

Chapter 7

Conclusions

Conclusions

INTRODUCTION

The primary purpose of this technical memorandum is to evaluate the scientific evidence on the validity of polygraph tests. The memorandum responds to concerns of the Committee on Government Operations, U.S. House of Representatives, about significant changes in Federal Government policy concerning polygraph testing. As discussed in chapters 1 and 3, National Security Decision Directive 84 (*NSDD-84*), issued by the President on March 11, 1983, authorized executive agencies and departments to require employees to take a polygraph examination in the course of investigations of unauthorized disclosures of classified information. On October 19, 1983, the Department of Justice announced that administration policy would also permit Government-wide polygraph use in preemployment, preclearance, periodic, and aperiodic personnel security screening of employees with access to highly classified information. Draft proposed revisions to Department of Defense (DOD) polygraph regulations (DOD 5210.48) would also authorize the expanded use of polygraph testing as part of personnel security screening of employees with highly sensitive access.

The combined effect of these changes is to authorize substantially expanded use of polygraph

examinations by the Federal Government for investigations of specific incidents (i. e., unauthorized disclosures), and, most significantly, for personnel security screening. In addition, *NSDD-84*, administration policy, and the DOD proposals authorize adverse consequences for refusal to take a polygraph examination.

By letter of February 3, 1983, the Committee on Government Operations asked OTA to assess the scientific evidence on the validity of polygraph testing, based primarily on a critical review and evaluation of existing research. In order to conduct this assessment, OTA studied the actual polygraph examination process, reviewed the results of prior research reviews, analyzed a wide range of relevant field and analog studies, and surveyed Federal agencies as to their polygraph use and any past, present, or planned polygraph research. This chapter highlights the overall scientific conclusions of the OTA evaluation and then discusses in some detail specific scientific conclusions and the implications for recent and proposed changes in Federal policy on polygraph testing.

OVERALL SCIENTIFIC CONCLUSIONS

OTA concluded that, as shown in chapter 2, polygraph testing is, in reality, a very complex process that varies widely in application. Although 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, polygraph tests require 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.

One general type of polygraph question technique (called the control question technique) is commonly used for investigations of specific criminal incidents and has received most of the research attention. Another technique (known as relevant/irrelevant) typically used for preemployment screening and periodic screening purposes has been only minimally researched. Based on a

detailed review of these and other question techniques in chapter 2, OTA concluded that there are significant differences, and that the results of research on one technique cannot be generalized to other techniques. Also, differences between techniques are so significant that the results of research on one technique in one application cannot necessarily be extrapolated to other applications. Chapter 2 also reviewed the Federal Government's use of polygraph testing and found that, with the exception of the National Security Agency (NSA) and Central Intelligence Agency (CIA), most current use, even in DOD, is for investigation of specific crimes using the control question technique.

In chapter 3, OTA reviewed the legal, governmental, and scientific controversies over polygraph testing. OTA found that previous debates at the Federal level have focused heavily on whether polygraph testing is scientifically valid. The conclusion of previous congressional inquiries has been that there is little or no scientific basis for the use of polygraph testing. Prior scientific reviews, on the other hand, have contradicted each other, some concluding that polygraph testing is almost 100 percent accurate, others that it is little better than chance. OTA determined that part of the problem in reaching conclusions about polygraph testing validity is that several scientific criteria must be taken into account when assessing validity. Also, previous scientific reviews have not been conducted systematically. In addition, previous reviews, whether legal, governmental, or scientific, have not differentiated polygraph use by type of question technique or application.

OTA conducted its own systematic review of prior research studies on the validity of polygraph testing (see ch. 4 for discussion of field studies of actual polygraph examinations and ch. 5 for discussion of analog or simulation studies). OTA found that there are almost no studies relevant to proposed Federal Government expansion of polygraph testing for preemployment, periodic, or aperiodic screening. This finding has major policy implications discussed later. OTA also found that, even among the rather extensive studies of the control question technique in criminal investigations, there is a wide range of accuracy (and thus, inconclusive and error) rates. OTA con-

cluded that this accuracy range could be partially explained by variations in research design but perhaps to a greater extent, as is discussed in chapter 6, by differences in examiners, examinees, question techniques, and conditions of testing.

