling the truth. It is not clear, however, why this theory would not explain false positives among suspects as well. Nonetheless, an analysis suggested by Barland, comparing results for victims and witnesses with those for suspects, would be of interest (see Giesen & Rollison [1980] for a comparison of innocent associations with guilty knowledge).

Despite the generally anomalous results of Horvath's (1977) study, an interesting finding may help to account for the results of the Kleinmuntz and Szucko (1982) study. Horvath found that suspects in crimes against property were less detectable than suspects in crimes against persons. This may be because crimes against persons are likely to have a greater amount of affect associated with them, and are, thus, more physiologically detectable. Barland and Raskin (1976), on the other hand, found no differences by type of crime. As noted previously (see Table 4.1), Kleinmuntz and Szucko's (1982) study selected only cases from the files of a polygraph firm involving crimes of theft. However, although the crimes against property hypothesis is suggestive, it may not fully explain the difference between Kleinmuntz and Szucko's and similar studies. The Davidson study, for example, only used theft cases, and it has a "0" false positive rate (although it has a substantial inconclusive rate). Analyses of other studies by crime type would be informative, although the number of cases would probably be too small for a meaningful analysis.

Szucko (personal communication, 1983) has suggested that one possible reason his results are so different from other polygraph firm studies' results, is that the individual who selected the charts in the Kleinmuntz and Szucko study could not read polygraph charts. Therefore, case selection may have been more variable than in some of the other studies. Alternative explanations are that: a) Kleinmuntz and Szucko only evaluated one chart

for each subject (at least three is standard); and b) their evaluators were examiner/trainees at the end of their internship period, not experienced examiners* (see Kricher, and Raskin, 1983).

Studies Using Judicial Outcomes and Original Examiners' Results

Barland and Raskin's (1976) analysis using judicial outcomes as a criterion has the lowest accuracy rate for innocent suspects--a 12.5 percent innocent correct and 75 percent false positive rate. The problems with judicial outcomes as a criterion have already been referred to, in particular, the fact that the judicial outcome is not a highly accurate measure of guilt because of such characteristics of the legal system as the necessity for proof beyond a reasonable doubt, and the prevalence of plea bargaining. These problems are illustrated here by the fact that only 41 of Barland and Raskin's original 92 cases were resolved by the criminal justice system. Again, there is clearly greater agreement on guilty subjects.

Other Considerations

Although the analysis above demonstrates that polygraph testing is better than chance at differentiating deceptive from nondeceptive subjects in criminal investiagtions, substantial false positive and false negative rates are obtained in several investigations. Although it is not possible to determine a "scientifically" acceptable rate of correct or incorrect judgments, clearly if error rates are between 10 and 25 percent, a large number of incorrect decisions would be made if the polygraph were widely employed. The base rate of guilt in actual situations may further complicate matters. It is not clear from the field studies conducted so far how many suspects were involved in the cases selected for polygraph testing, but if there were a large number of suspects, more false positives could be expected (see section 7).

Also problematic is the wide variability in accuracy rates across studies. Although some differences can be explained methodologically, other differences cannot. Of perhaps even greater importance than the accuracy rate variability and error rate problems is the observation that field studies of polygraph testing have only been conducted in criminal investigations. As is discussed more fully in section 6, criminal investigations may generate different levels of affect. In addition, different kinds of subject groups may be the focus of expanded Government use of polygraph testing. Only two field studies can be identified that relate directly to polgyraph testing in the national security area (Director of Central Intelligence, 1980; Edel & Jacoby, 1975). Neither of these is a validity study but because they are the only studies with any relevance to national security, they will be described below in some detail.

The DCI study consisted of a survey of 12 Government agencies (not including NSA). The study was conducted to evaluate the relative effectiveness of various means to conduct background investigations for purposes of applicant screening and security clearances for current employees. Background investigations are conducted through the use of personnel interviews, interviews with present and former neighbors, checks of educational

*Some maintain that the evaluators in Kleinmuntz and Szucko's study were even less experienced than that.

and work records and checks with a consortium of other national agencies (the so-called National Agency Check). Of the agencies surveyed, only the CIA used the polygraph to conduct background investigations.

In the 4 month period covered by the study, the CIA conducted 507 background investigations. Of these, adverse information arose concerning 47 percent of applicants or other individuals being investigated for security clearances. Thirty-five (83 percent) of the adverse cases were resolved against the individual (i.e., the applicant was not hired or clearance was not granted). In two-thirds of the instances of adverse information resolved against the individual with the use of the polygraph, subjects admitted to the adverse information. The kinds of issues admitted by subjects had primarily to do with drug and alcohol use (e.g., marijuana use, alcohol abuse, abuse of other drugs; approximately 55 percent of the cases) and immoral conduct (e.g., sexual deviance; 24 percent of cases). Four cases involved irresponsibility, a subcategory of which is violation of security regulations, and none involved the loyalty category. It is not clear whether any of the four irresponsibility cases involved violations of security regulations. Three of the 84 resolved against cases involved admissions of foreign connections, meaning in this case either that: 1) the subject was not a U.S. citizen; 2) the subject's spouse was not a citizen; 3) relatives potential "hostage"; 4) alien relatives, "hostage" unlikely; or 5) life abroad cannot be verified. The seriousness of the wrongdoings was not clear.

The crux of the DCI analysis was the construction of a productivity index for investigative techniques from the CIA data and data from other agencies. Based on the fact that a large number of cases were resolved against individuals by admission, and the polygraph was the "unique source" (DCI, 1980, p. 133) in all the CIA cases resolved against the subject, DCI tentatively concluded that the polgyraph was the most productive of all background investigation techniques. For admissions, for example, the polygraph had an index of 6.59 compared to 0.79 for "administrative screening," 1.08 for "investigative interviews," and 0.28 for "papers only."

Several aspects of the study should be noted. One is that the criteria for case selection and adverse information are not stated. Another issue, noted by the DCI study authors, is that even though the polygraph is reported as the sole source in resolving adverse information, it was only used after a thorough investigation using other sources had taken place. For this reason, it is difficult to assess its effectiveness separately from the effect of a thorough investigation. Furthermore, as a result of being conducted at the end of a background investigation, in this case the polygraph examinations could be considered a confrontation technique rather than an investigative tool, according to DCI. Agencies surveyed by DCI were asked not to include confrontation techniques in their responses. A third problem is that there was no independent verification of the cases that were resolved. Perhaps, most important, is that the effectiveness of polygraph examination cases involving most, if not all (i.e., irresponsibility) of the kinds of adverse information uncovered among applicants in the sample can probably not be generalized to investigations of unauthorized disclosures.

Edel and Jacoby (1975), in a study reported in a leading psychology journal, tested the reliability of polygrapher judgments of physiological

responsivity in applicants for positions with "a large Government agency." Forty cases were randomly selected from the agency's applicants in 1966. Ten practicing polygraph examiners acted as actual examiners in four cases each and raters in eight additional cases. In each case, examiners (raters) judged three physiological responses to each interview question as either "no specific reaction" or "a specific physiological reaction." The rate of agreement between examiners and raters as to whether a physiological reaction took place averaged 96 percent.

Of course, as the authors note, demonstrating consistency among examiners "is not equivalent to demonstrating consistency in interpretations based on these physiological reactions" (Edel & Jacoby, 1975, p. 634). For example, responses were not differentiated for relevant versus irrelevant questions. Therefore, although Edel and Jacoby's study indicates that the examiners in the Government agency can reliably detect physiological reactions, whether these physiological reactions indicate deception among applicants for positions in Government agencies has not been tested. Because of the potential adverse consequences for employment applicants (particularly in Government agencies where there is interagency checking (see, e.g., DCI, 1980) such tests have substantial practical significance.

Conclusions

Although there is some evidence from available field studies that polygraph testing is effective in detecting deception by guilty criminal suspects, there is also what in some cases might be regarded as a substantial error rate. This is particularly so for innocent subjects. There appears, as yet, to be no scientific field evidence that polygraph examinations can be effectively used to investigate unauthorized disclosures or that they represent a valid test to prescreen or randomly screen Government employees. Results of field studies are subject to additional problems of interpretation because of inadequate measures of ground truth.

The following section reports on the effectiveness of polygraph testing demonstrated by analog studies. As will be shown the construct and criterion components of validity are stronger in analog studies, but because of problems with external validity, they do not provide evidence about actual polygraph testing that is as direct as that form field studies. Nevertheless, reviewing such evidence is necessary to assess both the present and potential uses of polygraph testing.

Table 4.1
Characteristics of Field Studies

Type of Validity Affected

Study	Criterion	Basis of Examiner Decisions	Types of Cases
Bersh	Panel of legal professionals' assessment of investigative files	Original examiners' decisions	criminal investigations/military personnel
Barland & Raskin	" "	" "	sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes[8]
Barland & Raskin	Panel	Blind Evaluations**	" " "
Raskin	Confession	" "	" " "
Horvath & Reid	Confession	Blind Evaluation	theft, sexual misconduct, sabotage, bribery, criminal damage to property
Hunter & Ash	" "	" "	theft, official misconduct, brutality, sexual assaults, homicide
Slowik & Buckley	" "	" "	theft, industrial sabotage, drug abuse, rape
Wicklander & Hunter	" "	" "***	homicide, sexual assault, theft, official misconduct
Horvath	" "	" "	crimes against persons crimes against property
Davidson	" "	" "	crimes against property/military personnel
Kleinmuntz & Szucko	" "	" "	theft

Notes:

* Only 77 of 92 cases were analyzed as to type of crime.
** Not included in the analysis for reasons discussed in the text.
*** Wicklander & Hunter also included an evaluation in which evaluators were given additional case material.
\# Research design issues are discussed in the text.
\## All studies use some version of CQT.

* * * * * *

Table 4.2
Mean Detection Rates as a
Percentage of Total in Field Studies

"Ground Truth" per Criterion

Examiners or Evaluators' Diagnosis	Percent Guilty		Percent Innocent		
	Mean	S.D.	Mean	S.D.	
Deceptive	49.3	(12.7)	8.2	(7.2)	57.5
Nondeceptive	5.8	(5.1)	32.7	(16.7)	38.5
Inconclusive	2.0	(3.0)	2.1	(2.5)	4.0
	57.1		43.0		100%

lambda b = .65

* * * * * *

Table 4.3

OUTCOMES OF FIELD STUDIES OF VALIDITY

	GUILTY				INNOCENT				
tal N	N	Correct	Non Correct (False Neg.)	Inconclusive	N	Correct	Non Correct (False Pos.)	Inconclusive	Total N
rsh 1969									
anel of 4									
QT Unan.	32	96.9%	3.1%	0[1]	36	88.9%	11.1%	0[1]%	68
C Unan.	38	89.5	10.5	0	51	94.1%	5.9	0	89
g. Unan.		93.2	6.8	0		91.5	8.5	0	
jority									
OC & GQT)	34	70.6	29.4	0	25	80.0	20.0	0	59
		81.9	18.1	0		85.6	14.3	0	216
rvath & id, 1971 examiner, examiner/ aluators)	20	85.0	15.0	0[8]	20	90.5	9.5	0[8]	40
nter & Ash 73[2]--(1 aminer, 7 ex- iners/eval- tors)	10	87.1	11.4	1.4	10	86.4	14.1	0	20
owick & ckley, 75[3]--(ran- n selection; examiner/ aluators)	15	84.0	15.3	.7	15	90.7	6.6	2.7	30
cklander & ıter--(2 aminers/6 aluators)									
PG 4 20	10	98.6	1.3	0	10	86.6	8.3	5.0	20
PG Avg.		90.0	8.3	1.6		86.6	5.0	8.3	
		94.4	5.0	1.0		86.1	6.9	6.9	

Table 4.3 (Continued)

Horvath 1977[5] (10 examiner/ evaluators)									
Verified cases	28	77.1%	22.9%	0[10]%	28	51.1%	48.9%	0[10]%	56
Davidson 1979[6]--(random selection 7 examiners/ evaluators)	10	90.0	10.0	0	11	91.0	0	9.	21
Raskin[3,11] (1 examiner, 25 evaluators)	12			4				16	
Numerical		91.7	0	8.3	75.0		0	25.0	
Non-numerical		83.3	8.3	8.3		25.0	50.0	25.0	
Barland & Raskin, 1976 (1 examiner, panel of 5)									
Panel[7]	47	91.5	0	8.5	17	29.4	52.9	17.6	64
Judicial outcome	33	90.9	0	9.1	8	12.5	75.0	12.5	41
Kleinmuntz & Szucko 1983--(6 examiners/ evaluators)[12]	50	75.0	25.0	0[9]	50	63.0	37.0	0[9]	100[8]

Notes:

1. Data for inconclusives not reported; total inconclusives appear to total 27 (243 inital N--216 decisions reported).

2. Average of 2 blind chart analyses spaced at least 3 months apart, done by same examiners.

3. Average frequencies divided by number of examiners.

4. PG indicates evaluators had access to written information in addition to polygraph charts (e.g., case details, subject behavior during examinations, etc.) Both PE only and PG examinations were done by the same examinations two months apart.

5. Excludes Horvath's analysis of 28 unverified cases; because there is no criterion, this is a reliability study.

6. Majority decision only.

Excludes 28 cases for which the panel was unable to come to a decision as to guilt or innocence.

Examiner/evaluators were not allowed to judge as inconclusive as to overall deceptiveness. In another type of analysis on a question by question basis, judgments of doubtful or inconclusive were allowed.

Inconclusiveness not allowed.

There were apparently 15 (1.3 percent) inconclusive judgments out of 1120 total judgments (10 examiners X 112 cases) which the author excluded from further analyses.

Seven examiners used numerical scoring; 18 used non-numerical scoring procedures. It is clear, however, from a comparison of the Bersh and Barland and Raskin panel studies that selection and completeness of both polygraph records and case files may play an important part in research results, especially with respect to false positives.

Decisions were based on one polygraph chart: standard practice generally employs at least three. Also, the evaluations were made by students with little ploygraph experience.

REVIEW AND ANALYSIS OF POLGYRAPH ANALOG STUDIES

Introduction

Analog studies, for purposes of the present analysis, are investigations in which field methods of polgyraph examinations are used in simulated criminal or other situations. Such studies investigate either "mock" crimes set up by an experimenter (with the knowledge and collaboration of subjects) or actual small crimes "induced" by the experimenter. Such analog studies are not actual criminal investigations and subjects are usually aware that they are participants in polygraph research. Analog studies differ from other laboratory studies of polgyraph testing in that they simulate actual field examinations. Typical components of field examinations are replicated. Such studies test the validity of various polygraph techniqeus under controlled conditions. In Section 4, the results of a systematic review of field studies of validity were presented. As with the field studies, the studies concern the use of polgyraph examinations for investigation of crimes. The two exceptions (Barland, 1981; Correa & Adams, 1980) use analogs to the type of relevant/irrelevant (R/I) question technique typically used in the personnel screening situation.

The present section is organized as follows: first, the characteristics of analog studies and the varieties of ways in which they differ from field studies are discussed. Then, the criteria used for including studies in the analysis are described. The coding procedure, which is essentially the same as that used to code the field studies, is described briefly. Analog studies of the control question technique (CQT), general question technique (GKT), and personnel screening examination are then reviewed. The findings of a statistical analysis of the analog studies and a discussion of their findings complete the section.

Characteristics of Analog Studies

The "crimes" utilized in analog studies in order to establish ground truth have taken different forms. For the most part, they are "mock crimes"; that is, crimes in which subjects know they are role "playing" at being criminals for purposes of an experiment. Mock crime studies may be further differentiated by whether or not the experimenter controls the guilt or innocence of research participants. In some studies, subjects know that the crime is part of the experimental situation but they are more or less free to go through with the crime or not. Two analog studies have utilized actual small crimes. In these studies, apparently real situations were embedded in an experimental situation in which subjects were given an opportunity to commit a crime or not.

The consequences of failing a polygraph examination (e.g., a possible prison sentence) cannot be replicated in the laboratory. In analog studies, punishment takes such forms as losing the chance for a monetary reward. Some researchers have experimented with other punishments such as electrical shock (Lykken, 1959) or the threat of shocks (Bradley & Janisse, 1981). The analog studies that use real crimes provide another alternative, in that subjects can be threatened with real punishment (e.g., academic sanctions for cheating on an examination). In still other cases, subjects are led to believe that "stable" individuals can avoid detection.

Analog studies represent, thus, a "trade off" to the investigator interested in polygraph testing validity. On the one hand, because the researcher sets up the crime, ground truth is known; and because "ground truth" is established, analog studies are superior to field studies in terms of criterion validity. Furthermore, they provide the investigator with more control of the polygraph situation and conditions of testing. The experimenter can select particular subject groups, can standardize testing procedure for all subjects, and can systematically vary guilt or innocence. With this control, the experimenter can also directly compare the effects of variations in polygraph techniques, physiological measures, information given to subjects, and scoring methods.

On the other hand, although analog studies have greater criterion validity and offer greater experimental control, their use as indicators of polygraph testing validity is potentially problematic. The reasons have to do primarily with external validity (Raskin, 1982; Barland & Raskin, 1973; see, also, Abrams, 1973; Ansley, 1983a; Horvath, 1983; Lykken, 1981). That is, the crime situation differs, the testing situations in the field and the laboratory differ, the training of the examiners differs, the subject population differs, and, apparently most important, the consequences for "suspects" differ dramatically between the field and the laboratory.

Numerous specific differences can be noted. Perhaps most importantly, the laboratory crime and the consequences of detection are much less serious. In addition, in an analog study, demand characteristics (which suggest to the subject desirable responses) may create a somewhat different polygraph situation than found in typical field situations (Barland & Raskin, 1973). In terms of factors that may increase validity of analog studies, there is some evidence that laboratory researchers are, in general, able to use more sophisticated and stable equipment than portable machines often used in the field (Raskin, 1982). On the other hand, examinations in analog studies are often conducted by researchers who are primarily psychophysiologists (e.g., Dawson, 1980) or psychologists (Correa & Adams, 1981) with only limited training in field techniques. Field examinations, in contrast, are conducted by individuals whose primary training is as polygraph examiners and who are usually experienced. This would suggest that field examinations may be more accurate.

