Polygraph

VOLUME 27 1998 NUMBER 4

Contents

Published Quarterly

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

A Critical Analysis of Matte's Analysis of the Directed Lie

Charles R. Honts and Anne Gordon

Abstract

Matte (1998) reported that the directed-lie control (DLC) question test lacks empirical validation and fails to achieve an adequate level of construct validity. In the present article, it is argued that Matte's presentation was selective and biased. A thorough examination of the empirical literature reveals that there is ample evidence in support of the criterion validity of the DLC. Moreover, a critical examination of Matte's discussion of construct validity finds that his concerns relate more to the face validity than the construct validity of the DLC. Yet, even in that analysis many of Matte's assertions are shown by research to be untenable. Therefore, it is concluded that the DLC question test is a valid test for the psychophysiological detection of deception, and because of its many advantages in standardization should be considered for application in field settings.

Matte (1998) provided "an analysis of the psychodynamics of the directed lie control question" (p. 56). His review included a number of speculations concerning subjects' perceptions of the directed lie control (DLC) question test but included no new data. Among Matte's conclusions were the following:

1. The DLC question test lacks construct validity because it fails to demonstrate an adequate capacity to function as a control question of less intensity than the relevant question for the guilty examinee.

2. The DLC question test lacks criterion validity (accuracy).

3. There is an insufficient body of data to support the use of the DLC question test in application.

The purpose of this paper is to demonstrate that none of Matte's conclusions are supported in the scientific literature. Matte appears to have done a major disservice to the polygraph profession by presenting an inaccurate description of the modern use of the DLC question test. His analysis was based on a selective review of the empirical literature and was accompanied by a speculative analysis that has neither a scientific nor logical basis. The main goal of the present analysis is to correct the misinformation in Matte (1998) by: (1) providing an accurate description of the modern use of the DLC question test, (2) providing a comprehensive review of the published studies on the DLC question test and (3) by considering construct validity issues relevant to the DLC question test.

The DLC Question Test

A Brief History

The DLC question test was developed within the military intelligence system of the United States Government for use in multiple test counter-espionage situations. Fuse (1982) provided the first non-classified document describing the use of the DLC question test. His paper described how the DLC question test was then being used by military intelligence. Fuse noted that a general concern in testing in such situations was that adequate probable lie comparison questions could not be developed. Moreover, the repeated use of the same probable-lie questions was likely to result in their habituation (i.e., decreasing responses over presentations) and decreased effectiveness. Therefore, a directed-lie question, one to which subjects were instructed to lie, was developed to provide a type of comparison

Both authors are with Boise State University. Correspondence concerning this article should be addressed to: Charles R. Honts, Ph.D., Department of Psychology, Boise State University, 1910 University Drive, Boise, Idaho 83725, USA. Email to honts@truth.idbsu.edu.

question that could be used repeatedly with the same subject without the encountering the aforementioned problems. The psychological set of the directed-lie question was created in the context of an unknown number test. Subjects were told that anytime a person lies, physiological responses are produced and that those responses can be recorded with the polygraph. (In the original formulation, the successful detection of deception in the number test was required before the examination proceeded) Following review of the relevant questions of the examination (up to 8 in the original formulation), the DLC was presented in the following manner:

> The examinee is told that the purpose of the DLCs is to assure that his reaction capability has not significantly changed from that exhibited during the ACQ [number test]. The importance of the DLCs is emphasized by stating that they are used to assure that examinee still responses normally when attempting deception and when answering truthfully, thus the examiner avoids errors in the final interpretation of the charts. Stress that the DLCs are deliberate lies, known to both examinee and the examiner, and, as long as the examinee is responding normally, deceptive responses will appear on the test charts. (Fuse, 1982, pp. 20-21)

In the original formulation the exact question list, including question order, was revealed to the subject, and the subject was told that there would be three repetitions of the questions. After the first chart the examiner reviewed the relative responses to the questions. Then Fuse (1982) advised the following:

> If it appears that the DLCs are not generating at least some response activity, regardless of the response activity to the relevants mild interchart reinforcement may be utilized. For example, examinee may be told the test chart "looks good" and that when he lies, the responses are clear, and it is obvious that he has indeed engaged in the activities covered by the DLCs. (p. 25)

The group of researchers in David Raskin's laboratory at the University of Utah became interested in the DLC in the early 1980s, following Fuse's (1982) presentation and in light of the experiences reported by a local polygraph examiner, Larry Kelly. Kelly was conducting tests in a probation and parole setting and was facing many of the same problems with multiple tests that military intelligence had encountered. He began using the DLC questions and reported good success with it. One of us (Honts) began using the DLC in his private practice in January of 1984 when a defense-referred subject told Honts that he had been reading about polygraph tests and that he would not be able answer the control questions with a "No." Therefore, instead of using probable lie control questions, Honts used DLC questions. The subject failed the test and subsequently confessed to the crime of which he was accused. In a subsequent discussion with David Raskin we agreed that Fuse's original formulation of the DLC question test was too complex and time consuming. Moreover, we agreed that the procedure could be even more standardized than the Fuse approach allowed for.