OTA concluded, therefore, 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 design and conduct of specific studies and the particular application researched. Some applications (e.g., the use of the polygraph in criminal investigations) have been fairly heavily researched, while others (e.g., polygraph use in preemployment screening) have had very little research attention.

Further, regardless of whether polygraph testing is used in specific-incident investigations or personnel screening, OTA concluded that polygraph accuracy may also be affected by a number of factors: examiner training, orientation, and experience; examinee characteristics such as emotional stability and intelligence; and, in particular, the use of countermeasures and the willingness of the examinee to be tested. In addition, the basic theory (or theories) of how the polygraph test actually works has been only minimally developed and researched.

In sum, OTA concluded that there is at present only limited scientific evidence for establishing the validity of polygraph testing. Even where the evidence seems to indicate that polygraph testing detects deceptive subjects better than chance (when using the control question technique in specific-incident criminal investigations), significant error rates are possible, and examiner and examinee differences and the use of countermeasures may further affect validity.

More specific scientific conclusions and the implications for recent and proposed changes in Federal policy on polygraph testing are presented below. The discussion is organized in terms of conclusions and implications, first, for specific-incident investigations and personnel security screening use of the polygraph; second, for polygraph countermeasures and for the voluntary nature of testing; and finally, for further research.

SPECIFIC SCIENTIFIC CONCLUSIONS IN POLICY CONTEXT

Specific-Incident Criminal Investigations

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

The application of the polygraph to specific-incident criminal investigations is the only one to be extensively researched. OTA identified 6 prior reviews of such research (summarized in ch. 3), as well as 10 field and 14 analog studies that met minimum scientific standards and were conducted using the control question technique (the most common technique used in criminal investigations; see chs. 2, 3, and 4). Still, even though meeting minimal scientific standards, many of these research studies had various methodological problems that reduce the extent to which results can be generalized. The cases and examiners were often sampled selectively rather than randomly. For field studies, the criteria for actual guilt or innocence varied and in some studies were inadequate. In addition, only some versions of the control question technique have been researched, and the effect of different types of examiners, subjects, settings, and countermeasures has not been systematically explored.

Nonetheless, this research is the best available source of evidence on which to evaluate the scientific validity of the polygraph for specific-incident criminal investigations. The results (for research on the control question technique in specific-incident criminal investigations) are summarized below:

- Six prior reviews of field studies:
 - average accuracy ranged from 64 to 98 percent.

- Ten individual field studies:
 - correct guilty detections ranged from 70.6 to 98.6 percent and averaged 86.3 percent;
 - correct innocent detections ranged from 12.5 to 94.1 percent and averaged 76 percent;
 - false positive rate (innocent persons found deceptive) ranged from 0 to 75 percent and averaged 19.1 percent; and
 - false negative rate (guilty persons found nondeceptive) ranged from 0 to 29.4 percent and averaged 10.2 percent.
- Fourteen individual analog studies:
 - correct guilty detections ranged from 35.4 to 100 percent and averaged 63.7 percent;
 - correct innocent detections ranged from 32 to 91 percent and averaged 57.9 percent;
 - false positives ranged from 2 to 50.7 percent and averaged 14.1 percent; and
 - false negatives ranged from 0 to 28.7 percent and averaged 10.4 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. 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 error rates that could be considered significant.

The figures presented above are strictly ranges or averages for groups of research studies. Another selection of studies would yield different results, although OTA'S selection represents the set of studies that met minimum scientific criteria. Also, some researchers exclude inconclusive results in calculating *accuracy* rates. OTA elected to include the inconclusive 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 inconclusive would raise the overall accuracy rates calculated. In practice, inconclusive results may be followed by a re-test or other investigations.

Relevance to NSDD-84 and Administration Policy

While the results of the OTA review indicate that the control question technique has some validity in criminal investigations, there is only a limited scientific basis for generalizing the results of the OTA review to the context of NSDD-84 and the October 19, 1983, administration policy on polygraph use. NSDD-84 and administration policy authorize the use of the polygraph in administrative as well as criminal investigations of unauthorized disclosures of classified information.

First, there is no validity research directly on the use of the polygraph in unauthorized disclosure investigations. The subject matter and perhaps subjects of these investigations will vary from the typical criminal investigation as might the conditions and techniques of testing and use of countermeasures.