The characteristics of subjects who participate in analog studies also vary from subjects in field studies. Several use college students, others recruit community members through the newspaper, one uses police candidates, and another prison inmates. In many studies, subjects are probably better educated and more highly socialized than the average field examinee. In the case of student subjects, they are probably younger on the average and from a higher social class as well. Raskin (1983) notes that analog studies using students yield a lower accuracy rate than other studies. As will be discussed below, this may be due to subject differences between field and analog studies because a realistic fear of failure does not play a central role for subjects. The consequences of failure for analog studies are usually minimal in contrast to typical field investigations.

Study Selection

For present purposes, studies were only included as analog for the primary analyses if they employed actual field polygraph techniques to

detect deception or concealed information, and if the studies pertained to some use of polygraph testing in the real world. The studies selected are listed in Tables 5.1 and 5.2. Studies of components of the polygraph examinations, such as studies which used only card tests (Kugelmass et al., 1968; Lahri & Ganguly, 1978), number tests (Moroney & Zenhausern, 1972) or tests concerning concealed personal information (e.g., parents' first name; see, e.g., Lykken, 1960) were not included.

In addition, studies were excluded because their primary focus was on a theoretical factor thought to affect validity, such as variability in physiological recordings (Cutrow et al., 1972; Timm, 1982), or the role of "lying" (Kugelmass et al., 1967). Such studies will be referred to as laboratory investigations and are distinguished from analog studies.

Analog studies of the guilty knowledge test (GKT) have been included, although analyzed separately, because this form of the polygraph examination represents a serious alternative proposed for use in the field (Kleinmuntz & Szucko, 1982; Lykken, 1974, 1981), even though it has not been put into general practice.

Description of Studies

This section discusses each of the analog studies which have been organized into three categories according to questioning technique. The discussion of analog studies of the control question technique (CQT) begins this section. Studies of the CQT represent available studies, much like the case for field investigations (see section 4). Six studies of the concealed information or guilty knowledge technique (GKT) and of the relevant/irrelevant technique (R/I) follow. Only in one study (Barland, 1981), involving the R/I technique, were subjects Government employees. The results of individual studies are summarized in Table 5.1 (CQT and 5.2 (GKT). The description of the studies are followed by a systematic statistical analysis of the results of the CQT and GKT studies. The relevant/irrelevant studies were not analyzed as a group because of the paucity of the studies.

Essentially, as shown in Tables 5.1 to 5.4 the analysis of the analog studies yields similar conclusions as that of the field study analysis. This is, although there is a greater than chance probability of detecting deceptive and nondeceptive subjects, there is significant error rate, and a great deal of variation across studies. However, as has been found in some reviews (Abrams, 1973; Ansley, 1983a), analog studies of the CQT had lower accuracy rates than field studies of the CQT.

In the studies detailed below, the experiments also tested the effect of factors hypothesized to have an effect on validity. For example, Barland and Raskin (1976) examined the effect of validity of different types of feedback about the polygraph, and Dawson (1980) investigated the effects of countermeasures. These factors are examined more systematically in section 6; the emphasis of the present section is on the validity of different forms of polygraph examinations.

Control Question Technique

Fourteen analog studies of the control question technique were located. The largest group of these studies (six) emanate from the research

program of Professor Raskin at the University of Utah. The remaining eight studies were conducted at a number of settings in the United States and elsewhere. Raskin and colleagues have conducted a systematic analog research program, and these studies are described as a group helm. Other researchers have published individual studies testing specific hypotheses relevant to the validity of the polygraph. A description of these studies follows discussion of the University of Utah studies.

University of Utah Studies

Despite longstanding controversy about polygraph validity, the first research project conducting an analog study that simulated field polygraph techniques was not conducted until the 1970's (Raskin, 1982). It was then that an ongoing research program headed by Professor Raskin at the University of Utah began to study the validity of the polygraph through analog experiments. In addition, these studies also examined the relationship to validity of different polygraph techniques (e.g., the stimulation test), different physiological measures, different methods of assessing the results, different types of information provided to subjects and different subject and situation factors that could potentially affect polygraph validity.

The experiments conducted by Raskin and colleagues use similar procedures to set up the mock crime and to conduct polygraph testing. In each of their studies, subjects are randomly assigned to an "innocent" condition or to a "guilty" condition. the mock crime is the theft of a small amount of money or a valuable object from a desk in a nearby room. To increase their motivation, subjects are offered a financial bonus for convincing the polygraph they are innocent. In the testing the examiner employs the Federal ZOC control question technique, including a pretest interview. A numerical field scoring method developed by the Utah group (Barland & Raskin, 1975) is used to make the diagnosis of truthfulness or deception.

Barland and Raskin

In the initial analog study using the CQT (Barland & Raskin, 1975), 72 student "guilty" and "innocent" volunteers were randomly assigned to one of the three "feedback" conditions. The positive feedback subjects were instructed that the polygraph was effective, the negative feedback students were told that the machine was not working properly, and the other students received no feedback. Subjects then underwent a complete polygraph examination including a pretest interview. The Federal version of the zone of comparison (ZOC) technique was employed, with standard control questions used for all subjects. The CQT identified 53 percent of all subjects correctly. Twelve percent were identified incorrectly and 35 percent of the examinations were inconclusive. Of the errors, three (4 percent) were false negatives and six (8 percent) were false positives.

Podlesny and Raskin

Podlesny and Raskin (1978) conducted a more extensive experiment to examine the accuracy of the CQT using three different types of control questions. They also tested the accuracy of behavioral observations of the examinee (Horvath, 1973; Reid & Inbau, 1977) in detecting deception, since this type of information is used in many field examinations and could

possibly affect the validity of the technique (Lykken, 1974, 1981). They compared as well the capability of different physiological measures in differentiating between guilty and innocent subjects. A GKT was also conducted with 20 subjects (see below).

In Podlesny and Raskin's study, subjects were community members who responded to newspaper advertisements. The experimenters drew from the Reid method in their design of the pretest interview (see section 2). One experimenter asked the subjects three questions from Reid and Horvath's structured pretest interview designed to provoke the subjects into displaying "behavioral symptoms" of deception (Horvath, 1973; Reid & Inbau, 1977).

During the polygraph examination they included two special types of control questions among the set of questions asked of the subjects. One was a "guilt complex question," which asked the subject if he committed a fictitious crime of the same nature as the real crime. In this study, the guilt complex question was, "Did you take that watch from room 702?" (Podlesny & Raskin, 1978, p. 348). There was, of course, no watch stolen from room 702. The experimenters also varied the wording on some of the control questions, so that half the subjects received "nonexclusive" and half "exclusive" control questions.

In the pretest interview, the examiners followed the usual field procedure of reviewing the control questions with the subjects, and the questions were adjusted until they elicited a "no" response. The control question polygraph test then took place, with three or more charts obtained from each subject, although only the first three were used in the objective scoring. Immediately after testing, the first three charts obtained were scored blind on EDR, respiration and cardio measures. Later, an independent rater scored the tests, using the numerical scoring system devised by Barland and Raskin (1975). The experimenters also used objective measurements of all physiological response measures with the aid of computers and persons who had no knowledge of the field evaluations or treatments administered. The experimenters used the decisions made by the independent blind evaluator to assess the validity of the polygraph test. This was, however, equivalent to using the polygraph examiner's decision, because the independent rater and the examiner agreed on 100 percent of their decisions.

The results for both types of control questions combined (with an inconclusive zone used) were 80 percent correct, 10 percent incorrect and 10 percent inconclusive. There were three false negatives (8 percent) and one false positive (2 percent). The accuracy of the CQT differed depending on whether exclusive or inclusive control questions were used. When the exclusive control questions were used, 85 percent of the subjects were identified correctly, 5 percent incorrectly and 10 percent inconclusive. Of the assessments of the 20 subjects in this condition, one (5 percent) was a false negative and there were no false positives. When nonexclusive control questions were used, 75 percent were correct, 15 percent incorrect and 10 percent inconclusive. Of these 20, two (10 percent) were false negatives and one (5 percent) was a false positive. The evaluative scores for each physiological component were analyzed to determine if the scores differed between guilty and innocent subjects. Only the EDR and plethysmograph scores yielded significant differences.

Behavioral observations, by themselves, yielded a significant number of correct decisions, but this differed greatly between innocent and guilty subjects. Of the guilty subjects, 86 percent of the decisions made were correct (25 deceptive, 4 nondeceptive and 1 inconclusive); however, only 48 percent of the innocent subjects were correctly identified (12 deceptive, 11 nondeceptive, 7 inconclusive). An objective quantitative analysis for each physiological measure was employed to determine if each was effective in discriminating between guilty and innocent. Most of the measures yielded significant discriminations, with the exception of a few of the cardiovascular measures.

Raskin and Hare

A special population, prisoners, especially relevant to the field use of the polygraph, was the focus of a study by Raskin and Hare (1978). In their sample of 48 inmates of a Canadian prison, half were selected for high levels of psychopathy, and half for low levels. One purpose of their study was to investigate whether deceptive psychopaths could more easily escape detection than normal subjects (see section 6). Overall, assessments of deception from the field evaluations for all charts were 88 percent correct, 4 percent wrong, and 8 percent inconclusive. There were only two errors, both false positives. No significant differences were found between psychopaths and nonpsychopaths, suggesting that a CQT polygraph examination is equally valid for both. Also, a quanititative analysis showed that all the physiological measures were significantly different between guilty and innocent subjects. Psychopathy did not obscure these differences and in some cases enhanced them.

Rovner, Raskin and Kircher

Rovner, Raskin and Kircher (1979) studied the effect of information and practice on the accuracy of polygraph examinations. Seventy-two subjects recruited from the community took part in this mock crime experiment. One third of the subjects (12 innocent and 12 guilty) were given in-depth information about the polygraph and about countermeasures used to appear innocent (information condition). Another third received this information and underwent two practice polyraph examinations about which they received "feedback" (information and practice condition). The other third had no such intervention (standard). A blind field evaluation performed some time later produced the scores for decisions of guilty or innocence, and for analysis of the physiological responses. Accuracy for the Standard group and the Info group was identical: 88 percent correct, 4 percent incorrect, and 8 percent inconclusive. But accuracy for the information and practice condition was lower; 62.5 percent correct, 25 percent incorrect, and 12.5 percent inconclusive. There was one error in the standard group and one in the Information group--both false positives. The six errors in the Information and Practice conditions were three false positives and three false negatives.

Kircher

Some of the latest work of the Utah laboratory explores the use of computers in the analysis of polygraph recordings. Kircher (1983) compared the accuracy of a computer decisionmaking process to the accuracy of assessments of a field examiner. The computerized analysis cannot be included

in the statistical analysis of this report, because it is not presently a field scoring method, but the decisions of an independent evaluator who was used can be. This mock crime study followed the basic procedures of Podlesny and Raskin (1978) with 100 subjects from the community. The accuracy of the original examiner was not reported though the results of an independent evaluator were. The independent evaluator, who numerically scored the charts blindly, correctly diagnosed 87 percent of the subjects; misdiagnosed 6 percent; and made a judgment of inconclusive on 7 percent. The six errors were evenly divided between three false negatives and three false positives. In comparison, different computer decision models, on the average, correctly identified 84.9 percent of subjects, misidentified 7.85 percent, and placed 7.2 percent in an inconclusive category.

Other Studies

A range of other studies have been conducted in recent years to evaluate aspects of polygraph test validity. Such studies usually manipulate one or two variables that are hypothesized to be important determinants of polygraph validity. For the most part, these experiments use procedures that are similar to Raskin's mock crime paradigm. Some of the discussion of the procedures in each study is omitted, because they closely follow this paradigm.

Dawson

Dawson (1980), for example, focused on the effect of "cognitive countermeasures" on validity. His study was unique in that the subjects were actors trained in the Stanislawsky method of acting, which teaches actors to use their own experience to create emotional states appropriate for a role. Studying the attempts of "method" actors to foil the polygraph may help determine whether guilty subjects can be trained to use cognitive countermeasures to appear innocent (see section 6). Dawson was also interested in analyzing separately responses during two distinct phases of the questioning: while subjects listened to questions and while they responded.

Dawson's sample consisted of 24 student actors, half of whom were randomly assigned to the "guilty" group and half to the "innocent" group. They were instructed to use the Stanislawsky method to appear innocent on the polygraph examination. After the mock crime, four charts were obtained from ZOC control question test about the crime. On two of the charts, the subjects were instructed not to respond until they received a signal 8 seconds after a question. This served to separate responding associated with the questions from responding associated with answering. Numerical scoring based on Barland and Raskin's (1975) system was done separately on three different types of physiological responses: a) responses when the answers were immediate; b) responses during the questions when the answers were delayed; and c) responses during the answers when the answers were delayed.

Dawson found that the subjects' immediate physiological responses to the questions, whether they were answering immediately or not, led to decisions which were 88 percent correct, 8 percent incorrect, and 4 percent inconclusive (frequencies across two conditions were summed). The delayed answer response yielded a rate of 29 percent correct, 8 percent incorrect,

and 62 percent inconclusive. The incorrect decisions made were entirely false positives. A quantitative analysis revealed that the EDR and cardiovascular measures differentiated significantly between innocent and guilty, but respiration did not. The major outcomes of this study suggested that the polygraph was not susceptible to cognitive countermeasures of the sort used by the actors and that scorable responses generally occur immediately after questions.

This experiment does not, however, test cognitive countermeasures in a situation in which the subjects know the essentials of the CQT and apply cognitive countermeasures differentially to relevant and control questions. The average criminal subject is likely to attempt cognitive measures naively, but a sophisticated subject--perhaps more likely to appear in a national security investigation--may learn cognitive countermeasures along with the knowledge of the control question technique.

Widacki and Horvath

Widacki and Horvath (1978) designed an experiment to examine the polygraph's efficacy in comparison to other techniques in the mock investigation of a mock crime. They recruited 80 Polish student volunteers and had all of them provide writing specimens, photographs of themselves and fingerprints. Subjects were then assigned to 20 groups of four subjects each. Within each group, one subject was randomly assigned to be the perpetrator and the other three were innocent suspects. Each group was thus an "investigative case." Because of this feature of the design, the decisions of guilty and innocent were not independent. Therefore, Widacki and Horvath's findings could not be included in the statistical analysis of the control question analogues and must be considered separately. A similar situation holds for Kubis's (1962) mock crime experiment (see below).

The mock crime proceeded as follows: the guilty subject picked up a parcel from one of two persons acting as a "doorkeeper" of a building in the area. The perpetrator gave some experiment-related papers to the doorkeeper and then signed for the parcel. Thus, an eyewitness account (by the doorkeeper), fingerprints and handwriting specimens were all available. Blind polygraph examinations then were conducted using the Reid control question method (including the examiners' behavioral observations of the subject). Analysis of the three other sources of evidence was carried out.

Widacki and Horvath found that the polygraph produced the most correct decisions (n=18), the fewest (along with handwriting) incorrect decisions (n=1), and the fewest inconclusive decisions (n=1)., Widacki and Horvath note, however, that a direct comparison of these four investigative methods may be invalid because the experimental procedures could not insure a comparable level of quality of evidence for each method (e.g., fingerprints were not detectable in the majority of cases).

Because of its experimental design that had the examiner make decisions on four suspects as a group, the study produces data about the accuracy of the polygrpah that is difficult to interpret. But it does shed light on the efficacy of the polygraph relative to other investigatory techniques that might be the alternative. Certainly, it is crucial in policymaking to judge the validity of the polygraph relative to other techniques that would be used in its stead. More research is needed in which the

polygraph is compared to other investigatory techniques, and the quality of information across techniques is held constant. Such a comparative analogue study would be especially valuable if it included different techniques used in investigations of Federal personnel, such as those reported in the DCI survey mentioned above ("administrative screening," "investigative interviews," etc.).

Bradley and Janisse (1981) studied the effects of two other variables hypothesized to influence the validity of the polygraph: the degree of threat involved in the punishment being judged guilty, and successful demonstration to the subjects of the technique's accuracy. A mock crime was carried out using procedures similar to those used by Barland and Raskin (1975). subjects were also given a series of stimulation tests. Results of these tests were manipulated such that they made the polygraph test appear perfectly effective, partially effective, and ineffective. In addition, half the subjects were told they would receive a painful electric shock if found guilty, though no shock was ever given.

The degree of manipulated effectivenes had no direct effect on scores, but did tend to increase the accuracy of detection. Threat of punishment did not affect accuracy of detection, although it did have an overall effect on heart rate. EDR and heart rate change were significantly accurate in differentiating guilty and innocent, although another measure, pupil size change, was not.

Honts and Hodes

Two recent analog studies of the Backster ZOC method of testing (Honts & Hodes, 1982a, b) were conducted primarily for the purpose of testing whether polygraph examiners could detect the use of physical countermeasures by subjects. In the first study, subjects were college students who received extra credit toward their final grade for their participation. "Guilty" suspects participated in a mock crime (theft of an examination); innocent suspects were only told of the theft. All subjects were motivated to produce truthful outcomes on the polygraph test by an offer of twice the number of credits if the examiner reported them as truthful.

In addition to participation in the mock crime 24 of the guilty subjects participated in 15 minute training sessions in which they were told about the theory of the CQT and shown how to use either tongue biting (12 subjects) or toe pressing (12 subjects) as countermeasures during presentation of the control questions. They were also instructed to try to relax as much as possible during presentation of the relevant questions.