The revisions we made to the DLC question test at the University of Utah have been described elsewhere (Honts & Raskin, 1988; Honts, Kircher, & Raskin, 1995, Horowitz, Kircher, Honts & Raskin, 1997; Raskin, Honts, & Kircher, 1997; Raskin, 1989). In short, we made the following four refinements: The pretest interview was shortened by reducing the amount of time spent on describing physiology. A known number test was conducted. All questions were reviewed with subjects, but the order was not revealed and was varied on each repetition. Finally, both relevant and DLC questions were reviewed between each chart (as was standard practice at the University of Utah for probable-lie control question tests; see Honts, 1999). The Utah induction procedure was described by Raskin as follows:

> On this test I need to ask you some questions to which I want you to lie. Just as on the number test. I need to have questions to which you and I both know you are lying. That way, I can be sure that you continue to respond appropriately when you are lying and

> that you remain a suitable subject throughout this test. Therefore, I am going to ask you. "Before age 25, did you ever tell even one lie?" and I want you to lie to that question. Also, I want you to think of a particular instance when you did lie in the past, and I want you to have that in mind when you answer this question on the test. Do you have a particular instance in mind? . . All right, I do not want you to tell me what it is. When I ask you that question on the test, I want you to lie by answering no, and when you answer, I want you to think about the time when you lied. That way, you and I will be sure that you are lying when you answer that question on the test, and I can make sure that you react appropriately and that you continue to be a suitable subject. (p. 271)

The DLC test thus uses directed lie questions to create a comparison question. Scoring of a DLC is no different than the scoring of a traditional probable lie control test. The position of the researchers at the University of Utah was that the psychological rationale for using the DLC was the same as for the traditional probable-lie control. That is, for guilty subjects the relevant questions present a powerful set of stimuli. The inherent power of the relevant questions overwhelms all other stimuli, even the comparison questions to which it assumed (probable-lie) they are lying or to comparison questions to which they have been instructed to lie. For innocent subjects, although the relevant questions are recognized as important, these subjects find the set of comparison questions more powerful, either because they are uncertain of their answers, because they are actually attempting deception (probable-lie), or because of the examiners' instructions concerning the set (DLC). Thus, the discrimination of truth tellers and deceivers depends upon the valance and magnitude of the difference between the relevant and comparison questions. More on this topic will follow in the section on construct validity.

The Positive Control Test

The positive control test (Golden, 1969) has been suggested as another alternative to the probable-lie comparison question tests (Driscoll, Honts, & Jones, 1987). Driscoll et al. described the positive control test as follows.

> The positive control test essentially asks only relevant questions, but asks each question twice with different instructions regarding the answer. The first time the question is asked the subject is instructed to admit the acts in question; this is referred to as the subjective lie question. The second time each question is asked the subject is instructed to deny the acts in question; this is referred to as the subjective truth question. The two presentations of a relevant question are referred to as a positive control pair. *The rationale of the positive control test predicts that subjects will produce larger physiological responses when they are actually lying.* [italics added] Thus, if subjects show larger physiological responses to the subjective lie, that is, when they say they committed the crime, they are interpreted as truthful with regard to the issues of the examination. When subjects show larger physiological responses to the subjective truth question, that is, when they say they did not commit the crime, they are interpreted as deceptive to the issues of the examination. (p. 218)

Thus, the methods and rationale underlying the positive control approach are dramatically different than those underlying the DLC approach. Positive control questions are not in any meaningful way similar to directed lie control questions. Despite Matte's assertions to the contrary, any similarities between the positive control approach and the DLC question approach are at best superficial and are at worst spurious. Moreover, despite Matte's assertion that the positive control test is prone to false negative errors, the research actually indicates that the positive control test is simply poor at discriminating truthful from deceptive subjects. However, in positive control tests, the examiner can manipulate the relative predominance of either false positive or false negative errors by controlling the order of presentation of the subjective truth and subjective lie responses. Research shows that the first question of the positive control pair

produces the largest response regardless of the answer (Honts & Driscoll, 1989). That outcome is consistent with the general psychophysiological principle of habituation and is thus lawful and expected, but it has nothing to do with the DLC question test. Thus, Matte's use of data from studies of the positive control approach to criticize the DLC question test is inappropriate and meaningless.