Second, the investigative conditions authorized by NSDD-84 and administration policy may be quite different from conditions under which prior research was conducted. NSDD-84 does not specify what type of investigative procedures will be followed, how subjects will be selected or identified, who will conduct the examinations, or what question techniques will be used. Administration policy provides some specific guidelines such as requiring that polygraph testing be used only when "other information or means of investigation have produced a substantial objective basis for seeking to examine the employee" and there is "no other reasonable means of resolving the matter" (185a). However, in general, the extent to which employees will be requested or required to take polygraph examinations in unauthorized disclosure investigations is largely left to the discretion of agency heads.

Third, even the Federal Bureau of Investigation (FBI) has concluded that, "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." The FBI is planning its own validity research.

On the other hand, to the extent polygraph use in unauthorized disclosure investigations is similar

to the way the polygraph is used in criminal investigations, there is at least some although far from conclusive scientific basis for polygraph validity,

Large-Scale Screening

The polygraph test is used by some private firms and on rare occasions by some Federal agencies to screen a large number of people in connection with the investigation of a crime. Unlike the typical specific-incident criminal investigation, in a large-scale screening investigation, typically the suspect list has not been narrowed down to one or a few persons and only limited information about the crime is available.

NSDD-84 appears to permit such use of the polygraph in unauthorized disclosure investigations, although the actual extent of NSDD-84 is unclear. Administration policy appears to be ambivalent. While on the one hand providing guidelines for "carefully limited use of the polygraph," the policy implies that DOD polygraph regulations are acceptable. DOD regulations have been used, albeit infrequently, to authorize polygraph screening of large numbers of individuals (ranging from about 2 dozen up to 80) in investigation of specific incidents.

There is no scientific basis for generalizing the results of the OTA review to establish polygraph validity in this large-scale screening application. First, no scientifically acceptable research has been conducted on large-scale specific-incident screening use of the polygraph. Second, the screening conditions here are likely to vary even more from the conditions of the research studies reviewed by OTA. For one thing, much less information is likely to be known about circumstances surrounding an unauthorized disclosure and possible suspects. 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 use of polygraph testing theoretically can be expected to result in significantly higher error rates than when the list of suspects is narrowed down to a very small number, as in a typical criminal investigation. The screening use of polygraph tests is most dependent on the so-called base rate of guilt, i.e.,

the percentage 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, two persons would be incorrectly identified (false positives).

These are, of course, hypothetical examples, and have not been systematically investigated in either field or analog research, although some reviewers (e. g., Ben-Shakhar (28)) have carefully worked through a number of possibilities. Also, operating procedures of Federal agencies (e.g, quality control review, consideration of other investigatory information) might catch, correct, or minimize erroneous polygraph decisions.

Nonetheless, the FBI, which outside of DOD and CIA, is the principal Federal agency that conducts polygraph examinations, believes that large-scale screening is not an appropriate use of polygraph testing. FBI regulations prohibit 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],

Personnel Security Screening

Draft revisions to the DOD polygraph regulations would authorize the use of polygraph tests to determine initial and continuing eligibility of DOD civilian, military, and contractor person-

nel for access to highly classified information (Sensitive Compartmented Information and/or special access). The use of polygraph tests to determine continuing eligibility would be on an aperiodic (i.e., irregular) basis (181). These are all known as personnel security applications of the polygraph. In addition, administration policy announced on October 19, 1983, would permit Government-wide use of polygraph tests in personnel security screening of employees (and applicants for positions) with access to highly classified information. The new policy provides agency heads with the authority to give polygraph examinations on a periodic or aperiodic basis to employees with highly sensitive access.

Results of OTA Review

Personnel security screening involves a different type of polygraph test than specific-incident investigations, and very little screening research has been conducted. Three studies were cited by the intelligence agencies (NSA and CIA) as providing support for personnel security use of polygraph tests.

A 1975 field study (6) of polygraph screening of government job applicants (from an unidentified Federal agency) showed high consistency in readings of physiological arousal by different examiners. But this study concluded nothing about validity.

In a 1981 analog study (43) of preemployment 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, the questions and context had nothing to do with national security, and the test format was atypical of personnel screening examinations.

A 1980 survey conducted by the Director of the 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. For example, the criteria for case selection were not stated and there was no independent verification of the cases that were resolved. Also,

the polygraph was used only after a thorough investigation based on other sources had taken place (see ch. 4 for further discussion).