The actual polygraph examinations took place a week after the theft and training sessions. All guilty subjects were instructed to have the "stolen" examinations with them, presumably to enhance subject involvement. Four charts were obtained from each subject using a standard Backster examination administered by an experienced polygraph examiner. The examiner was aware of the details of the experiment, including a knowledge of the base rates of guilt and the countermeasures that would be attempted, but was blind to the group assignment of individual subjects. At the end of each examination, the examiner made a yes/no decision regarding the subject's use of countermeasures. After all subjects had been tested, the original examiner made a decision as to deception by blindly evaluating the

charts using the Backster numerical scoring technique, and made another decision about the use of countermeasures based on inspection of the charts. Charts were also examined and scored by a second examiner who was blind to all aspects of the experiment.

As shown in Table 5.1, while there was a low rate of false negatives (5.5 percent), examiners were not able to make a decision on one-third of countermeasure and no countermeasure guilty subjects, and half of the innocent subjects. There was a large (16 percent) false positive rate. Examiners were not able to detect the use of countermeasures.

In their second experiment on countermeasures, Honts and Hodes used approximately the same procedures and subject pool, with the exception that subjects were asked to employ both countermeasures simultaneously, were given 30 minutes of training, including a practice session, and were asked to practice at home. A cardio cuff was added to the polygraph instrument, and a card test was conducted prior to the administration of the first test.

Overall, results of the second study replicated the first. The categorizations of the original examiner were 51 percent correct, 14 percent incorrect, and 35 percent inconclusive. Twenty-six percent of the countermeasure subjects compared to none of the guilty/no countermeasure subjects were incorrectly classified as truthful. Examiners were not able to detect successful countermeasure users.

Only two CQT analog studies have used real, albeit small, crimes. Another study using a real crime tested the GKT technique and is discussed in the section on that technique. An experiment by Ginton and others (Ginton et al., 1982) aimed to create even more verisimilitude than usual in the analog study. An auxiliary purpose was to test the relative effectiveness of behavioral observations, global evaluations, including behavioral observations, and numerical scoring based on the charts alone.

Ginton, et al.

Subjects in Ginton et al.'s investigation were 21 Israeli policemen. They were given paper and pencil tests that were presented as required aptitude tests. Subjects were asked to score their own tests, which provided an opportunity to cheat, that is, to revise their initial answers. The test answer sheets, however, were chemically treated so that cheating could be detected. Seven of the 21 subjects actually changed their initial answers. Later, subjects were told they were suspected of cheating, were offered an opportunity to take a polygraph examination, and were told their careers might depend on the outcome. Fifteen subjects actually underwent the polygraph testing, only two of whom had actually cheated.

A CQT was administered, and each subject was evaluated by three polygraph experts who had conducted or witnessed the particular examination being evaluated. One examiner (an observer) relied on behavioral observation, another (a rater) used only the charts, and a third (the actual examiner) used both sources of information. The evaluations were made globally. Five other polygraph examiners evaluated the charts later using both the Utah group's scoring system (Barland & Raskin, 1975) and global evaluations. The original three performed a second analysis in this way, too.

Conclusions about this study are limited because of a large no-show rate among the guilty subjects. Both guilty subjects who took the test were correctly detected. However 15 percent of the noncheaters were incorrectly identified as deceptive.

Heckel, et al.

Another analog study (Heckel, Brokaw, Salzberg & Wiggins, 1962) used a sort of staged actual crime to investigate the differential accuracy of the CQT with psychotic, neurotic and normal subjects. Fifteen subjects (five from each of the above three groups) were given the opportunity to steal money from the wallet of an experimenter who was staging a session of psychological testing. The experimenter later alleged that $20 had been stolen, and arranged for polygraph examinations of the 15 subjects by a field examiner. No money had actually been stolen, so the subjects were actually innocent. Four polygraph experts later rated the charts. Averaging the results for these independent evaluators, 11 of the subjects were correctly labeled innocent, 1 was called guilty and 3 were placed in an inconclusive category. The one error and one inconclusive were with psychotic subjects, and the other two inconclusives were with neurotic subjects. Since psychotic subjects would generally be excluded from field polygraph examinations because it is thought their disorders would interfere with accuracy of results, only the 10 normal and neurotic subjects are included in the statistical analysis in this report.

Szucko and Kleinmuntz

A somewhat different approach to assessing the validity of the polygraph was taken by Szucko and Kleinmuntz (1981). They compared directly the ability of polygraph examiners to assess deception against the ability of computers to do the same using a digitalized form of the same data. They had a sample of 30 psychology undergraduate volunteers and randomly assigned them to the guilty or innocent conditions. The mock crime involved the "theft" of a $5 bill. Polygraph tests were administered by four examiner-trainees from a polygraph firm near Szucko and Kleinmuntz's university. The recordings of the physiological measure were transformed into digital form for computer analysis.

Six experienced polygraph examiners independently evaluated the charts. No inconclusive category was allowed in the study. Digital polygraph data was evaluated by computer. A lens model equation drawn from studies of human judgment was used. The results of this analysis indicated that five of the six polygraph raters were able to detect deception significantly better than chance, but four of them also had fairly high rates of false positives. Szucko and Kleinmuntz estimate that the judges detected on the average 71 percent of guilty subjects, but also called 49 percent of the innocent subjects deceptive (false positive). Szucko and Kleinmuntz state that 80 percent of the protocols could be classified correctly using a purely statistical analysis, but they do not state the detection rate, false positive rate and false negative rate of their statistical analysis.

Kircher and Raskin (1983) contend on the other hand that evaluators using numerical evaluations can be "at least as accurate as those produced by any known statistical decision model and that the accuracies of both clinical and statistical methods exceed 90 percent." Kircher and Raskin

reanalyzed charts from the Rovner et al. (1979) study described above and used as a lens model, similar to that employed by Szucko and Kleinmuntz. The studies, however, differed in a number of ways, which could account for the variation in their results. Probably the most important difference is that Kircher and Raskin's interpreters were trained in numerical scoring procedures (Raskin, 1982), whereas interpreters in the Szucko and Kleinmuntz study used global evaluation procedures (Reid & Inbau, 1977).

Concealed Information Tests

Although the largest number of analogue studies investigate the CQT, several analog studies have examined the validity of the guilty knowledge test, one type of concealed information test. A search of the literature revealed no analog studies of the peak of tension test as a distinct technique.

Lykken

In one early invstigation of the GKT, Lykken (1959) attempted to demonstrate that the detection of incriminating knowledge about a crime can be done more accurately than the detection of a lie about the crime. In Lykken's study, 49 male college students were randomly assigned to 4 categories of guilty in conducting 2 mock crimes. Subjects either committed a staged "theft," a staged "murder," both or neither. An experimenter then conducted two GKT polygraph examinations with each subject, one for each crime.

Each test in Lykken's (1959) consisted of six questions about details related to the "murder" situation and "theft" situation (e.g., asking the subject to identify an object present in the "murder" room). To make subjects anxious about the accuracy of their responses during the examination, they were told they would be given an electric shock if the examiner felt their responses indicated guilt; in fact, subjects received an electric shock after every question. The relevant alternative in each question was randomly varied among an average of five possibilities. If the question about the relevant detail produced the EDR with the greatest amplitude, it recieved a "2." If it was the second largest in amplitude, it received a "1." A percent guilty score on each test was "12", and a perfect innocent score was "0". A score of seven or greater was categoried as guilty for the purpose of analysis, and a score of six or less was categorized as innocent. The guilty knowledge test was accurate to a signficant degree in identifying subjects who committed both, either or neither of the crimes. On the basis of this experiment, Lykken argued that the GKT, with some refinements, could be applicable in criminal investigations.

Davidson

Other researchers have used Lykken's GKT paradigm to explore further its utility as a polygraph examination technique. Davidson (1968) examined the GKT's validity under conditions that varied motivation level and that he claimed were, in general, more "ego-involving" for subjects. In Dawson's study, 48 college students were recruited and assigned randomly to 12 groups of 4. Three of the four were instructed to attempt to commit specific mock murders, and the fourth served as a control. The mock crimes were arranged such that one subject would "commit" the crime, one would try to

fail, one was motivated but never had the opportunity and one (the control) had no knowledge of the crime. Half of the subjects who "committed" the murders received a large amount of money ($25 to $50) and half received a small sum ($10 to $1). The different amounts were presumed to create a different level of motivation in the subjects. The subjects then were examined with the use of the GKT. Six multiple choice questions with five alternatives were presented to the subjects, and the EDR was recorded. The scoring method followed Lykken's (1959) exactly (see above). Overall 98 percent of the classifications were correct against a chance level of 25 percent. The only error was one false negative.

Podlesny and Raskin

Podlesny and Raskin (1978) included GKT in their study of a variety of polygraph techniques and physiological measures. Their experiment was unique in that it employed the GKT in the same context as the CQT (see above). Thus, they were able to compare the accuracy rates of the two techniques, although they claimed that a different statistical comparison was impossible because the two techniques use very different methods to assess guilt. Podlesny and Raskin also were the first to test the GKT with physiological measures other than the EDR. To make assessments of guilt, they used the traditional polygraph respiration and cardio measures, and another vascular measure that was a composite of finger blood volume and finger blood amplitude. This latter measure was recorded by the photoplethysmograph mentioned above. In addition, Podlesny and Raskin performed a quantitative analysis of differences between guilty and innocent subjects on several other physiological measures.

The GKT was conducted after the same mock theft Podlesny and Raskin (1978) used to study the CQT. Twenty subjects (10 guilty and 10 innocent) were examined with the GKT, which included 5 questions with 6 alternatives each. The relevant alternatives were placed among the others alternatives in a "pseudo-random" order (Podlesny & Raskin, 1978, p. 349). The GKT charts were scored by the same method used by Lykken (1959) and Davidson (1968). Podlesny and Raskin also scored the charts in another way, with the addition of an inconclusive zone of scores five or six. This scoring system for assessing guilt was used with the photoplethysmograph, respiration and cardio measure as well as the EDR. Their fingings were that the GKT with the EDR was correct for 90 percent of the subjects and incorrect for 10 percent, all false negatives. Using an inconclusive zone did not add significantly to the accuracy of the technique, however: 80 percent of assessments were correct, 10 percent incorrect (all false negatives), and 10 percent inconclusive.

Giesen and Rollison

Giesen and Rollison (1980) studied the effects on the GKT of the subjects' trait anxiety levels and of the possibility that crime-related details could be relevant to innocent subjects because of associations unrelated to the crime. Trait anxiety is anxiety that is characteristic of one's personality and would be relatively stable over time. Both trait anxiety and "innocent associations" could conceivably confound the detection of guilt with the GKT.

Giesen and Rollison selected 40 female undergraduates who responded

positively to a questionnaire item on "palmar sweating." EDR is related to sweating. Thus, this sample may have tended to produce higher EDR's than the norm. This group was divided into two groups of 20: those who scored higher on a questionnaire measure of anxiety (Lykken's activity preference questionnaire) and those who scored lower. Ten subjects in each group were then assigned to the guilty knowledge condition, and to the "innocent associations" condition. The guilty subjects were told to pretend to be secret agents who had committed a murder. They read a narrative about the crime, and role-played the act of burning an incriminating picture. Innocent subjects also played secret agents, but read a narrative containing several identical details (e.g., how much money was involved), which in the guilty condition were related to the crime. They had, therefore, as much exposure to this information as the guilty subjects, but in an innocent context. Using the GKT with the EDR, experimenters asked subjects eight crime-related questions, each with five alternatives. Those details common to both conditions were used as the crime-relevant items in the GKT questions. Scoring followed Lykken's (1959) method.

Giesen and Rollison found that the GKT was highly accurate, correctly classifying all of the innocent subjects and detecting all but one of the guilty subjects (97.5 percent correct assessment). In addition, they found that the EDR measure was significantly different between guilty and innocent subjects. Trait anxiety level had no effect on EDR by itself, but the more anxious subjects in the guilty condition had significantly greater EDR than the less anxious, especially in response to the relevant items. These findings would suggest that anxiety alone does not confound GKT results, but anxiety in guilty subjects might indeed augment the accuracy of the technique. The study also suggests that the GKT may be accurate even when innocent subjects have greater associations with the crime-relevant items than with neutral items. This finding, however, must be tempered by the fact that the entire sample was selected for their tendency for palmar sweating under stress and thus, may be unrepresentative of polygraph subjects in general.

Balloun and Holmes

Balloun and Holmes (1979) used the GKT to detect guilt in a "real" crime arranged by the experimenters. They were also interested in the effect of psychopathy and of repeated examinations on the accuracy of the GKT. They selected 18 male college students with high scores on the psychopathic deviate (Pd) scale of the Minnesota Multiphasic Personality Inventory (MMPI) and 16 with low scores. The Pd scale was originally designed to make the diagnosis of psychopathic personality and was used as a scale to measure relative "amounts" of psychopathy. The experimenters acknowledge, however, that the Pd scale may be an inadequate measure of this diagnosis. These subjects took a fake intelligence test with two other students (actually confederates of the examiner). The confederates urged subjects to cheat and supplied test answers to those who were willing. Eighteen of the 34 students cheated. Later, the subjects underwent a polygraph examination using the guilty knowledge test. They were reminded that cheating on exams could lead to academic dismissal, and that the experimenters knew that some had cheated on the "intelligence test." Information from the intelligence tests that only the cheaters would know served as the incriminating details on the GKT. Another GKT with the same content, but a different order of questions was then administered to see if the subjects would adapt to the GKT and, thus, reduce its accuracy.

Balloun and Holmes scored the GKT using Lykken's (1959) method with three physiological measures (EDR, heart rate, and finger pulse volume), but only the EDR produced significant results. On the first test, guilty subjects scored significantly higher and were detected with significant accuracy. However, on the second test, though the guilty subjects had significantly greater scores, they were not great enough for significantly accurate detection of guilt at the criterion level (5.5 out of 10) used. There was no difference between the high and low Pd subjects on either administration of the GKT.

In their study of the influence of threat and demonstrations of accuracy on the polygraph examination (see above), Bradley and Janisse (1981) also tested the 192 subjects with the GKT after the CQT had been conducted. The questions concerned four relevant details. They were scored using the Lykken (1959) method. With the EDR data, the GKT classified 74 percent of subjects correctly, and 26 percent incorrectly with 11 false positives and 39 false negatives. With the measure of heart rate change, the GKT correctly categorized 63.5 percent of subjects correctly and 36.5 incorrectly, with 17 false positives and 53 false negatives. Neither the degree of threat nor the demonstrated effectiveness of the polygraph test had a significant effect on the discrimination between deceptive and truthful subjects.

Timm

Timm (1982) examined the effect of the administration of a placebo on the validity of the GKT. Also included in the experiment was an investigation of the effect on GKT accuracy of differential feedback from the stimulation test. In the experiment all 270 college student subjects committed a mock crime. There were no "innocent" subjects. Before the mock crime, subjects were either: a) given a placebo and told it would help them "beat" the test; b) given a placebo and told it would make it more difficult to deceive the examiner; or c) not given a placebo. The stimulation or number test was arranged to produce three different feedback conditions. One third of the subjects' numbers were detected, one third were not, and one third did not receive the results of the stimulation test. After the GKT was conducted on each subject, charts were scored according to the Lykken (1959) method. Adequate charts were obtained for 237 subjects. Of these subjects, 70.4% to 80.8% of them produced scores indicative of guilt, depending on how conservative a cut-off-point for the score was used. Neither the placebo condition nor the feedback condition produced a significant effect on detection ability. Because of the absence of "innocent" subjects in this study (i.e., a base rate of guilty of 100 percent), the study tells us nothing about the accuracy of the GKT with the innocent subjects. And even the results with guilty subjects are difficult to interpret when there is no comparison to results with innocent subjects. Also, without innocent subjects, a lambda is impossible to calculate.

PreEmployment Screening

Despite its widespread use in the field, there are few analogue studies of the preemployment screening polygraph examination. The two that are known to employ post-1960 polygraph screening techniques are reviewed. Correa and Adams (1981) conducted an analogue investigation of this type of examination with 40 undergraduate subjects. Barland (1981) conducted an analogue study with government personnel.

Correa and Adams

Like the usual preemployment screening test, the examination in Correa and Adams' study included a number of relevant questions. Subjects were interviewed prior to the polygraph exmaination and completed a questionnaire about their background. Half the group was instructed to lie to nine relevant questions and half to tell the truth. The polygraph test was conducted and 3 charts of 32 questions each were recorded. Most of the relevant questions concerned information from the questionnaire, but also included were three questions about events staged by the researcher in the initial interview (e.g., giving the subject a glass of water). These latter questions served as a check on the honesty of subjects in completing the questionnaire and were considered relevant questions in the evaluation of deception or nondeception. The examiner subjectively made assessments of veracity based on the polygraph recordings. When questions about the staged events and the application were diagnosed by the examiner, all 40 of the subjects were correctly identified as being deceptive or truthful.

Correa and Adams' conducted a question-by-question analysis of the charts of deceptive subjects. A mean of 75 percent of the relevant items from the screening application were correctly classified, and a mean of 25 percent were incorrectly classified. When change scores were calculated for each physiological response, all physiological measures (EDR, respiration, cardiovascular) significantly discriminated truthful from deceptive subjects. Correa and Adams suggest that these findings provide evidence for the validity of prescreening polygraph examinations. There are, however, a number of problems with the Correa and Adams study that may compromise its validity. Several features of the experiment are probably highly unrepresentative of or unrelated to field preemployment polygraph examinations: the length of the interview (96 questions); the number of deceptive responses subjects made (9); and the inclusion of questions about the experiment itself. Furthermore, the experimenters fail to discuss the criteria by which the assessments of veracity were made, so it is difficult to ascertain whether these assessments correspond to field assessments.

Barland

The Barland (1981) study is important for several reasons. One, subjects were actual military personnel who in Barland's opinion might be the types screening for counterintelligence purposes. All subjects were assigned to intelligence duties. It is, thus, unique in being the only validity study of preemployment screening in an intelligence context. However, because it did not ask any questions related to security interests, it cannot be considered a full analog to field situations.