Criterion Validity of the Directed Lie Control

In their respected book on psychometric theory, Nunnally and Bernstein (1994) define criterion (predictive) validity as follows, "Predictive validity concerns using an instrument to estimate some criterion behavior that is external to the measuring instrument itself." (p.94). For example, polygraph examiners attempt to predict (diagnose) whether an individual is telling the truth or being deceptive with respect to some set of issues. Our instruments, in Nunnally and Bernstein's use of the term, are the various polygraph tests (e.g., the DLC question test) that are administered. In psychometrics, predictive validity is usually assessed by establishing the degree of correlation between the outcome on a test (pass/fail) with a criterion (truth teller/deceiver). This is the approach that was taken by the Utah group many years ago (for a discussion see, Raskin, 1989). Thus, in a psychometric analysis of the criterion validity of the various polygraph tests it is not useful to consider the accuracy rates for either innocent or guilty in isolation, as was done in Matte 1998[1]. For example, research on the relevant-irrelevant (RI) test has indicated that it is very good at detecting attempted deception, with accuracy rates of identifying guilty subjects often approaching 100% (Horowitz et al, 1997; Horvath, 1988). If only those data were considered one might think that the RI had high criterion validity. However, those same studies found that the RI correctly identified very few (~20%) of the innocent subjects. Thus from a psychometric analysis, the RI was shown to be a poor discriminator. In reanalyzing the Horowitz et al (1997) data, Honts (1994) found the predictive correlation for the RI to be only r = .38. That is, the RI was able to account for only 14% (r^2 X 100) of the variability between innocent and guilty subjects. A meaningful criterion validity analysis of the DLC question test must look at its ability to discriminate truth tellers from deceivers.

Honts and Raskin (1988)

The Utah group of researchers published two studies comparing the validity of the DLC and probable lie control questions. The first of those studies was a field study published by Honts and Raskin (1988). That study reported the results of an exhaustive sample of confirmed field cases from Honts and Raskin's private practices over a 4-year period. During that period Honts and Raskin were substituting a DLC for one probable lie question in their standard comparison question format (three relevants, three comparisons, and three neutrals; see Kircher & Raskin, 1988). The Honts and Raskin recordings were subjected to independent analysis. The independent evaluators scored the data in two ways. First, each relevant question was scored to the preceding comparison question, regardless of type. After all tests were evaluated, the evaluator went back, without reference to his previous scores, and reevaluated the data substituting the nearest (in time) probable-lie for the DLC. Decisions were then made using the rules developed at the University of Utah (Raskin & Hare, 1978). The results of the Honts and Raskin Study are illustrated in Table 1. Honts and Raskin (1988) report the criterion validity of the tests as follows: When the DLC was scored, tau c = 0.90. When the DLC was not scored, tau c = 0.81. In terms of the amount of criterion variability accounted for by the test, this represents an improvement of approximately 15% when the DLC was scored.

In discussing the Honts and Raskin (1988) study Matte (1998) repeated a criticism of that study that has often been raised by Abrams in his efforts to keep polygraph tests from being admitted as evidence in courts of

[1] Although examining the differential error rates for truth tellers and deceivers is generally not part of assessing criterion validity, it can be useful for policy analyses that consider costs and benefits surrounding the particular application (for example see Honts, 1991).

Table 1
Decisions With And Without The DLC.

	Decisions		
	Truthful	Inconclusive	Deceptive
With DLC			
Innocent	11	2	0
Guilty	1	0	11
Without DLC			
Innocent	8	3	2
Guilty	0	1	11