OTA inquiries to all DOD components using the polygraph identified only one DOD research study on personnel screening use of the polygraph (16). 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. OTA inquiries to other executive agencies and departments using the polygraph identified no research on personnel security screening use of the polygraph.

OTA recognizes that the administration as well as NSA, CIA, and DOD 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, CIA agreed that the cumulative unclassified research evidence reviewed by OTA is not directly relevant to national security applications. However, CIA does claim to have classified research to support their use of polygraph tests. 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 polygraph screening situation might involve a base rate of one guilty person (e.g., one person engaging in unauthorized disclosure) out of 1,000 employees. Assuming that the polygraph is 95 percent valid, then, the one guilty person would be identified as deceptive but so would 50 innocent persons. The predictive validity would be about 2 percent. Even if 99 per-

cent 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 systematically studied in field or analog research. NSA claims that they in fact have experienced a very low false positive rate and that, in any event, polygraph test results are only one factor in making decisions and are subject to quality control checks and other reviews. It appears that NSA (and possibly 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 v. 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 pretest or posttest phase of the examination” (187).

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 (e. g., with respect to type of questions asked and question techniques employed) that results of the OTA research reviewed previously discussed cannot be generalized to the NSA situation.

False Negatives/Countermeasures

The primary purpose of polygraph testing under NSDD-84, the DOD revised regulations, and administration policy 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). A concern with false negatives (guilty persons incorrectly identified as nondeceptive) 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.

Theoretically, polygraph testing—whether for personnel security screening or specific-incident investigations—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 polygraph countermeasures has been limited and the results—while conflicting—suggest that validity may be affected. Further, some research (e.g., 75) suggests that polygraph examiners may not be able to easily detect certain physical countermeasures. The research results for drug and psychological countermeasures are mixed. The possible effects of countermeasures are particularly significant to the extent that the polygraph is used and relied on for national security purposes, since even a small false negative rate could have serious consequences. In addition, those individuals who the Federal Government would most want to detect (e.g., for national security violations) may well be the most motivated and perhaps the best trained to avoid detection.

Voluntary v. Involuntary

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, CIA) and, under certain 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” (187). 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.”

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 nonelection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position. *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. And administration policy authorizes denial of clearance, transfer or reassignment, and, under some circumstances, termination of employment for refusal to take a polygraph test.

Under these conditions, 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, which OTA did not address, conducting polygraph tests on this basis could affect test validity. It is generally recognized that, for the polygraph test to be accurate, the voluntary cooperation of the individual is important. For example, NSA has stated that, in conducting screening examinations, “[t]he full cooperation of the individual taking the test is essential or the results will be inconclusive.” The polygraph only detects physiological arousal, and under involuntary conditions, the arousal response of the examinee may be *very* difficult or impossible to interpret. However, no direct research on this topic was identified. Overall, OTA concluded that imposing penalties for not taking a test may create a de facto involuntary condition that increases the chances of invalid or inconclusive test results.

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.

Polygraph Theory

The basic theory of polygraph testing is only partially developed and researched. 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. However, this theory has been challenged by some psycholo-

gists and others who believe that various factors—e.g., the examinee's intelligence level, psychological health, emotional stability, and belief in the "machine"—may, at least theoretically, affect the physiological response.

OTA concluded that a stronger theoretical base is needed for the entire range of polygraph applications, including current and proposed Federal Government 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.

Criminal Investigation Validity

There are still many unanswered questions about the validity of use of the polygraph in specific-incident criminal investigations. A planned FBI-Secret Service validity study is intended to meet this need. However, OTA did not review the research plan, which would benefit from an independent review by the scientific community and others before the research approach is finalized. Such a review would help ensure that the

research design is as scientifically sound as possible. Also, the U.S. 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 by the Army research until the late 1980's.

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.

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 U.S. Army Intelligence and Security Command 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.

CONCLUDING COMMENT

A major reason why scientific debate over polygraph validity yields conflicting conclusions is that the validity of such a complex procedure 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 on polygraph testing is hard to design and conduct.

Advocates 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 technical memorandum is 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 typical criminal investigations of specific incidents, and more limited evidence when such investigations extend to incidents of unauthorized disclosure. However, there is very little research or scientific evidence to establish polygraph test validity in large-scale screening as part of unauthorized disclosure investigations, or in personnel security screening situations, whether they be preemployment, preclearance, periodic or aperiodic, 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*, DOD, and administration polygraph proposals.