Second, it tested the validity of a type of control question technique, the directed lie control question (DLCQ) technique, in a screening situation. The DLCQ is part of a counter-intelligence screening test developed by Army Intelligence examiners in 1971. Subjects typically answer "yes." When they answer yes, the examiner instructs them that when they are asked such questions they should respond with a "no" rather than a "yes." Thus, they are directed to lie, and their lies to these questions constitute the control questions against which responses to relevant questions are compared. The directed lie control question differs from the control question discussed previously (see section 2). With the DLCQ

technique, the control questions are not designed to provoke the subject to lie or be concerned about the telling the truth. The "lies" do not constitute deception since the examiner instructs the subject to tell lies that they both know are false. Thus, it is difficult to see how they could generate the same concern as standard control questions, since subjects are not deceiving the examiners (or worrying that they might be), and there also is, therefore, no possibility of fear of detection.

The question of whether the control question technique can be used outside of specific issue investigations (e.g., in preemployment or a periodic screening) is controversial. It is difficult to construct standard control questions when much of a person's past is irrelevant to the purpose of the examination, since past misdeeds (i.e., other than the specific issue being investigated) typically comprise the subject area of control questions.

In this 1981 study, Barland solicited volunteers from a military intelligence community. Subjects were told the purpose of the study and that testing would be limited to the subject's date of birth, place of birth, education, employment and residences (these were the relevant items), and that some subjects would be instructed to furnish the examiner with false information. Approximately half the subjects were instructed to lie to one of the relevant items; these subjects were offered a $20 reward if they could appear truthful on the polygraph examination. Unlike the data in the Correa and Adams study, the experimenter was able to check the information given by the subjects against data obtained from background investigations. The three polygraph examiners in the study had 3, 6 and 9 years of polygraph experience and had been trained at the U.S. Army Military Police School polygraph course.

Examiners used three methods of chart interpretation: zone comparison, greatest control method, and relevant-irrelevant method. As explained in section 2, in the zone method, relevant questions are evaluated against the larger of either control question response in a zone. In Barland's (1981) zone method, each physiological measure for each relevant/control question pair was rated on a point scale using interpretive criteria taught at USAMPS. In the relevant/irrelevant method, each relevant question was evaluted without making specific reference to the control question nearest it; emphasis "was placed on the size and consistency of reactions at the relevant questions" (p.8) and scored globally rather than numerically. The "greatest control" method consisted of evaluating all five relevant questions against the single control question on that chart which had the largest overall reaction. In addition to the comparisons of the three chart interpretation methods, charts were analyzed globally and on a question-by-question basis.

In the global method of analysis subjects were categorized as either deception indicated, no deception indicated, or inconclusive on the basis of appearing deceptive to any of the relevant questions. That is, if a subject was in fact deceptive to any relevant question, and he reacted deceptively to any of the questions, it was considered a hit even though the examiner may have misidentified which relevant question the subject was deceptive to. Using this method of assessing deceptiveness, the three methods of chart interpretation achieved the following results:

Zone:

62% correct identification of truthful subjects;

19% incorrect;

19% inconclusive

70% correct identification of deceptive subjects;

17% incorrect;

13% inconclusive

Greatest control:

77% correct identification of truthful subjects;

23% incorrect;

8% inconclusive

50% correct identification of deceptive subjects

23% incorrect

27% inconclusive

Relevant-Irrelevant

73% correct identification of truthful subjects

23% incorrect

4% inconclusive

80% correct identification of deceptive subjects

13% false negative

7% inconclusive

Presumably, the correct identification rates would be lower if only those cases in which the truly deceptive relevant response was counted as a "hit." To test this hypothesis, the authors conducted a question by question analysis. In this method, identification of truthful responses increased but identification of deceptive responses declined quite a bit. Using the zone technique, 77 percent of the truthful questions and only 57 percent of the deceptive question were correctly identified. With the greatest control scoring method, 85 percent of truthful responses and less than half (43) of deceptive questions were correctly identified. The relevant/irrelevant scoring technique showed the best results. With this method, 88 percent of the truthful subjects and 67 percent of deceptive questions were correct identified (although global results were better with the R/I technique). This interpretation should be modified by the fact

that each examiner used all three scoring techniques and the relevant/irrelevant technique was the last used. Thus, the interpreter had the benefit of his previous judgments. The results of a blind analysis using other interpreters were not ready to be reported by Barland at the time his 1981 report was submitted.

The results of the Barland study raise serious questions about the usefulness of directed lie control questions in screening procedures as well as, in general, the validity of polygraph testing for preemployment and counterintelligence purposes, especially if used alone. Of course, the limitations of analogue studies should be taken into consideration. Because of these limitations Barland considers his results a "worst case" scenario. Finally, interpretations must depend on the false positive and false negative rates which are deemed acceptable for particular purposes.

Findings

Separate statistical analyses were performed for the guilty knowledge and control question analog studies. The following data for the analog studies discussed above were reviewed: a) percentage of guilty subjects judged deceptive; b) percentage of guilty subjects judged nondeceptive (false negatives); c) percentage of guilty subjects judged inconclusive; d) percentage of innocent subjects judged deceptive (false positives); e) percentage of innocent suibjects judged truthful; and f) percentage of innocent subjects judged inconclusive.

As with the field studies, an index of predictive association lambda, guilty knowledge analog studies combined (see Tables 5.3 and 5.4). The results indicate that the control question test provides a 30 percent improvement in prediction over the base rate for these analog studies, and the guilty knowledge test a 57 percent improvement in prediction over the base rates, although because the studies differed so much, more accurate conclusions would be achieved by calculating lambdas for different studies or conceptually related groups of studies (see section 4). These figures should be interpreted with caution as in real life the base rate of guilt will vary considerably from approximately 50/50 distributions in laboratory experiments. Thus, it is difficult to draw unqualified conclusions from the analog studies given the wide variety of designs used.

The false negative rate for the analog studies of the control question technique ranged from 0 to 28.7 percent. Inconclusives ranged from 0 to 44 percent for guilty subjects and from 0 to 47 percent for innocent subjects. There is a wide range of false positives (4 to 51 percent). Global evaluations by the examiners, field scoring techniques, and purely statistical analyses of the data all seem to produce high detection rates in most studies. One exception is Kleinmuntz and Szucko's (1981) study, which found the validity coefficients of polygraph examiners' judgments markedly inferior to a purely statistical analysis of the charts. However, it is unclear how comparable their method of measuring validity is to the usual method of using an accuracy rate, and it is also not clear how applicable the lens model they use is to the question of the validity of the polygraph.

Another exception is Ginton et al.'s study (1982), in which field numerical scoring was found to be inferior to the global evaluation method

in detecting deception. However, the examiners in that study were Israeli polygraph professionals who may characteristically use a global method of assessment, and for whom the Utah numerical scoring system may have been somewhat foreign.

Accuracy of detection differed sizably between control question analog studies using students as subjects (Barland & Raskin, Bradley & Janisse, Szucko & Kleinmuntz; Widacki & Horvath is exluded as discussed above) and other control question analog studies (Podlesny & Raskin, Raskin & Hare, Rovner et al., Dawson, Ginton et al.). Experiments using students had lower percentages of correct decisions for both guilty and innocent, and more false negatives and false positives. Given the small number of studies in each category when the studies are divided in this way, it is unclear whether this difference is attributable to the nature of the subjects (student vs. nonstudent) or other characteristics of these experiments.

As shown in Tables 5.5 and 5.6, GKT analog studies tended to detect a somewhat lower average percentage of the guilty subjects (16.7 to 95 percent). They also had a relatively high proportion of false negatives (5 to 83.3 percent) but a lower rate of false positives, except for the Bradley and Janisse study. It should be noted, however, that the GKT was not assessed under conditions that deviated as much from the ideal as the control question test deviated. Nor were there as many studies testing the GKT as the CQT. This suggests that the confidence one can have in the GKT findings is, in general, less than the confidence one can have in the CQT analog findings.

In summary, there exist a number of studies of the control question technique; a smaller number of the concealed information test, all using the guilty knowledge test; and only two studies of the preemployment screening interview, one of them with government personnel. The analog studies systematically explored many of the technical variables associated with the polygraph (cf. the Utah group's study of control question techniques), and also studied the effect of several situational variables on the validity of the polygraph. The control question test was found to detect guilty subjects with a high degree of accuracy, but also to be subject to false positive errors. There was a large amount of variability among the control question analogs, especially the more they diverged in technique from the field method. The guilty knowledge test had a slightly lower rate of detection of guilt, more false negatives, but fewer false positives.

Table 5.1

Outcomes of Control Question
Analogue Studies of Validity

	GUILTY				INNOCENT				
	N	Correct	Incorrect (False Neg.)	Inconclusive	N	Correct	Incorrect (False Pos.)	Inconclusive	Total N
Barland & Raskin 1975	36	63.9%	8.3%	27.8%	36	41.7%	16.7%	41.7%	72
Podlesny & Raskin 1978	20	69.0	16.0	15.0	20	91.0	4.0	5.0	40
Raskin & Hare 1978	24	87.5	0	12.5	24	75.0	4.0	20.8	48
Rovner et al.[1] 1978	36	77.8	8.3	13.9	36	80.5	13.9	5.5	72
Dawson 1980	12	91.7	0	8.3	12	58.3	25.0	16.7	24
Widacki & Horvath[2] 1978	20	90.0	5.0	5.0	-	-	-	-	20
Bradley & Janisse	96			96					192
EDR		60.4	13.5	26.0		58.3	9.4	32.3	
Heart rate		35.4	20.8	43.8		33.3	19.8	46.9	
Szucko & Kleinmuntz[3] 1981	15	71.3	28.7	4	15	49.3	50.7	4	30

Table 5.1 (continued)

	GUILTY				INNOCENT				
	N	Correct	Incorrect (False Neg.)	Inconclusive	N	Correct	Incorrect (False Pos.)	Inconclusive	Total N
Ginton et al. 1982	2	100.0 %	0 %	0 %	13	84.6%	15.4%	0 %	15
Honts & Hodes 1982[a]									
No countermeasures	9[6]	67.0	0.0	33.	12	33.0	1 .0	50.0	
Countermeasures	19[6]	58.0	5.5	36.6	—[8]	--	--	--	
	28				12				40
Honts & Hodes 1982b									
No countermeasures	19	84.2	0.0	15.8		15.8	31.6	52.6	
Countermeasures	19	36.8	26.3	36.8	19				
	38				19				57
Heckel et al 1962									
Normals	--[7]				5	100.0	0.0	0.0	
Non-del. psychiatric	--				5	70.0	10.0	20.0	
(Delusional psychiatric)	[9]--				5	45.0	35.0	20.0	
Hammond 1980	32	71.9	3.0	25.0	30	40.0	20.0	40.0	62

Notes:
1. Summed across conditions.
2. Examiner's task was to detect the one guilty person in each of 20 groups of 4 suspects.
3. Based on ratings of 5+ on 1 to 8 scale of certainty of non-deception/deception.
4. Examiners were not allowed to categorize an examination as inconclusive.
5. 9 guilty subjects used pain countermeasures (tongue biting) and 10 used a muscle (toe pressing) countermeasure.
6. Original subject assignments, 12 to each of 4 (including 2 countermeasure) conditions. A total of 8 subjects were eliminated from the analysis of results for guilty subjects for failure to follow countermeasure instructions.

7 There was no guilty condition.
8. Innocent subjects used no countermeasures.
9. Not included in analysis reported in Table 5.3.

Table 5.2

Outcomes of Guilty Knowledge
Analogue Studies of Validity

	GUILTY				INNOCENT				
Total N	N	Correct	Non Correct (False Neg.)	Inconclusive	N	Correct	Non Correct (False Pos.)	Inconclusive	Total N
Lykken[1] 1959	50	88.0	12.0	-	48	100.0	0	-	98
Davidson 1968	12	91.7	8.3	-	36	100.0	0	-	48
Podlesny & Raskin 1978	10	80.0	20.0	0	10	80.0	0	20.0	20
Balloun & Holmes 1979	18			16				34	
Test 1		61.1	38.9	-		87.5	12.5	-	
Test 2		16.7	83.3	-		93.7	6.3	-	
Giesen & Rollison 1980	20	95.0	5.0	-	20	100.0	0	-	40
Bradley & Janisse 1981	96			96				192	
EDR		59.4	40.6	-		88.5	11.5	-	
Heart Rate		44.8	55.2	-		17.7	82.3	-	

Footnotes

[1]Frequencies for detection of two mock crimes were combined.

Table 5.3

Mean Detection Rates
as a Percentage of Total
in Analog Studies of
Control Question Technique

	Ground Truth	
	Percent Guilty	Percent Innocent
Examiners' Diagnosis	Mean	Mean
Deceptive	33.43	7.71
Nondeceptive	5.92	26.25
Inconclusive	12.77	13.92
	52.12	47.88

Lambda $_b$ = .406

Table 5.4

Mean Detection Rates
as a Percentage of Total
in Analog Studies of
Guilty Knowledge Test

	Ground Truth		
	% Guilty	% Innocent	
Examiners Diagnosis	Mean	Mean	
Guilty	27.9	2.2	
Not Guilty	17.3	52.6	
Inconclusive	0	0	0
	45.2	54.8	

Lambda $_b$ = .69

FACTORS AFFECTING POLYGRAPH EXAMINATION VALIDITY

Introduction

The analyses of both field and analog studies reported in sections 4 and 5 indicate that there is considerable variability in accuracy rates of polygraph examinations. To interpret these variations, numerous factors, such as the restricted range of techniques and applications tested in these studies, need to be considered. In addition, researchers have attempted to explain the variability in accuracy scores by proposing a number of factors that theoretically may affect polygraph test validity. These include characteristics of examiners, settings, and subjects. In addition, subjects have been known to use, or might be trained to use, a number of countermeasures to "beat" the polygraph. For many of these factors the research evidence is contradictory. For others, there has been little or no empirical testing. This section describes evidence from field and analog studies, as well as from laboratory investigations, on factors that may affect the accuracy of polygraph tests. The goals of this section are to assess the empirical evidence for the conditions under which polygraph testing is expected to be most valid and the conditions under which polygraph testing can be made ineffective. The section concludes by suggesting which factors affecting validity need additional research.

Examiner, Subject, and Setting Characteristics

The previously described analyses of field and analog studies (see sections 4 and 5) emphasize the characteristics of polygraph tests and their relation to accurate or inaccurate outcomes. In the present section, the focus shifts away from the tests, themselves, to additional factors that may affect validity. These factors are sometimes referred to as dimensions of external validity and aid in the assessment of the generalizability of research findings. Considerations of these factors will enable evaluation of the conditions under which various levels of validity may be expected from polygraph examinations. Differential validity in polygraph tests may be obtained with different examiners, subject populations, and with examinations conducted in different settings.

Examiner

It has long been recognized (cf. Lykken, 1981; Orne, 1975; Raskin, 1981; Slowik, 1979) that the examiner's skill has an important effect on the validity of polygraph tests. Examiner experience is an essential element reported by investigators and has often been used to explain differences in accuracy rates (Raskin, 1978; Raskin & Podlesny, 1979). There are some data to indicate that experienced examiners have better accuracy rates. In recognition of this outcome, training has been accorded a high priority both within and outside Government agencies which conduct polygraph examinations and by state legislators and polygraph examiner groups (cf. American Polygraph Association, 1983). An extensive array of training facilities now exists, offering a somewhat diverse set of orientations to polygraph testings.

Experience. A number of studies have tested how examiner experience relates to validity of polygraph examinations. Horvath and Reid (1971), for example, had charts utilized in their validity study reexamined by a

group of 10 polygraph examiners. Seven of the examiners were experienced and three of them were examiner-interns (each with less than 6 month's experience). According to Horvath and Reid (p. 279, see Table 1), experienced examiners made an average of 91.4 percent correct judgments, while the average for inexperienced examiners was 77.5 percent.

Training. Experience in conducting polygraph examinations suggests that there are a number of clinical components to detection of deception. To some extent, training programs capture these clinical elements by extensive training in "proper" examiner attitude and relationship with subjects. Increasingly, however, training programs emphasize standardized techniques for constructing questions and scoring examinations. In this respect, the U.S. Army Military Police School (USAMPS) is perhaps the best example. The school serves as the central training site for almost all Government agencies which maintain polygraph examiner staffs. The USAMPS teaches several versions of the CQT (including what they call the modified general question technique or MCQT and the original Backster's zone of comparison or ZOC method) and several specific protocols for selecting question sets and scoring polygraph charts. Trainees receive both didactic classroom training and supervised experience conducting polygraph examinations. The current curriculum for the USAMPS uses Reid and Inbau's (1977) text on polygraph testing, supplemented by materials prepared especially for its trainees (Army, 1977). The USAMPS is one of the number of training programs certified by the American Polygraph Association (cf. American Polygraph Association, 1983).

On the basis of presently available data, it is not possible to determine whether types of training have an effect on outcomes. A study by Raskin (1976) indicates that examiners trained in schools that emphasize numerical scoring were significantly more accurate than examiners who attended other schools (97.1 percent v. 86.9 percent). It is difficult to determine, however, if training in numerical scoring is more efficient or if better examiners/schools select such techniques. The fact that examiners who were trained in numerical techniques, but who did not use them, did more poorly than examiners trained in numerical techniques who used them (88.5 percent v. 98.9 percent) suggests that numerical evaluation rather than examiner selection (or some other aspect of the training) provides an advantage.

Subjects

Much effort in recent years has been devoted to development of systematic training. Less attention appears to have been paid to the characteristics of subjects of polygraph testing. Frequently, research reports of polygraph examination do not report even the most easily available data on subject characteristics (e.g., proportion of males and females). There have, however, been a number of studies of specific population groups (e.g. psychopaths) hypothesized to be less detectable. This section will consider, in addition to subjects' psychopathy, other diagnostic categories and subject variables such as gender, intelligence, motivation, and responsivity to arousal.