Adapted from Honts & Raskin 1988

law (e.g. U. S. v. Gilliard, 1996, 1998). Abrams's criticism of the Honts and Raskin study is as follows: Half of the subjects in that study were suspects in child sex abuse cases. One of the criteria used for confirming ground truth in the sexual abuse cases was the retraction of the allegation by the alleged victim. Retractions are sometimes false; therefore, the study cannot be trusted. However, even a cursory examination of the Honts and Raskin (1988) report reveals this to be a "straw man" criticism that is without merit. There were 25 confirmed cases in the Honts and Raskin study. Of those 25 cases 11 were from child sex abuse cases. Of those 11 cases 5 were guilty subjects who were confirmed by their own confessions. Thus, Abrams's criticism cannot be applied to those five cases because there was no retraction in those cases. Thus, there are only 6 potential cases to which Abrams criticism might apply. In response to Abrams's testimony, we have re-examined the case facts from those 6 cases. Abrams's criticism is applicable to only two of those cases. Even if all six of those cases were dropped from the study the central finding would not change; the DLC improved discriminability. Moreover, Abrams's statement that child sexual abuse victims frequently recant, is based on speculation and purely anecdotal evidence. No quantitative data exist that demonstrate that false recantations are any more frequent than false confessions. If we are unwilling to put weight on recantations, then we should also dismiss confessions as a useful criterion. In any event, the statistical analysis reported in the Honts and Raskin (1988) paper indicates that there were no differences between suspects in sex abuse cases and those who were suspects in other criminal matters. Abrams's (1991) criticism of Honts and Raskin (1988) thus deserves little attention

Horowitz, Kircher, Honts, & Raskin (1997)

In a complex mock-crime experiment Horowitz et al. contrasted directed lies that involved personally relevant information (e.g., Have you ever told a lie even one time in your life?) and trivial directed lies that did not involve the subject personally (e.g., Does 2 + 2 = 4?) with the probable lie test and a relevant-irrelevant test. The results of that experiment are summarized in Table 2. To assess the ability of each of the various techniques to discriminate truthful from deceptive subjects, we calculated a detection efficiency coefficient (r) for each technique. The square of the correlation coefficient can be used to give an index of the percentage of variance accounted for in the guilt criterion. The DLC test using the personal directed lies produced the greatest absolute discrimination between the innocent and guilty subjects r = 0.69. Decisions with a personal directed lie accounted

Table 2
Results of the Horowitz et al., (1997) study of comparison questions.

Technique / Guilt	Examination Outcome			Detection
	Truthful	Inconclusive	Deceptive	Efficiency r
Probable-Lie				
Innocent	12	1	2	.56
Guilty	3	4	8	
DLC - Personal				
Innocent	13	0	2	.69
Guilty	2	3	11	
DLC - Trivial				
Innocent	10	3	2	.50
Guilty	3	4	8	
Relevant-Irrelevant				
Innocent	3	1	11	.38
Guilty	0	0	15	

for 16% more of the variance in the guilt criterion than decisions with the probable lie CQT. The Personal DLC test produced one fewer false negative outcome than did the probable lie control test.

Matte (1998) has paid particular attention to the respiration data of Horowitz et al. (1997). Unfortunately, Matte's discussion of those findings was misleading and in some ways nonsensical. The data in the Horowitz et al. study were collected with laboratory instrumentation and were digitized. Features of the physiological waveforms were extracted with computer algorithms. Only one feature was extracted for respiration and that was respiration length (Timm, 1982) during the first 10 seconds following question onset. Innocent subjects in the DL conditions produced significantly longer respiration lengths than did innocent subjects in the probable lie condition. The respiratory responses of guilty subjects were as expected in all conditions with the least (but not significantly so) deceptive response being given by subjects in the probable lie condition. Since Matte's (1998) thesis was that the DLC test produces an unacceptable number of false negative outcomes, it is puzzling that he chose to highlight this particular effect in the Horowitz et al study. The Horowitz et al. respiration length findings with DLC should have increased the tendency toward false positive outcomes not false negative outcomes. Matte (1998) further confuses the issue by mixing the results of computer-derived features with those that are the result of numerical scoring. In numerical scoring many other aspects of the respiration tracing are considered. Whatever the effects with the

computer-generated features, the numerical scores given by the evaluator in the Horowitz et al. study were highly successful in discriminating innocent from guilty subjects when the DLC question test was used. The greatest separation of mean numerical scores for innocent and guilty occurred with the personal DLC question. This finding suggests that Matte's protestations about the respiration findings are without merit.