Subject factors are often described in the literature as personality or individual difference factors (Raskin, 1982; Waid & Orne, 1981). They refer to traits associated with individuals which may make them differentially detectable in a polygraph examination. Understanding these effects

should enable determination of the conditions under which polygraph testings will yield particular levels of validity. The mechanism by which subject variables affect polygraph examination validity has to do with differential autonomic arousal. Validity is affected when an interaction results between arousal and polygraph testing.

Psychopathy and level of socialization. One aspect of potential subject effects which has received considerable attention is the effect of level of socialization and psychopathy on detectability. In a series of studies by Waid and his colleagues (Waid, 1976; Waid, et al., 1979, a,b) significant relationships were found in the laboratory between socialization and autonomic responsiveness. An initial finding (Waid, 1976) was that college students who scored low on socialization (on a standard psychological inventory), gave smaller electrodermal responses (EDRs) to stimuli than did high scoring subjects. In a more directly relevant investigation (Waid et al., 1979), a group of college students was asked to deceive or not to deceive a professional polygraph examiner. Results indicated that subjects who were not detectable were significantly less socialized than those who were detectable. Susceptibility to detection seemed to be mediated by socialization and low socialization subjects showed reduced EDRs. Highly socialized subjects were more responsive electrodermally, and as a result, several of them were misclassified as deceptive.

Raskin (1982) has criticized Waid et al.'s (1979) research as not having practical significance for evaluations of polygraph validity. According to Raskin, simply demonstrating that there is a difference in responsibility on the first set of questions does not mean that subjects would not be correctly detected in an actual polygraph examination (which may involve 3 to 4 charts). Some of Raskin's own studies (e.g., Barland & Raskin, 1975; Raskin & Hare, 1978) suggest that psychopathic individuals are not less detectable than nonpsychopathic individuals. In Raskin and Hare's study, convicted felons, half of whom were diagnosed as psychopathic, performed a mock crime. These subjects were then administered a polygraph examination and offered a substantial monetary bonus if they could produce a truthful outcome. In contrast to Waid et al.'s findings, Raskin and Hare found that individuals diagnosed as psychopathic and/or low in socialization were more reactive and easily detectable than those not psychopathic and high in socialization. Earlier research by Raskin (Barland & Raskin, 1975) supports this finding. Barland and Raskin's (1976) field study, on the other hand, found that subjects who scored high on the Pd scale of the MMPI (a measure of psychopathy) had smaller cardio (but not respiration or SRR) scores than low Pd subjects.

In a previously described study, Balloun and Holmes (1979) conducted an analog study of college students using a "cheating" situation. Their results indicated that subjects who scored high on the psychopathic deviate scale of the MMPI (a standard psychopathology inventory) were just as easy to detect as were those individuals who scored low on the scale. It is important to note, however, that the polygraph test was a concealed information type of test, not a control question test (CQT) or relevant/irrelevant test (R/I). A doctoral dissertation by Hammond (1980; see section 5) also found no difference between normal and psychopaths.

Other Psychopathology. Guilty psychopaths may escape detection because they are not concerned enough about a misdeed to create interpretable

physiological responses. Individuals with other forms of psychopathology may escape detection or be classified as false positives for other reasons (e.g.. emotional instability, delusional thinking). The one study that has investigated this possibility (Heckel et al., 1962; see section 5) found, in fact, that innocent neurotics and particularly psychotics were likely to be identified as deceptive. There were no guilty subjects in this "real crime" analog study.

Gender. One of the most obivous subject differences is gender. Males and females may have different patterns of autonomic arousal, and such differences may affect polygraph testing validity. Unfortunately, few data exist to examine this hypothesis; most research only studies male subjects. In one study (Giesen & Rollison, 1980), all female subjects were tested in a mock-crime situation using the guilty knowledge test (GKT). The GKT was found to be highly accurate, but because males were not also tested, it is impossible to determine if males would have been less detectable. The 2 Honts and Hodes (1982a, b) analog studies included both females and males; the researchers do not report any gender differences in detectability.

Intelligence. Intelligence is an additional variable which potentially might affect detectability. Particularly if an R/I or CQT test is involved, the ability of intelligence subjects to anticipate questions may affect polygraph accuracy. One possibility is that intelligence subjects are less detectable because, if trained, they are able to anticipate questions and employ countermeasures. Another possibility is that because intelligence subjects better understand the implications of a polygraph examination, they will respond to relevant questions with heightened arousal when they are attempting to deceive (Barland & Raskin, 1973). There has been relatively little research on intelligence and polygraph testing. In one of the few experiments which look at intelligence and detectability, Kugelmass(1967) found no correlation between intelligence and responsivity on a peak of tension (POT) card test. On the other hand, research by Gustafson and Orne (1963) found that motivation to deceive increased the probability of detection. Barland and Raskin (1973) feel this is evidence of the potential role of intelligence. Barland and Raskin's study, in 1976, which compared detection rates among subjects of different educational levels found no difference. However, a separate analysis of the sources of false positive errors by Raskin (1976) found that the majority of false positives occurred among subjects who had college degrees. Level of education, of course, is not a perfect indicator of intelligence, and there is a need to better understand these relationships.

Ethnic and group differences. Another category of subject differences that may affect subject polygraph validity has to do with ethnic and group differences in physiological response. Research conducted cross culturally (e.g., Kugelmass & Lieblich, 1968; Laarus, 1966; Sternbach & Tursky, 1965), indicates that there are ethnic differences in response to stress. Such differences may, in turn, affect detection of deception. As noted earlier, these effects may interact with the ethnic identification of the examiner. However, effects of ethnic differences have not been directly tested with respect to polygraph examinations.

Autonomic lability. A final individual difference is what Waid and Orne (1981) have referred to as autonomic lability. Regardless of other

differences among subjects, there may be consistent individual differences connected with their level of autonomic arousal.

Although there is considerable variance for an individual in autonomic responses to most physiological measures of autonomic nervous system (ANS) arousal, electrodermal lability may be different. Given the importance of the EDR for polygraph examinations, it may be essential to understand more about this factor. Unfortunately, most of this research (e.g., Waid, Wilson & Orne, 1981) has been conducted with concealed information tests and not with the CQT or R/I.

Setting

One theory underlying lie detection using the polygraph is that the threat of punishment leads an individual to manifest a physiological reaction (Davis, 1961). This suggests, then, that settings in which an individual is more certain of being detected and in which the consequences are greatest, will permit higher levels of detection. Further, in order to be certain of being detected, a subject must believe in the efficacy of the polygraph procedures in order for it to function. According to some (e.g., Waid & Orne, 1981), the polygraph is often used somewhat like a "stage prop" and its presence is meant to "enhance the subject's concern" (p. 74). Stimulation tests, used in almost all field polygraph examinations, serves the same function, albeit more directly. There is considerable discussion (e.g., Weir, 1978), in fact, about how frequently within a polygraph examination such stimulation tests should be utilized in order to increase the validity of the examination.

Instrument presence. Some research, reported by Orne and his colleagues, addressed the question of the situational features necessary for a polygraph examination. In one component of a study reported by Orne et al. (1972), subjects were led to believe that the polygraph recording equipment was not operative. There was some indication that the pretest condition in which subjects were led to believe that the polygraph instrument was inoperative produced a lower detectability; however, results were not statistically significant. In an earlier study (Thackray & Orne, 1968), detectability was not affected by subjects' belief in whether the machine was recording. Both of these studies involved use of concealed information tests.

A more recent study by Orne's group (Waid, Orne & Wilson, 1979) tested a similar hypothesis using a different procedure. In this study, subjects saw the polygraph machine turned off, although the experimenters actually ran the leads to a second polygraph device and were able to record responses during a pretest review of questions. The results indicated that subjects who were aware of being recorded had significantly higher responses to relevant questions and not significantly difference responses to control questions.

Bogus pipeline. An interesting and potentially important aspect of how the polygraph achieves valid results is based on what social psychologists call the "bogus pipeline" (Jones & Sigall, 1971). The bogus pipeline is a procedure used to elicit veridical attitudes in situations where social desirability effects (i.e., subjects' desire to express socially acceptable opinions) may mask actual attitudes. The procedure involves

attaching subjects (via skin electrodes) to an ostensible physiological recording device called the "electromyograph" (EMG) and providing subjects with a "steering wheel" device to record their attitudes. In a typical study (Jones & Sigall, 1971), subjects were told that the EMG measured implicit muscle potentials and that it was an improved polygraph or "lie detector." The recording device is actually "electrical junk" (Jones & Sigall, 1971), and the purpose of the procedure is simply to convince subjects that their actual attitudes are detectable.

Results from a number of investigations which have used the bogus pipeline procedure (e.g., Quigley-Fernandez & Tedeschi, 1978; Sigall & Page, 1971) support Jones and Sigall's premise. Several studies indicate that when subjects believe that their attitudes are detectable by a physiological recording device, they more readily express their actual attitudes. Although it is difficult to know what "actual" attitudes are, higher veridicality is assumed with the bogus pipeline because the procedure yields more socially undesirable responses than when it is not used. For example, in Sigall and Page's (1971) initial experiment, they found that subjects in the bogus pipeline condition would admit to negative attitudes about "Negroes." Similar subjects in nonbogus pipeline conditions using paper-and-pencil tests would not reveal such attitudes. Later research has shown that this finding holds for attitudes toward handicapped individuals and for "confessing" to having prior knowledge about a psychological experiment.

Although the bogus pipeline research suggests that the conditions of testing may have important effects on polygraph subjects, it is not clear that these effects interact with the test itself. In a substantial number of criminal investigations subjects voluntarily confess after having the polygraph procedure explained or being shown the results of the examination. In personnel screening, subjects often admit to errors in their job application or past indiscretions (Barton, 1976; DCI, 1980). Most available field and analog research does not permit determination of the extent to which the polygraph is used in this way.

Specific settings. Although there are a number of settings in which polygraph examinations take place, and it is reasonable to assume that the setting may interact both with subject and examiner characteristics to affect the validity of polygraph tests, there is no research that directly tests the impact of different settings on polygraph testing validity.

Countermeasures

Countermeasures are deliberate techniques used by deceptive subjects to avoid detection during a polygraph examination (Reid & Inbau, 1977; Barland & Raskin, 1978; Lykken, 1981; Waid & Orne, 1981, 1982). Countermeasures can range from simple physical techniques, to so-called mental countermeasures, to the use of drugs and biofeedback techniques. There is a potentially large list of such countermeasures, and there are a number of plausible, but not yet validated, techniques to avoid deception. The research on polygraph countermeasures is summarized below by type of countermeasure.

Physical

Physical measures taken by a subject during a polygraph examination

are, perhaps, the most frequently discussed countermeasures used by subjects (Barland & Raskin, 1973; Lykken, 1981). Any physical activity which could affect physiological responses is a potential problem for interpretation of a polygraph test record. There is no question that physical measures, from tensing muscles to biting the tongue, to squeezing toes, to shifting one's position can affect physiological response.

There are frequent references to the use of such measures (see e.g., Commerman, 1981; Lykken, 1981a). But little systematic research has been conducted to establish the impact of the use of such measures on polygraph decisions. Kubis (1962) found that when subjects press their toes towards the floor they were able to reduce the probability of detection from 75 to 10 percent. A replication of this experiment, however, by More (1966) found that there was no decrease in detectability caused by toe movements. In two more recent studies, by Honts and Hodes (1982a,b), the efficacy of two physical countermeasures was tested in analog situations. Both studies found that countermeasures allowed subjects to "beat" the polygraph. Furthermore, there was a large percentage of inconclusives. In addition, both studies found that experienced examiners were not able to detect use of the countermeasures. A recent study by Honts, Raskin and Kircher (1983) also found that the use of physical countermeasures decreased detectability; the false negative rate for countermeasure subjects was 78 percent. However, examiners using a separate electromyographic (EMG) analysis were able to detect the use of countermeasures 80 percent of the time.

Thus, the evidence, while limited, is that deceptive subjects who use physical countermeasures and who can distinguish non-relevant from relevant questions (in a CQT or R/I test) can increase their chances of avoiding detection.

Drugs

In contrast to physical measures, which potentially may be detected by an observant polygraph examiner by running multiple charts or by the use of particular physiological measures, the use of various pharmacological agents is probably more difficult to detect. Not only may drugs be difficult to detect by observation, but they may also not be detected by multiple polygraph tests. Some theorists have suggested that such pharmacological agents have the potential to produce incorrect or uninterpretable polygraph records.

Research on drugs factors is only beginning to be conducted. Recent research by Waid (Waid, Orne, Cook & Orne, 1981) indicates that the tranquilizer, meprobamate (Miltown®), permits subjects who are being deceptive to increase their ability to avoid detection in a polygraph examination. One feature of tranquilizers such as meprobamate is that they suppress autonomic activity which may not be accompanied by any observable psychomotor differences. In Waid et al.'s study a GKT was used in a polygraph test. Subjects were all male and divided into three groups: a) a tranquilizer group; b) a placebo group; and c) a control group. Only 3 of 11 guilty subjects who had taken meprobamate were scored as deceptive.

It should be noted that Waid et al.'s investigation involved the GKT. According to Raskin (1982), a different problem would be encountered by attempts to utilize tranquilizers to defeat an examination employing the

CQT. The use of such drugs in a CQT polygraph examination might be more likely to yield inconclusive findings, rather than errors. This interpretation is supported by the recent analog study of Gatchel et al.(1983), which found that the use of propranolol, a beta-blocking drug, resulted in a 32.2 percent inconclusive rate, although the overall error rate was low. An additional finding was that examiners could not tell which subjects had used the drug. Conclusions drawn from this study must be limited by the fact that subjects were students. Other studies using college students (e.g., Hammond, 1980; Honts & Hodes, 1982a, b) have also resulted in large numbers of inconclusives.

A recent study by Iacono et al. (unpublished) found that ingestion of neither 10 milligrams of diazepam (Valium®) nor 20 milligrams of methylophenidate (Ritalin®) affected the accuracy of detection. Results in both active drug conditions were more accurate than when subjects ingested a placebo (a capsule containing lactose).

Research on other psychoactive drugs has not been reported in the literature, although such research is now being planned under the auspices of the National Security Agency and the Army Intelligence and Security Command. There are also no data as to the use of common drugs by actual polygraph examination subjects. Although examiners normally ask subjects to report use of medications or other drugs, blood samples or other detection means are typically not employed. It is thus difficult to assess the magnitude of drug use by subjects in previous research on the validity of polygraph testing.

In addition to drugs, there have also been reports of the use of various chemicals to confuse physiological recording (see Barland & Raskin, 1973). Placing antiperspirant powder, clear nail polish, or other agents on the balls of one's fingers may make the electrodermal responses less reliable. Such measures, however, should be detectable by a trained examiner.

Hypnosis/Biofeedback

There is a substantial literature in psychology about the use of hypnosis and biofeedback to alter and condition physiological responses. There is some evidence (see Sarbin & Slagle, 1979) that hypnosis, for example, induces declines in skin conductance levels. A number of investigations have attempted to show that hypnotically suggested amnesia is an effective countermeasure. Such research seems to indicate that hypnosis is not effective (see Barland & Raskin, 1973).

Recent research by Corcoran, Lewis and Garver (1978) has examined the effects of biofeedback training on suppressing electrodermal responses. They found that both hypnosis and biofeedback groups were able to reduce detectability after training as compared to a control group. In another study, Rovner, Raskin and Kircher (1979) reported that subjects who received extensive information about the nature of lie detection and practiced using countermeasures were detected significantly less than subjects without such training. It seems clear that if hypnosis or biofeedback operate as countermeasures, especially with commonly used tests such as the CQT, that extensive training would have to accompany their use.

Mental

Another category of countermeasures involves those that get the subject to think differently about the test. As noted earlier, most polygraph examinations rely on the subject's motivation to avoid detection rather than on any response directly connected with "telling a lie." Simple cognitive countermeasures include patterns of thinking that suppress responses to control or irrelevant questions. More complex cognitive countermeasures are based on knowledge of the results of the examination and lack of belief in one's detectability.

Controlling thoughts. Any individual who understands the basic structure of a particular polygraph examination should be able to differentiate irrelevant and control from relevant questions (when an R/I or CQT examination is conducted.) Given that when a CQT is employed, the questions are reviewed prior to their presentation, a subject intent on deception could try to alter cognitively responses to various questions (although since the order of presentation is varied, this is made difficult during an actual test). The procedure would be to try to dissociate oneself from the relevant questions and heighten response to control questions. Various means of such mental dissociation have been described (cf. Barland & Raskin, 1973; Lykken, 1981). They range from concentrating on an irrelevant object or by convincing oneself that the question means something different than intended.

There is little research evidence available to ascertain the potential of these techniques. Only one study on the use of cognitive countermeasures appears to have been conducted, and it does not provide much information. In this study, Dawson (1980) used as subjects actors trained in the "method" school of acting. Subjects were instructed to try to defeat the polygraph in a CQT examination. The results indicated that without exception the actors were detected and were not able to control their affective responses. It should be noted, however, that the actors in Dawson's study were merely told to "appear innocent" and were not told what would make them appear nondeceptive on a polygraph test.

Knowledge of results. Another set of countermeasures is based on subjects having knowledge of the results of their polygraph examinations. In criminal situations, particularly when an inconclusive outcome is obtained or when the subject disagrees with the outcome of an initial polygraph test, subjects are retested. In noncriminal situations, individuals are often tested at the beginning of their employment and at a number of subsequent points. In addition, subjects who know that they will receive a polygraph examination may seek training in methods to avoid detection. For all of these reasons, it is important to understand how feedback about polygraph examinations affects validity and whether prior experiences represent a potential countermeasure.

In an early laboratory study (Gustafson & Orne, 1965), subjects were given a stimulation test and feedback concerning its outcome. Feedback was manipulated so that some subjects thought they had successfully avoided detection and other thought that they had not. Subjects were motivated on a subsequent trial to avoid detection (they were told, "only mature and stable individuals are able to fool the lie detector"). The results indicated that subjects who believed that they had avoided detection were much less

detectable on the second trial (13 out of 16 were not detected, while only 1 of 16 in a control group were able to avoid detection). It should be noted that a stimulation test is a form of a concealed information test and the results may be due to lowered overall arousal. "Beating" a CQT represents a somewhat different problem.

A recent study, by Rovner et al. (1979), tested a similar hypothesis in a CQT examination. Several groups of subjects were placed in a mock crime situation. One group was given information about the nature of a CQT examination and information on what physiological reactions they should try to simulate. Another group was given information plus two practice tests involving actual physiological recordings after which they were told whether or not they had beat the polygraph. A third group served as a control and was given a typical polygraph examination. The results indicated that the information only and control groups were not able to avoid detection; however, 25 percent of the guilty subjects in the information plus practice group were able to avoid detection. Raskin (1982) maintains that this 25-percent error rate should be considered the "upper limit" because, in actual field situations, motivation would be much higher. Although Raskin is, perhaps, correct, it is also possible that in actual situations (where motivation is high), subjects might engage in more practice.

Belief in "machine". A final countermeasure is based on research about the bogus pipeline (Jones and Segall, 1971) and the role of the setting in inducing valid outcomes. If the validity of polygraph testing is dependent on the belief by subjects in the efficacy of the procedure, then it is possible that a countermeasure would involve training subjects to believe that the polygraph does not work. This might be done, for example, by providing subjects with false feedback on a polygraph examination. Unfortunately there is little research in this area, and the two studies that have been conducted come to different conclusions about the effect of belief in the techniques' effectiveness. In one study, Bradley and Janisse (1981) tested the hypothesis by rigging a stimulation test at various levels of effective detection. Depending on the condition, subjects were "detected" on one, two, or three trials to create the impression that the detection measures were ineffective, sometimes effective, or perfectly effective. For the electrodermal response measure, the more effective the apparatus appeared to be, the more the innocent subjects scored as nondeceptive and the more the guilty subjects scored as deceptive. In an earlier study, however, Timm (1979) found that feedback about the techniques effectiveness had no effect on whether subjects deceptiveness or nondeceptiveness could be detected. The theoretical support provided by research on the bogus pipeline indicates that subjects' belief in the technique needs additional reseach.

Research Implications of Factors Affecting Validity

Regardless of the established scientific validity, and in the absence of Federal legislation, it is likely that polygraph examinations will continue to be conducted as part of criminal investigations. Although, perhaps less likely, it is possible that critiques of polygraph examinations may result in more circumscribed use. Whatever the situation, however, understanding how polygraph examinations function may conceivably be important for national security purposes, if for no other reason than for purposes of counterintelligence. If one assumes that national security

requires that every means available be used to maintain the security of classified information, continued basic and applied research may be justified. If further research on polygraph testing is carried out, a number of research priorities can be identified on the basis of the present analysis. These priorities include research on countermeasures, comparative research on question techniques, and field-based research.

Examiner training and polygraph techniques. There are a number of research priorities for developing better polygraph techniques and training examiners to use them. Although probably not most important, these include development of standardized scoring techniques and development of more reliable measures of physiological arousal. Perhaps greater importance is research on the use of question techniques in different situations.

Scoring. Research is currently being conducted by the Army on development of computer scoring systems. There is some evidence (e.g., Kleinmuntz & Szucko, 1981) that the validity of polygraph examination decisions is improved if the clinical judgment of examiners is removed (see also, Ben-Shakhar et al., 1983) and related evidence that numerical scoring is more accurate (Kircher & Raskin, 1983; Raskin, 1976) than non-numerical scoring. Research should proceed on developing analogs to digital scoring systems. Such research, however, would not address the impact of examiner-examinee interaction, and this area also needs further study.

Question techniques. Another research priority is to develop additional protocols for question development. Perhaps the most important research need in this regard is to develop and field-test the concealed information test. Basic research and theory (see, e.g., Ben-Shakhar et al., 1983; Lykken, 1981; Raskin, 1982) suggests that such examinations have the highest likelihood of detecting deception, although no field research has yet been conducted to examine their use. Such research should both establish means of constructing GKTs and their validity in actual use.

An additional priority is to develop and test question techniques that may be employed in screening situations. Some examiners for example claim to use a version of the CQT for screening examinations (see section 2). This application of the CQT has not been subjected to scientific tests of validity. In addition, efforts should be devoted to testing the use of the CQT with different subject groups and in national security investigations.

Countermeasures. If the polygraph is to be more widely employed in national security investigations, there is an urgent need for research on countermeasures. Particular priorities would be research on drugs, biofeedback training, and subject gullibility and motivation. Such research needs to be carried out both in field situations and in the laboratory.

There are a number of drugs that are suspected of lowering ANS arousal and that theoretically may be able to invalidate the results of a polygraph examination or compel an "inconclusive" finding. A first priority is to extend Waid et al.'s (1981) research on meprobamate (which reduced detectability) to other psychoactive drugs.

Biofeedback training, as well as other forms of training have not been investigated, yet their effects on polygraph examinations may be

substantial. Subjects' beliefs about the accuracy of the polygraph may also be critical. As suggested by the research on the "bogus pipeline," individuals who believe their underlying thoughts are detectable, are more likely to provide veridical responses. The reverse phenomenon seems feasible and it would seem possible to train individuals to believe that the polygraph is ineffective. Such training might be accomplished by providing individuals with false feedback on the polygraph as well as by specific instructions during simulated polygraph examinations. Similarly, subjects who can be easily trained to beat the polygraph may be more desirable as intelligence agents.

Comparison with other techniques. Only one study in the available literature (Widacki & Horvath, 1978) systematically compares the polygraph with other investigatory tools. There is a need to examine whether the polygraph provides independent or corroborative evidence and whether the judgments made by polygraph examiners are merely a function of their clinical judgment as investigators, or whether it is a function of the polygraph examination itself.

A complication with this research is that the polygraph functions, in many situations, as a threat. Individuals' fear of taking the examination, in many instances, may lead them to confess or provide incriminating evidence. The threat potential, however, is in part a function of theirs and others' knowledge of research results. If, for example, it became widely known that the polygraph was "beatable," it is likely that this threat would be reduced and, hence, the validity (and utility) of the polygraph would be reduced.

Field studies. As described in sections 3 and 4, there are numerous problems with the available field and analog evidence. Field studies are problematic because they can only be conducted where an independent criterion of guilt or innocence is available. As such, these studies may be a select sample of cases (e.g., where guilt is overwhelming) and is certainly a select set of examiners. Analog studies have a different set of problems and have not adequately motivated subjects or may not have the appropriate number of cases. What is needed is research which deals with the problems of the available field and analog research.

One "theoretical" solution to the problem of conducting systematic field studies is to conduct "AMSCAM"-like investigations using bogus unauthorized disclosures instead of bribes; that is, for investigators to set-up a situation where they have knowledge of who is guilty or innocent. The polygraph could be used to select guilty from innocent with a known base rate and ground truth. Such methods, of course, raise a number of ethical, legal and pragmatic questions and it is not clear whether they could provide definitive answers. They could not be used frequently nor with a wide range of techniques/situations. Conducting polygraph research presents serious conceptual and methodological problems; in the absence of such research, however, it will not be possible to develop fully an assessment of the validity of polygraph examinations.

Conclusions

The description in this section of factors affecting validity and potential countermeasures suggests that there is a great deal more to

understand about polygraph tests before one can be assured of its validity. Despite our lack of knowledge in certain areas, there are clearly things that can be done to improve validity. In part, the history of polygraph development over the past 15 to 20 years has been to systematize and improve procedures. One central problem, not adequately addressed by either the literature on improvements in validity or countermeasures, is the extent to which these factors affect error rates or affect numbers of inconclusives. For policy purposes, clearly such distinctions and a sense of the magnitude of false decisions is needed. Substantial research, beyond what is currently available, would have to be conducted in order to answer such questions.

CONCLUSIONS

Introduction

The primary purpose of this study is to evaluate the scientific evidence on polygraph validity. In order to do this, OTA studied the actual polygraph exmaination process, reviewed the results of prior research reviews, analyzed a wide range of field and analog studies, and surveyed Federal agencies for any relevant research.

A major conclusion is that polygraph testing is, in reality, a very complex process that varies widely in application. While the polygraph instrument itself is essentially the same for all applications, the purpose of the examination, type of individual tested, examiner training, setting of the examination, and type of questions asked, among other factors, can differ substantially. The instrument cannot itself detect deception. Therefore, the polygraph test requires the examiner to develop questions to be asked in each case, compare the physiological response (as measured by the instrument) to the different questions, and infer deception or truthfulness based on these comparisons.

A second major conclusion is that no overall measure or single statistic of polygraph validity can be established based on available scientific evidence. The amount and quality of the evidence depends on the particular application. Some applications (e.g., the use of the polygraph in criminal investigations) have been fairly heavily researched, while others (e.g., polygraph use in pre-employment screening) have had very little research attention.

Therefore, conclusions about polygraph validity can only be made in the context of specific applications. The three types of applications discussed below are: specific-issue criminal invstigation, specific-issue screening investigation, and personnel security screening (including pre-employment or pre-clearance and aperiodic).

Specific-Issue Criminal Investigation

A principal use of the polygraph test is as part of an investigation (usually conducted by law enforcement or private security officers) of a specific situation in which a criminal act has been alleged to have, or in fact has, taken place. This type of case is characterized by a prior investigation that both narrows the suspect list down to a very small number, and that develops significant information about the crime itself. When the

polygraph is used in this context, the application is known as a specific-issue or specific incident criminal investigation.

Results of OTA review. This type of application of the polygraph is the only one to be extensively researched. OTA identified 6 prior reviews of such research (summarized in section 3), as well as 10 field and 12 analog studies that met minimum scientific standards and were conducted using the control question technique (the most common technique used in criminal investigation; see section 2). Still, many of these research studies had various methodological problems that reduce the extent to which results can be generalized. In addition, only some question techniques have been researched, and the effect of different types of examiners, subjects, and settings has not been systematically explored. Finally, the basic theory (or theories) of how the polygraph test actually works has only been minimally researched.

Nonetheless, this research is the best available source of evidence on which to evaluate the scientific validity of the polygraph for specific-incidence criminal investigations. The results (for research on the control question technique) are summarized below:

- 6 prior reviews of field studies
  - -- Average accuracy ranged from 64 to 98 percent.
- 10 individual field studies
  - -- Correct guilty detections ranged from 70.6 to 98.6 percent.
  - -- Correct innocent detections ranged from 12.5 percent to 94.1 percent.
  - -- Average correct guilty detection rate was 86.3 percent.
  - -- Incorrect guilty detections (false positives) ranged from 0 to 75 percent.
  - -- Incorrect innocent detections (false negatives) ranged from 0 to 29.4 percent.
  - -- Average false positive rate was 19.1 percent (percent of innocent persons found deceptive.)
  - -- Average false negative rate was 10.2 percent (percent of guilty persons found non-deceptive.)
- 12 individual analog studies
  - -- Correct guilty detections ranged from 35.4 to 100 percent.
  - -- Correct innocent detections ranged from 5 to 91 percent.
  - -- Average correct guilty detection rate was 64.1 percent.
  - -- Average correct innocent detection rate was 54.8 percent.
  - -- False positives ranged from 4 to 50.7 percent.
  - -- False negatives ranged from 0 to 28.7 percent.

-- Average false positive rate was 16.1 percent.

-- Average false negative rate was 11.4 percent.

The wide variability of results from both prior research reviews and OTA's own review of individual studies makes it impossible to determine a specific overall quantitative measure of polygraph validity. These variations in results in part reflect differences in research design.

The preponderance of research evidence does indicate that, when the control question technique is used in specific-incident criminal investigations, the polygraph detects deception at a rate better than chance, but with significant error rates. Taking the averages for the 10 field and 12 analog studies, the range of polygraph validity results can be summarized as follows:

- o correct guilty detections
  68 to 86 percent

- o correct innocent detections
  49 to 76 percent

- o false positives (innocent persons found deceptive)
  19 to 28 percent

- o false negatives (guilty persons found innocent)
  10 to 13 percent

These figures are strictly averages for two groups of research studies. Another selection of studies would yield different results. Also, some researchers exclude inconclusive results from the base rate. OTA elected to include the inconclusives on the grounds that an inconclusive is an error in the sense that a guilty or innocent person has not been correctly identified. Exclusion of inconclusives would raise the overall accuracy rates calculated. In practice, inconclusive results may be followed by a Pre-test or other investigations.

Relevance to NSDD-84. However, there is only a limited scientific basis for generalizing the results of the OTA review to the context of National Security Decision Directive 84. NSDD-84 authorizes the use of the polygraph in administrative as well as criminal investigations of unauthorized disclosures of classified information.

First, the Federal Bureau of Investigation has concluded that, even in a criminal context, "to date, no methodologically adequate study of control question techniques has been reported ... Inferences regarding the validity of control question examinations ... rest upon the results of laboratory studies conducted under highly dissimilar conditions" (FBI, 1983). The FBI is planning its own validity research.

Second, the investigative conditions authorized by NSDD-84 may be quite different from conditions under which prior research was conducted or, for that matter, under which the FBI currently uses the polygraph. NSDD-84 does not specify what type of investigative procedures will be followed, how subjects will be selected or identified, who will conduct the exams, or what question techniques will be used.

Third, there is no validity research directly on the use of the polygraph in unauthorized disclosure investigations. The subject matter and perhaps subjects of these individuals will vary from the typical criminal investigation.

On the other hand, to the extent polygraph use in unauthorized disclosure investigations is similar to the way the FBI uses the polygraph in criminal investigations, there is at least some although not conclusive scientific basis for polygraph validity.

Specific-Issue Screening Investigation

The polygraph test is used by some private firms and on very rare occasion by some Federal agencies to screen a large number of people in connection with the investigation of a crime.

NSDD-84 appears to permit such use of the polygraph in unauthorized disclosure investigations, although the actual extent of NSDD-84 is unclear.

There is no scientific basis for generalizing the results of the OTA review to establish polygraph validity in this application. First, no research has been conducted on specific-issue screening use of the polygraph. Second, the screening conditions here are likely to vary even more than specific-issue criminal investigations from the conditions of the research studies reviewed by OTA. For one thing, much less information is likely to be known about the circumstances surrounding the unauthorized disclosure and possible suspects if the polygraph testing is not preceded by a conventional investigation. This could translate into differences in the questions used, the behavior of the polygraph examiner, the motivation and response of the subject, and the effectiveness of countermeasures.

Third, the large-scale screening of the polygraph theoretically can be expected to result in significantly higher error rates. The screening situation is most dependent on the so-called base rate of guilt, that is, the percentge of the group of persons being screened that has engaged in the criminal (or otherwise proscribed) activity. If the percentage of guilty is small, say 5 percent (1 guilty person out of every 20 persons screened, or 50 out of 1,000), then even assuming a very high (95 percent) polygraph validity rate, the predictive value of the screening use of the polygraph would only be 50 percent. That is, for each 1,000 individuals screened, about 47 out of the 50 guilty persons would be correctly identified as deceptive, but 47 out of the 950 innocent persons would be incorrectly identified as deceptive (false positives). Thus of the 94 persons identified as deceptive, one-half would be innocent persons. For every person correctly identified as deceptive, another person would be incorrectly identified.

As another example, if a lower polygraph validity rate is assumed (say 90 percent), then the predictive value would be expected to drop to about 33 percent. That is, for every person correctly identified as deceptive, 2 persons would be incorrectly identified (false positives).

These are, of course, hypothetical examples, and have not been studied in field or analog research. Also, operating procedures of Federal

agencies (e.g., quality control review, consideration of other investigatory information) might override, catch, correct, or minimize false positive polygraph decisions.

Nonetheless, the FBI, which outside of DOD and CIA, is the principal Federal agency that conducts polygraph examinations, prohibits the "use of the polygraph for dragnet-type screening of large numbers of suspects or as a substitute for logical investigation by conventional means" (FBI Polygraph Regulation 13-22.2(2), 1980). Furthermore, the FBI permits a polygraph examination to be administered only when "there is reasonable cause to believe that the person to be examined has knowledge of or was involved in the matter under inquiry or investigation, or if the person is withholding information relevant to the inquiry or investigation" (FBI Polygraph Regulation 13-22.4(2), 1980).

Personnel Security Screening

Draft revisions to the DOD polygraph regulations would authorize the use of the polygraph for determining initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (Sensitive Compartmented Information and/or special access). The use of the polygraph in determining continuing eligibility would be on an aperiodic (i.e., irregular) basis (DOD, 1983). These are all known as personnel security applications of the polygraph.

DOD research. OTA inquiries to all DOD components using the polygraph identified only one DOD research study on personnel screening use of the polygraph (Barland, 1981). The results of this study raise more questions than they answer, and certainly do not provide support for high polygraph validity in a screening situation. The limitations of the study reduce its applicability, but it is the only DOD polygraph screening research known to OTA.

Other research. Three other studies were cited by the intelligence agencies (National Security Agency and Central Intelligence Agency) as providing support for personnel security use of the polygraph test.

A 1975 field study (Edel and Jacoby) of polygraph screening of government job applicants (from an unidentified Federal agency) showed high consistency among the polygraph chart interpretations of different examiners. But this study concluded nothing about validity and, furthermore, was not conducted in a national security context.

A 1981 analog study (Correa and Adams) on pre-employment screening use, 75 percent of the responses of deceptive individuals were detected accurately. Twenty-five percent were detected incorrectly. Any conclusions based on this study must be limited by the fact that the subjects were students, and the questions and context had nothing to do with national security.

A 1980 survey conducted by the Director of Central Intelligence Security Committee concluded that the polygraph was the most productive of all background investigation techniques. However, this was a utility study not a validity study, and had many limitations and qualifications.

OTA recognizes that NSA and CIA believe that the polygraph is a useful screening tool. However, OTA concluded that the available research evidence does not establish the scientific validity of the polygraph for this purpose.