Barland (1981)

The Barland (1981) study was only briefly considered by Matte (1998), and the results were not described in detail, other than to say that they were better than chance. Barland examined the validity of the Military Intelligence version of the DLC in a mock screening setting with 26 truthful subjects and 30 subjects who attempted deception. All subjects were tested with the DLC; no other techniques were examined. The results of the Barland study are illustrated in Table 3. Excluding inconclusive outcomes, Barland's evaluators correctly classified 79% of the subjects. Barland did not calculate a correlation with the criterion statistic for these decisions, but one can be calculated from the data in Table 3. The correlation with the criterion for the DLC test in Barland was significant, tau c = 0.56, $p < .001$. Although this might be considered modest performance in comparison to that obtained in the University of Utah studies, it must be remembered that Barland's (1981) study was in a screening setting. When compared with other mock-screening studies, that have often produced near chance performance with probable lie tests (e. g. Barland, Honts, & Barger, 1989; Honts, 1992), the performance of the DLC in Barland (1981) was actually quite strong.

Reed (1994) also published as DODPI Research Staff (1995)

There are two empirical papers concerning the DLC question test that Matte did not discuss. The first of those is Reed (1994). Reed reports three laboratory mock screening studies that were part of the validation of a new national security screening technique, the Test for Espionage and Sabotage (TES). Following a series of studies that indicated that the national security screening tests of the time were making an unacceptably high number of false negative errors (Barland, Honts, & Barger, 1989; Honts, 1991; 1992; 1994) the DODPI attempted to develop a more accurate screening test. It should be noted that the primary concern in conducting national security screening tests is a desire not to make false negative errors. Following a series of studies that have not been published, Reed (1994) describes the product of the DODPI's efforts. In the first study reported in Reed (1994), the TES, a test format with only DLC questions for comparison, was tested against two versions of the Counterintelligence Scope Polygraph (CSP) test. One version of the CSP used probable-lie comparison questions while the other used directed-lie comparison questions. The TES outperformed both of the CSP formats in terms of correctly identifying guilty subjects. The CSP with directed-lie comparisons was slightly, but not significantly, better at identifying guilty subjects than was the CSP with probable-lie comparisons. A second study was reported that produced even higher accuracy for the TES, a DLC test format. Little information is provided about the third study, but it also appears to show considerable discriminability for the DLC based TES.

Table 3
Accuracy of Decisions in Barland (1981).

	Decision		
Guilt	Truthful	Inconclusive	Deceptive
Truth Teller	16	5	5
Deceiver	5	4	21

DODPI Research Staff (1998)

Ironically, the article immediately following Matte (1998) in this journal was an empirical paper from DODPI describing their success with the TES, a DLC-based technique. DODPI (1998) reported a mock espionage/ sabotage study that involved 82 subjects. All subjects were tested with the TES. Excluding one inconclusive outcome, the examiners correctly identified 98% of the innocent subjects and 83.3% of the guilty subjects. Although DODPI did not report a correlation measure of discriminablity one has been calculated; We calculated the detection efficiency r for the DODPI study to be, $r = 0.83$, $p < .001$. This statistic indicates that the DLC-based TES is extremely successful in discriminating between innocent and guilty subjects.

Abrams (1991)

Abrams (1991) reports the results of a study of 10 confirmed field cases from Abrams' private practice. Abrams included a single DLC question as the last question of the final chart of the examination. Abrams found that comparisons made to this single DLC question produced more positive numerical scores than comparisons to the probable lie questions of the study. No statistical tests were reported, perhaps because of the extremely small sample size. In any event, without statistical information it is impossible to evaluate whether the effects reported by Abrams were reliable or not. In addition, Abrams's peculiar use of a single DLC question at the end of the test is not representative of the current uses of that technique. The major discrepancy between Abrams's technique and the techniques used by others greatly minimizes the generalizability of Abrams findings. Finally, even if Abrams's findings are reliable, they are not surprising. Including the DLC question at the end of a test makes the DLC question in Abrams's study a novel stimulus. It is well-established that novel stimuli cause marked physiological responses. It is almost certain that the novelty of the DLC question in Abrams's study is responsible for causing subjects' observed responses. Moreover, it is likely that any question, even a neutral or previously unasked probable-lie question would have produced the same result under the conditions of the Abrams study. Abrams acknowledges the limitations of his (1991) study and had the following to say under cross-examination in United States v. Gilliard (1996):

> Q. Would you agree that your study represents too small a sample to make generalizations from?
>
> A. Worse than that. It's that the directed lie is only -- only occurs one time at the end of the test, and that certainly weakens it, and it's indicated in that paper.
>
> Q. In fact, you only used ten subjects?
>
> A. That's correct.
>
> Q. And of those ten subjects, you only used the directed lie on one of the three charts that you ran on each subject?
>
> A. That's correct.
>
> Q. And because of the very small amount of data, Professor Honts felt like that the generalizations you were making in your paper and which you have made today were not justified. Isn't that what he indicated?
>
> A. That's, that's true. . . . .
>
> Q. So, you would characterize the study that you published which in 1991 as a pilot study?
>
> A. That would be a good description.
>
> Q. And you would agree with someone who said more work needed to be done?
>
> A. Yes.
>
> Q. Before generalizations could be made from the research?
>
> A. Well, it certainly provides some suggestion of what needs to be done.
>
> Q. This is 1996, Doctor. How many studies have replicated your results in that study?