In comments to OTA, the CIA agreed that the cumulative unclassified research evidence reviewed by OTA is not directly relevant to national security applications. However, the CIA does claim to have classified research to support their use of the polygraph. OTA did not review this research. No other Federal agency, including NSA, has claimed to have relevant research results that were not available for OTA review on an unclassified basis.

False positives. One area of special concern in personnel security screening is the incorrect identification of innocent persons as deceptive. All other factors being equal, the low base rates of guilt in screening situations would lead to high false positive rates, even assuming very high polygraph validity. For example, a typical DOD polygraph screening situation might involve a base rate of 1 guilty person (e.g., 1 person engaging in unauthorized disclosure) out of 1000 employees. Assuming that the polygraph is 95 percent valid, then, the 1 guilty person would be identified as deceptive but so would 50 innocent persons. The predictive validity would be about 2 percent. Even if 99 percent polygraph validity is assumed, there would still be 10 false positives for every correct detection of a guilty person.

Again, these are hypothetical examples that have not been the subject of field or analog research. NSA officials have agreed with the previous analysis as hypothetically correct. But NSA claims that they in fact have experienced a very low false positive rate and that, in any event, polygraph results are only one factor in making decisions and are subject to quality control checks and other reviews. It appears that NSA (and possibly the CIA) use the polygraph not to determine deception or truthfulness per se, but as a technique of interrogation to encourage admissions. NSA has stated that the agency "does not use the 'truth vs. deceptive' concept of polygraph examinations commonly used in criminal cases. Rather, the polygraph examination results that are most important to NSA security adjudicators are the data provided by the individual during the pre-test or post-test phase of the examination" (NSA, 1983).

The validity of the polygraph as used by NSA has not been researched. And, in general, this kind of application is potentially different in so many ways from the polygraph use in specific-incident criminal investigations that results of the OTA research review previously discussed cannot be generalized to the NSA situation.

False negatives/countermeasures. The primary purpose of the polygraph when used for personnel security screening is to detect persons who have or intend to participate in proscribed activities (e.g., unauthorized contact with a foreign agent, disclosure of classified information). The concern with false negatives (guilty persons identified as non-deceptive) is that, apart from any errors inherent in the polygraph test itself, the guilty person may be able to escape detection through the use of countermeasures.

The research on polygraph countermeasures (e.g., physical movement or pressure, drugs, hypnosis, biofeedback) has been limited and the results conflicting. This lack of research is particularly significant to the extent that the polygraph is used and relied upon for national security purposes, since even a small false negative rate could have serious consequences. Some research (e.g., Honts & Hodes, 1982) suggests that polygraph examiners may not be able to easily detect certain physical countermeasures. The research results for drug and psychological countermeasures are mixed.

However, in any event, those individuals who the Federal Government would most want to detect (e.g., for legitimate national security violations) may well be the most motivated and perhaps the best trained to avoid detection.

Ethical issues. The use of the polygraph for personnel security screening could exacerbate ethical concerns. The ethical aspects of polygraph use are a matter of debate that is outside the primary focus of the OTA study. However, the proposed revisions to DOD polygraph regulations include certain provisions that have significance both for ethics and validity.

As currently used in the Federal Government, with few exceptions, polygraph examinations are voluntary. That is, a person cannot be forced to take a polygraph test against his or her will. A refusal to take a polygraph test does not, or at least is not supposed to, result in adverse consequences. The only exceptions are NSA (and by extension, the CIA) and, under very limited conditions, the FBI. NSA notes that "the polygraph examination is part of the Agency's security processing. Failure to complete processing may result in failure to be accepted for employment" (NSA, 1983). FBI regulations require that "polygraph examinations will be administered only to individuals who agree or volunteer to take an examination" (FBI Regulation 13-22.2(3)). The only exception is for certain FBI employees and applicants under specified circumstances where "a refusal to be examined by polygraph may lead to an adverse inference being drawn"(FBI, 1983).

The DOD proposal would provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may result in adverse consequences for the individual. These include nonselection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position.

Under these circumstances, polygraph examinations would not be voluntary in the strict sense, since a refusal could result in penalties. Apart from the ethical and perhaps legal implications, conducting polygraph tests on this basis could affect test validity. NSA has stated that, in conducting screening examinations, "[t]he full cooperation of the individual taking the test is essnetial or the results will be inconclusive"(NSA, 1983).

Further Research

OTA concluded that, to the extent that polygraph testing is going to continue to be used by the Federal Government, further research is needed. Possible research priorities include the following.

Research on polygraph countermeasures. Since NSA and CIA are already heavily dependent on the polygraph, their use alone justifies an intensified research effort on countermeasures. NSA and the Army Intelligence are planning such research, but the level of effort appears low (e.g., $65,000 pilot study in NSA) considering the consequences of false negatives.

Personnel security screening validity. Given the almost total lack of research on this application, further research is clearly necessary if there is to be any possibility of establishing a scientific basis for the personnel security screening use of polygraph testing. The Army's current 10-year research program to develop a new state-of-the-art polygraph instrument should be reevaluated to determine if research priorities and direction need adjustment. As it stands now, validity issues will not be addressed until the late 1980s.

Criminal investigation validity. There is no definitive study on the validity of Federal agency use of the polygraph in specific-incident criminal investigations. The planned FBI-Secret Service validity study is intended to meet this need. However, the research plan should be subjected to extensive review by the scientific community and others before the research approach is finalized.

Polygraph theory. The basic theory of polygraph testing is only partially developed. The testing process is complex and not amenable to easy understanding. A stronger theoretical base is needed for the entire range of polygraph applications. Basic polygraph research should consider the latest research from the fields of psychology, physiology, and medicine; comparison among question techniques; and measures of physiological response.

References

Abrams, S., "Polygraph validity and reliability: a review," Journal of Forensic Sciences. 17:313-327, 1973.

Addington v. Texas, 441 U.S. 418, 1979.

American Polygraph Association, Manual for polygraph school inspections. American Polygraph Association, P.O. Box 74, Linthicum Heights, MD. 21090, 1983 a.

American Polygraph Association, Polygraph school accreditation standards. 1-15, 1983b.

American Psychological Association, Standards for educational and psychological tests. Washington, D.C.: APA, 1974.

American Psychological Association, Draft joint technical standards for educational and psychological testing. Unpublished draft, Washington, D.C., 1983.

Ansley, N. A review of the scientific literature on the validity, reliability and utility of polygraph techniques, National Security Agency, Ft. Meade, MD., 1983a.

Ansley, N., Quick reference guide to polygraph admissibility, licensing laws, and limiting laws. Severna Park, MD: American Polygraph Association, 1983b.

Ansley, N., Horvath, F., & Barland, G.H., Truth and Science: A Bibliography. A comprehensive index to international literature on the detection of deception and the polygraph (lie detector) technique (2nd edition). American Polygraph Association, Linthicum Heights, MD, 1983.

Backster, C., Methods of strengthening our polygraph technique. Police, 6:61-68, 1962.

Backster, C. Standardized polygraph notepack and technique guide. San Diego, CA: Author, 1979.

Balloun, K.D., & Holmes, D.S., "Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime," Journal of Applied Psychology. 64: 316-322, 1979.

Barefoot v. Estelle, SILW 5190 (6-28-83).

Barland, G.H., "A Survey of the Effect of the Polygraph in Screening Utah Job Applicants: Preliminary Results," In U.S. Congress, 1978.

Barland, G.H., An introduction to the number test. Polygraph. 7:203-208, 1978.

Barland, G.H., A validation and reliability study of the counterintelligence screening test. Unpublished manuscript. Security Support Battalion, Military Intelligence Group, United States Army, Fort George Meade, MD., 1981.

Barland, G.H., On the accuracy of the polygraph: An evaluative review of Lykken's "Tremor in the Blood," Polygraph. 11:258-272, 1982.

Barland, G.H., 1982b.

Barland, G.H., Detection of deception theory. Unpublished, November 1982.

Barland, G.H., & Raskin, D.C., Detection of Deception. In W.F. Prokasy & D.C. Raskin (Eds.), Electrodermal Activity in Psychological Research (pp. 418-471). New York: Academic Press, 1973.

Barland, G.H., & Raskin, D.C., "An evaluation of field techniques in detection of deception," Psychophysiology. 12:321-330, 1975.

Barland, G.H., & Raskin, D.C., "Validity and reliability of polygraph examinations of criminal suspects." (Report No. 76-1, Contract 75-NI-99-0001) Washington, D.C.: National Institute of Justice, Department of Justice, 1976.

Barland, G.H. & Raskin, D.C., 1978.

Barton, M., A Study of the Admissions Made During Pre-Employment Polygraph Examinations (1964-1975) and Their Significance to the Business Community. Unpublished Masters Thesis, Sam Houston State University, 1976.

Belt, J.A. & Holden, P.B., "Polygraph usage among major U.S. corporations," Personnel Journal. 57(2)(1978): 80-86.

Ben-Ishai, A., Some remarks on polygraph research. Paper presented at the Ninth Annual Meeting of the American Academy of Polygraph Examiners, Chicago, Illinois, August 1962.

Ben-Shakhar, G., Bar-Hillel, M., & Lieblich, I., Trial by polygraph: Scientific and juridical issues in lie detection. Hebrew University, 1983.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M., An evaluation of polygraphers' judgements: A review from a decision theoretic perspective. Journal of Applied Psychology 67:701-713, 1982.

Bersh, P.J. (1969). A validation study of polygraph examiner judgements. Journal of Applied Psychology 53: 399-403, 1969.

Bitterman, M.E. & Marcuse, F.L., "Cardiovascular responses of innocent persons to criminal interrogation," American Journal of Psychology 60: 407-412, 1947.

Blaylock, 1983 Testimony before joint hearings, 1983.

Blaylock, K.D., Testimony before Joint hearing of House Post Office and Civil Service Subcommittee on Civil Service and House Judiciary Subcommittee on Civil and Constitutional Rights on use of polygraphs and prepublication review. April 21, 1983.

Blum, R.H. & Osterloh, W., "The polygraph examination as a means for detecting truth and falsehood in stories presented by police informants," The Journal of Criminal Law, Criminology and Police Science 59: 133-137, 1968.

Borchard, E., (1932) Convicting the innocent. Garden City, NY: Garden City Publishing Company.

Bradley, M.T. & Janisse, M.P., "Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures," Psychophysiology 18: 307-314, 1981.

Brown, C.C., Methods in psychophysiology. Baltimore: The Williams & Wilkins Company, 1967.

Buckley, J., Letter to Fred B. Wood, Office of Technology Assessment 1983.

Carlucci, 1982.

Carlucci, F.C.,. Memorandum for Secretaries of Military departments. Chairman, Joint Chiefs of Staff; Directors of Defense, Agencies. "Degradation of operational readiness/mission accomplishment due to

personnel security investigation shortfalls." August 6, 1982.

Cimmerman, A., "They'll let me go tomorrow." Criminal Defense 8, 7-10, 1981.

Cook, T.D., & Campbell, D., Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1979.

Corcoran, J.F.T., Lewis, M.D., & Garver, R.B., "Biofeedback conditioned galvanic skin response and hypnotic suppression of arousal: A pilot study of their relation to deception," Journal of Forensic Sciences 23: 155-162, 1978.

Correa, E.I., & Adams, H.E., "The validity of the pre-employment polygraph examination and the effects of motivation," Polygraph 10: 143-156, 1981.

Cureton, R., "A consensus as to the validity of polygraph procedures," Tennessee Law Review 22: 18-32, 1953.

Cutrow, R.J., Parks, A., Lucas, N., & Thomas, K., "The objective use of multiple physiological indices in the detection of deception," Psychophysiology 9: 578-588, 1972.

Davidson, P.O., "Validity of the guilty-knowledge technique: The effects of motivation," Journal of Applied Psychology 52: 62-65, 1968.

Davidson, W.A., Validity and reliability of the cardio activity monitor. Polygraph 8: 104-111, 1979.

Davis, R.C., "Physiological responses as a means of evaluating information." In A.D. Biderman & H. Zimmer (ed.), The manipulation of human behavior. New York: Wiley, pp. 142-169, 1961.

Dawson, M.E., "Physiological detection of deception: Measurement of responses to questions and answers during countermeasure manuevers." Psychophysiology 71: 8-17, 1980.

Decker, R.E., "The Army stimulation test; a control procedure," Polygraph 7: 176-178, 1978.

Edel, E.C., & Jacoby, J. "Examiner reliability in polygraph chart analysis: Identification of physiological responses," Journal of Applied Psychology 60: 632-634, 1975.

Edwards, R.H., A survey: Reliability of polygraph examination conducted by Virginia polygraph examiners (Report for the Commonwealth of Virginia). Richmond, VA: Bureau of Forensic Science, Dept. of General Services. 1981.

Ekman, P., Levenson, R.W., & Friesen, W.V. (1983). Autonomic nervous system activity. Science 221: 1208-1210.

Elaad, E., & Schahar, E., Polygraph field validity. In I. Nachson (ed.), Scientific interrogation in criminal investigation. Selected papers

presented at the First National Conference on Scientific Interrogation in Criminal Investigation at Bar-Ilan University & Ramat-Gan, Israel. November 1976.

Ennis, B.J., & Litwack, T.R., "Psychiatry and the presumption of expertise: Flipping coins in the courtroom," California Law Review 62: 693-752, 1974.

Ferguson, R.J., The polygraph in Private Industry. Springfield, ILL: Charles C. Thomas, 1966.

Fingerhut, K.R. Use of the stimulation test in pre-employment testing. Polygraph 7:185-188, 1978.

Frye v. U.S., 293 F.1013 (DC Cir. 1923).

Gatchel, R.J., The effect of propanolol on polygraphic detection of deception. Unpublished manuscript. 1983.

Giannelli, P.C., The admissibility of novel scientific evidence: Frye v. U.S., a half century later. Columbia Law Review 80: 1198-1250, 1980.

Giesen, M., & Rollison, M.A., "Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance," Journal of Research in Personality 14: 1-11, 1980.

Ginton, A., Netzer, D., & Elaad, E., "A method for evaluating the use of the polygraph in a real-life situation," Journal of Applied Psychology 67: 131-137, 1982.

Glass, G.V., McGaw, B., & Smith, M.L. Meta-analysis in social research. Beverly Hills, Calif: Sage, 1982.

Goodman, L.A., & Kruskal, W.H., "Measures of association for cross-classification," Journal of the American Statistical Association 49: 732-764, 1954.

Gustafson, L.A., & Orne, M.T., "Effects of heightened motivation on the detection of deception," Journal of Applied Psychology 43: 408-411, 1963.

Gustafson, L.A., & Orne, M.T., "The effects of task and method of stimulus presentation on the detection of deception," Journal of Applied Psychology 49: 412-417, 1964.

Harnon, E., Evidence obtained by polygraph: An Israeli perspective. The Criminal Law Review 340-349, 1982.

Harrelson, L.H., The Keeler technique; 2nd ed. Chicago: Keeler Polygraph Institute, 1964.

Hayden, 1978.

Hayden, T., Testimony of Trudy Hayden, ACLU Projective Privacy. In U.S. Congress Senate Judiciary Committee Hearings, 1978.

Hayden, T., "Employers who use lie detector tests," Business & Society Review 41: 16-21, 1982.

Hays, W.L., Statistics (3rd edition), New York: Holt, Rinehart & Winston, 1981.

Heckel, R.V., Brokaw, J.R., Salzberg, H.C., & Wiggins, S.L., Polygraphic variations in reactivity between delusional, non-delusional, and control groups in a "crime" situation. Journal of Criminal Law, Criminology and Police Science, 53: 380-383, 1962.

Honts, C.R., & Hodes, R.L., The effect of simple physical countermeasures on the detection of deception. Paper presented at the Society for Psychophysiological Research, Minneapolis, MN, 1982.

Honts, C.R., & Hodes, R.L., The effects of multiple physical countermeasures on the detection of deception. Psychophysiology, 19: 564-565 (Abstract), 1982b.

Honts, C.R., Raskin, D.C., & Kircher, J.C., "Detection of deception: Effectiveness of physical countermeasures under high motivation conditions." Paper presented at meetings of Society for Psychophysiological Research, September 1983.

Horvath, F.S., The Police Candidate Polygraph examination: Considerations for the Police Administrator, Police 33-39, 1972.

Horvath, F.S., "Verbal and nonverbal clues to truth and deception during polygraph examinations," Journal of Police Science and Administration 1: 138-152, 1973.

Horvath, F.S., "Detection of deception: A review of field and laboratory research," Polygraph 5: 107-145, 1976.

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

Horvath, F.S., Detecting deception in eyewitness cases: Problems and prospects in the use of the polygraph. In Wells, J. & Loftus, B. (eds.) Advances in the psychology of eyewitness testimony. New York: Cambridge University Press, in press.

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

Hunter, F.L. & Ash, P., "The accuracy and consistency of polygraph examiners' diagnoses," Journal of Police Science and Administration 1: 370-375, 1973.

Iacono, W.G., Boisvenu, G.A., & Fleming, J.A. (1983). The effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. Unpublished manuscript, University of British Columbia, Vancouver, Canada.

Jones, E.E., & Sigall, H., "The bogus pipeline: A new paradigm for measuring affect and attitude," Psychological Bulletin 76: 349-364, 1971.

Keeler, 1934.

Keeler, L., "Debunking the 'lie detector.'" The American Journal of Police Science XXV: 153-159, 1934.

Kircher, J.C., & Raskin, D.C., "Computerized decision making in the detection of deception," Psychophysiology 18: 204-205, 1981.

Kircher, J.C., & Raskin, D.C., Clinical versus statistical lie detection revisited: Through a lens sharply. Unpublished manuscript, Department of Psychology, University of Utah, Salt Lake City, Utah, 1983.

Kleinmuntz, B., & Szucko, B., On the fallibility of lie detection. Law and Society Review, 17: 84-104, 1981.

Kubis, J.F., Studies in lie detection: Computer feasibility considerations (Contract AF-30-(602)-2270. Project No. 5534) U.S. Air Force, Arlington, VA., 1962.

Kubis, J.F., Analysis of polygraphic data. Technical Report No. RADC-TDR-64-101, January 1965.

Kugelmass, 1967.

Kugelmass, S., Lieblich, I., & Bergman, Z. The role of "lying" in psychophysiological detection. Psychophysiology, 3: 312-315, 1967.

Kugelmass, S., & Lieblich, I., "The relation between ethnic origin and GSR reactivity in psychophysiological detection." Journal of Applied Psychology 52: 158-162, 1968.

Kugelmass, S., Lieblich, I., Ben-Ishai, A., Opatowski, A., & Kaplan, M., "Experimental evaluation of galvanic skin response and blood pressure change indices during criminal interrogation," The Journal of Criminal, Criminology and Police Science 59: 632-635, 1968.

Larson, J.A., Lying and its detection. Chicago: University of Chicago Press, 1932.

Lahri, S.K. & Ganguly, A.K., "Experimental study of the accuracy of polygraph technique in diagnosis of deception with volunteer and criminal subjects," Polygraph 7: 89-94, 1978.

Larson, 1932.

Larson, J.A., "Deception tests and lie detectors," Journal of Criminal Law and Criminology 12: 390-399, November 1921.

Lazarus, R.S., Psychological stress and the coping process. New York: McGraw-Hill, 1966.

Lykken, D.T., "The GSR in the detection of guilt," Journal of Applied Psychology 43: 385-388, 1959.

Lykken, D.T., "The validity of the guilty knowledge technique: The effects of faking," Journal of Applied Psychology 44: 258-262, 1960.

Lykken, D.T., "Psychology and the lie detector industry," American Psychologist, 29: 725-739, 1974.

Lykken, D.T. A tremor in the blood: Uses and abuses of the lie detector. New York: McGraw Hill, 1981a.

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

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

Lyon, V.W., "Deception tests with juvenile delinquents," Journal of Genetic Psychology 48: 494-497, 1936.

Macy, John W., Jr. (Chairman, U.S. Civil Service Commission), letter to Honorable John E. Moss, Chairman, Foreign Operations and Government Information Subcommittee, House Committee on Government Operations dated November 22, 1965.

MacNitt, R.D., In defense of the electrodermal response and cardiac amplitude as measures of deception. Journal of Law and Criminology, 33: 266-275, 1942.

Marston, W.M., "Systolic blood pressure changes in deception," Journal of Experimental Psychology, 2: 143-163, 1917.

Matte, J.A., The art and science of the polygraph technique. Illinois: Charles Thomas, 1980.

McEvoy, J.P., "The lie detector goes into business," The Reader's Digest, pp. 69-72, January 1941.

Minchew, 1979.

Monahan, J., The clinical prediction of violent behavior. U.S. Government Printing Office, 1981.

More, H.W., "Polygraph research and the university," Law and Order 14: 73-78, 1966.

Moroney, W.F., & Zenhausern, R.J., "Detection of deception as a function of galvanic skin response recording methodology," The Journal of Psychology, 80: 255-261, 1972.

Mullenix, P.A., & Reid, J.A., "The pretest interview and its role in the detection of deception," Polygraph 9: 74-85, 1980.

Orne, M.T., "Implications of laboratory research for the detection of deception," In N. Ansley (Ed.), Legal admissibility of the polygraph, (Springfield, Illinois: Charles C. Thomas, 1975): 94-119.

Orne, M.T., Thackray, R.I., & Paskewitz, D.A., "On the detection of deception: A method for the study of the physiological effects of psychological stimuli," In N. Greenfield & R. Sternbach (Eds.), Handbook of psychophysiology (pp. 743-785). New York: Holt, Rinehart and Winston, 1972.

Peters, R.B., "A survey of polygraph evidence in criminal trials," American Bar Association Journal, 68: 162-165, 1982.

Pillemer, D.B., & Light, R.J., "Benefiting from variation in study outcomes," In R. Rosethal (Ed.), New directions for methodology of social behavior science: Quantitative assessment of research domains. San Francisco: Jossey-Bass, 1980.

Podlesney, J.A., & Raskin, D.C., "Physiological measures and the detection of deception," Psychological Bulletin, 84: 782-799, 1977.

Podlesney, J.A., & Raskin, D.C., "Effectiveness of techniques and physiological measures in the detection of deception," Psychophysiology, 15: 344-358, 1978.

Privacy Protection Study Commission. (July 1977). Personal privacy in an information society (Catalog No. Y3.P93/5:1:1/9777). Washington, D.C.: U.S. Government Printing Office.

Pyle, C.H., "The defense department's regulations of 1982," Statement before the Subcommittee on Civil and Constitutional Rights, 1-25. (More info not right). (1982).

Pyle, C.H., The defense department's polygraph regulations of 1982. Statement Before the Subcommittee on Civil and Constitutional Rights, Committee on the Judiciary U.S. House of Representatives, 1-25, 9, 1982.

Quigley-Fernandez, B., & Tedeschi, J.T., "The bogus pipeline as lie detector: Two validity studies," Journal of Personality and Social Psychology, 36: 247-256, 1978.

Raskin, D.C., personal communicatio, August 23, 1983.

Raskin, D.C., Reliability of chart interpretation and sources of errors in polygraph examinations (Report No. 76-3, Contract 75-NI-99-0001, National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice, University of Utah), 1976.

Raskin, D.C., "Scientific assessment of the accuracy of detection of deception," Psychophysiology 15: 143-147, 1978.

Raskin, D.C., "Science, competence, and polygraph techniques," Criminal Defense, 8: 11-18, 1981.

Raskin, D.C., The scientific basis of polygraph techniques and their uses in the judicial process. Proceedings of the Stockholm Symposium on Witness Psychology, p. 319-371, 1982.

Raskin, D.C., & Hare, R.D., "Psychopathy and detection of deception in a prison population," Psychophysiology, 15: 126-136, 1978.

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

Reid, J.E. & Inbau, F.E., Truth and deception - The polygraph technique (3rd ed.). Maryland; Wilkins, 1977.

Robbins, N.E., & Penley, W.J., Review of polygraph charts on non-deceptive subjects, Polygraph, 4:199-207, 1975.

Rosenthal, R., Experimenter effects in behavioral research (enlarged ed.) New York: Irvington, 1976.

Rosenthal, R., & Rubin, D.B., "Comparing effect sizes of independent studies," Psychological Bulletin 2:500 -504, 1982.

Rovner, L.I., Raskin, D.C., & Kircher, J.C. Effects of information and practice on detection of deception. Paper presented at the meeting of the Society for Psychophysiological Research, Wisconsin, 1979.

Sackett, P.R., & Decker, P.J., "Detection of deception in the employment context: A review and critical analysis," Personnel Psychology, 32: 487-507, 1979.

Saks, M.J., & Van Duizend, R., The use of scientific and technological evidence in litigation: Report and indexed bibliography, 1983.

Sarbin, T.R., and Slagle, R.W., Hypnosis and psychophysiological outcomes. In E. Fromm & R.E. Shor (eds.) Hypnosis: Developments in research and perspectives. New York: Aldine, 1979.

Saxe, L., & Fine, M., Social experiments methods for design and evaluation. Beverly Hills: Sage, 1981.

Schroeder, J. (1983). Zone comparison test construction (ZCT) summary sheet. Unpublished manuscript.

Senese, L., "Accuracy of the polygraph technique with and without card test stimulation," Journal of Police Science and Administration, 4: 274-276, 1976.

Sigall, H., & Page, R., "Current stereotypes: A little fading, a little faking." Journal of Personality and Social Psychology, 18: 247-255, 1971.

Signall, H., & Page, R., Reducing attenuation in the expresison of interpersonal affect via the bogus pipeline. Sociometry, 35: 629-642, 1972.

Simon, M.A., Shall we ask the lie detector? Science, Technology & Human Values, 8: 3-13, 1983.

Skolnick, J.H., "Scientific theory and scientific evidence: An analysis of lie detection," Yale Law Journal, 70: 694-728, 1961.

Slowik, S.M., "Ethics in pre-employment polygraph examinations," Polygraph, 8: 144-149, 1979.

Slowik, S.M., & Buckley, J.P., "Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings," Journal of Police Science and Administration, 3: 305-309, 1975.

Smith, E.S., "Compilation of state and federal privacy laws," Privacy Journal, 1: 79, 1981.

Smith, E.S., "Supplement to compilation of state and federal privacy laws," Privacy Journal, 1982, 1983.

Sternback, R.A., & Turskey, B., "Ethnic differences among housewives on psychophysiological and skin potential responses to electric shock," Psychophysiology, 1: 241-246, 1965.

Szucko, J.J., personal communication, 1983.

Szucko, J.J., & Kleinmuntz, B., "Statistical versus clinical lie detection," American Psychologist, 36: 488-496, 1981.

Thackray, R.I., & Orne, M.T., "Effects of the type of stimulus employed and the level of subject awareness on the detection of deception," Journal of Applied Psychology, 52: 234-239, 1968.

Timm, H.W., "The effect of placebos and feedback on the detection of deception." Unpublished doctoral dissertation, College of Social Science, Michigan State University, 1979.

Timm, H.W., "Analyzing deception from respiration patterns," Journal of Political Science and Administration, 10: 47-51, 1982a.

Timm, H.W., "Effect of altered outcome expectancies stemming from placebo and feedback treatment on the validity of the guilty knowledge technique," Journal of Applied Psychology, 4; 391-400: 1982b.

U.S. Central Intelligence Agency, Director of Central Intelligence. Investigative scope and adjudicative procedures among intelligence community agencies. Personnel Security Survey, 1980.

U.S. Civil Service Commission. Letter to The President dated July 29, 1966.

U.S. Congress House Committee on Government and Foreign Operations. Use of polygraphs as "lie detectors" by the Federal Government. Subcommittee on Government Operations (89th Congress, 1st Session. House Report No. 198.) Washington, D.C.: U.S. Government Printing Office, 1965.

U.S. Congress, House Committee on Government Operations and Foreign Operations Subcommittee on Government Operations. Use of polygraph as lie detectors by the Federal Government (Part 2). 89th Congress, 2nd session. House Report No. 280V. Washington, D.C.: U.S. Government Printing Office, 1966.

U.S. Congress, Senate Committee on the Judiciary. Privacy, polygraphs, and

employment. A study prepared by the Staff of the Subcommittee on Constitutional Rights. Washington, D.C.: Government Printing Office, 1974a.

U.S. Congress. House Committee on Government Operations. The Use of Polygraph and Similar Devices by Federal Agencies. Hearings before Foreign Operations and Government Information, 93rd Congress, 2nd Session. Washington, D.C.: Government Printing Office, 1974b.

U.S. Congress, House Committee on Government Operations 1976. The Use of Polygraphs and Similar Devices by Federal Agencies. Thirteenth Report by the Committee on Government Operations Together with Separate and Dissenting Views. 94th Congress, 2nd Session. Washington, D.C.: Government Printing Office, 1976.

U.S. Congress Senate Committee on the Judiciary. Polygraph control and civil liberties protection act. Hearings before the Subcommittee on the Constitution. Washington, D.C.: Government Printing Office, 1978.

U.S. Congress, House Permanent Select Committee on Intelligence, Subcommittee on Oversight. Security clearance procedures in the intelligence agencies. Washington, D.C.: U.S. Government Printing Office, 1979b.

U.S. Congress. House Committee on Education and Labor. Pressures in today's workplace, vol. III. Hearings before the Subcommittee on Labor-Management Relations, December 15, 1979, Washington, D.C.: U.S. Government Printing Office, 1980a.

U.S. Congress. (1980b). House Permanent Select Committee on Intelligence on Security Clearance Procedures Pre-employment security procedures of the intelligence agencies. Hearings before the Subcommittee on Oversight. May 16, 17, 24, June 21, 1979, Washington, D.C.: Government Printing Office, 1980b.

U.S. Department of Commerce. The cost of crimes against business, 1976.

U.S. Department of Defense. Department of Defense Directive Number 5210.48, Conduct of Polygraph Examinations and the Selection, Training and Supervision of DOD Polygraph Examiners, July 13, 1965.

U.S. Department of Defense. Department of Defense Directive Number 5210.48, Polygraph Examinations and Examiners, October 6, 1975 (1/14/77).

U.S. Department of Defense, Department of the Army. Law Enforcement Investigations. Army Field Manual 19-20. Washington, D.C.: Author. April 1977.

U.S. Department of Defense. "Select Panel Review of the Department of Defense Personnel Security Program" April 16, 1982.

U.S. Department of Defense. Department of Defense Polygraph Program. DOD Directive, June 1983.

U.S. Department of Defense, Department of the Army (1983). The validity of polygraph testing. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Defense. Present Status of DOD Research on the Polygraph Report of the DOD Joint services Group on a coordinated R & D Program of Lie Detection Research, August 28, 1968.

U.S. Department of Justice, Background briefer on Presidential Directive (1983a).

U.S. Department of Justice. Presidential directive on safeguarding national security information. 1983b.

U.S. Department of the Treasury, Response to OTA survey (1983).

U.S. National Security Agency, no date.

U.S. Office of Personnel Management. Federal Personnel Manual System. Chapter 736. Appendix D. Use of the polygraph in Personnel Investigations of Competitive Service Applicants and Appointees to Competitive Service Positions. Washington, D.C.: U.S. Office of Personnel Management, 1973.

U.S. v. Brown, 557 F.2d 541 (cir. 1977).

U.S. v. Stifel, 433 F.2d 431 (6th Cir. 1970)

U.S. v. Williams, 583 F.2d (2nd Cir. 1978).

"Uproar over news leaks and lie detectors." U.S. News & World Report, 54: 8, August 15, 1963.

Waid, W.M., "Skin conductance response to both signaled and unsignaled noxious stimulation predicts level of socialization," Journal of Personality and Social Psychology, 34, 923-929, 1976.

Waid, W.M. & Orne, M.T., "Cognitive, social, and personality processes in the physiological detection of deception," Advances in Experimental Social Psychology, 14: 61-106, 1981.

Waid, W.M. & Orne, M.T., "The physiological detection of deception." American Scientist, 70: 402-409, 1982.

Waid, W.M., Orne, E.C., Cook, M.R., & Orne, M.T., "Effects of attention, as indexed by subsequent memory, on electrodermal detection of information," Journal of Applied Psychology, 63: 728-733, 1978.

Waid, W.M., Orne, E.C., Cook, M.R., & Orne, M.T., "Meprobamate reduces accuracy of physiological detection of deception," Science, 212: 71-73, 1981.

Waid, W.M., Orne, M.T., & Wilson, S.K., "Effects of level of socialization on electrodermal detection of deception," Psychophysiology, 16: 15-22, 1979a.

Waid, W.M., Orne, M.T., & Wilson, S.K., "Socialization, awareness and the electrodermal response to deception and self-disclosure." Journal of Abnormal Psychology, 88: 663-666, 1979b.

Waid, W.M., Wilson, S.K., & Orne, M.T., "Cross-model physiological effects of electrodermal lability in the detection of deception." Journal of Personality and Social Psychology, 40: 1118-1125, 1981

Weir, R.J., In defense of the relevant-irrelevant polygraph test. Polygraph 3: 119-166, 1974.

Weir, R.J., "Stimulation procedures - A conservative view." Polygraph, 7: 209-214, 1978.

Weir, R.J., & Atwood, W.F., Applicant screening polygraph examinations. Polygraph 10: 129-142, 1981.

Whiteside, G., Profits vs. Polygraph. Phoenix, AZ: Associated Grocers, 1979.

Wicklander, D.E., & Hunter, F.L., "The influence of auxiliary sources of information in polygraph diagnoses." Journal of Police Science and Administration, 3: 405-409, 1975.

Widacki, J., "Analiza przestanek diagnozowania w badanich poligraficznych (The analysis of diagnostic premises in polygraph examinations.) Katowice: Uniwersytetu Slaskiego, 1982.

Widacki, J., & Horvath, F., "An experimental investigation of the relative validity and utility of the polygraph technique and three other common methods of criminal identification." Journal of Forensic Sciences. 23: 596-601, 1979.

Ziskin, J., Coping with Psychiatric and Psychological Testimony. Venice, California: Law and Psychology Press, 1982.

* * * * * *

POLYGRAPH VALIDITY ADVISORY PANEL

Christopher H. Pyle, Associate Professor of Politics, Mt. Holyoke College
David C. Raskin, Professor of Psychology, University of Utah
Harold Sigall, Professor of Psychology, University of Maryland
George B. Trubow, Professor of Information Law and Policy, The John Marshall Law School
Althea M.I. Wagman, Research Associate of Psychiatry, Maryland Psychiatric Research Center, University of Maryland School of Medicine

OTA POLYGRAPH VALIDITY PROJECT STAFF

John H. Gibbons, Acting Assistant Director, OTA Health and Life Sciences Division(Since Sept. 1983; David Banta served as Associate Director until August 1983).
Clyde B. Wood, Project Director, (Communications and Information Technologies Program)
Leonard Saxe, Principal Investigator and Author (Boston University)
Denise Dougherty, Co-author and Analyst (Health)(OTA Analyst since September 1983; Boston University Research Assistant until August 1983).
Theodore Cross, Co-author (Boston University).
Jack Langenbrunner, Analyst (Health)
Katherine Locke, Research Assistant (Health)

Administrative Support:

Ginny Cwalina, Administrative Assistant (Health)
Elizabeth Emanuel, Administrative Assistant (CIT)
Jennifer Nelson, Secretary (Health)
Shirley Gayheart, Secretary (CIT)
Winky Marlar, Secretary (Health)
Marie Calabrese, Secretary (Boston University)
Molly Zane, Secretary (Boston University)

Other Contributors

Michael Saks, Boston College
Daniel Ozer, Boston University
Yoram Bar-Tal, Boston University
Mary Beasley, Boston University

* * * * *