A. None as far as I know.

Thus even Abrams acknowledges the weaknesses of his study and that his results are unique in the empirical literature. Interestingly, Matte (1998) failed to mention any of these problems with the Abrams (1991) study.

Construct Validity of the Directed Lie Control

Unlike criterion validity, which can be evaluated directly with statistics, construct validity is a much less tangible concept. According to a current text on research methods in the social sciences, "Construct validity concerns the question of whether the results support the theory behind the research. Is there another theory that would predict the same result?" (McBurney, 1998, p. 83). Graziano and Raulin state the following, "Construct validity refers to how well the study's results support the theory or constructs behind the research and asks whether the theory supported by the findings provides the best available theoretical explanation of the results" (1993, p. 171). Cook and Campbell (1979) in their classic text on validity and quasi-experimentation state the following:

> Construct validity is what experimental psychologists are concerned with when they worry about "confounding." This refers to the possibility that the operations which are meant to represent a particular cause or effect construct can be construed in terms of more than one construct, each of which is stated at the same level of reduction. Confounding means that what one investigator interprets as a causal relationship between theoretical constructs labeled A and B, another investigator might interpret as a causal relationship between constructs A and Y or between X and B or even between X and Y. (p. 59)

Construct validity thus has to do with theory, and with the accurate labeling of constructs developed from that theory.

Matte (1998) states, "The Directed Lie Control Question however appears to lack construct validity in that it fails to demonstrate adequate capacity to function as a control question of less intensity than the relevant question for the guilty subject." (p. 64). There are two problems with that assertion. The first, and most important, is that the research previously discussed makes it clear that Matte's statement is false. With the exception of the seriously flawed study by Abrams (1991), all of the studies described above had low false negative rates. Were Matte's (1998) assertion about the DLC question test true, the false negative rates for the DLC should have been significantly higher than those for probable-lie comparison question test. In those studies that included both techniques, the false negative rates for the DLC were either no different than those for the probable-lie or were lower for the DLC.

The second problem with Matte's assertion is that it is not clear that his concerns relate to construct validity at all. A discussion of construct validity must include a theory within which the construct is evaluated. Matte's (1998) paper does not state a theoretical context within which to evaluate the DLC question test. To make a scientific assertion that the DLC question test lacks construct validity Matte would have to formally specify the theory he was using and present data showing that the DLC does not function as expected by that theory. Preferably those data would come from multiple methods and multiple sources (Cook & Campbell, 1979). However, Matte has done neither of these things. Moreover, Matte's (1998) statements indicate that he is more concerned with face validity than construct validity. Face validity is more a legal concept then a scientific one and is concerned with the question, "Does the instrument appear to measure what it is designed to measure?" Thus, judgments concerning face validity are more subjective than scientific, whereas judgments concerning construct validity are more scientific than subjective

Matte's Table 1

It may be that Matte's Table 1 was intended to embody some theoretical notions, but these are not stated explicitly. Moreover, no data are provided to support the categorization of the DLC question test with Matte's

Table 1. The information in Matte's Table 1 represents speculation and personal opinions rather than empirical data. Thus, Matte's Table 1 is meaningless except as a statement of Matte's opinion. The content of Table 1, should be given no more weight than any other statement of opinion until the studies have been conducted to evaluate the hypotheses contained therein.

If Matte's Table 1 is analyzed, the content of the items seems to assert that the emotion of fear is a necessary state for the psychophysiological detection of deception. However, all of the major scientists in the area have rejected that notion. Interested readers should see the discussion of this topic in Rosenfeld (1994).

Aside from questions concerning of the role of fear in the psychophysiological detection of deception, there are data that directly address some of the other assertions presented in Matte's Table 1, and those data contradict his assertions. Horowitz et al. (1997) gave a post-test questionnaire to their subjects that asked them to rate the importance of each of the test questions to the outcome the examination. These subjects also rated the degree of physiological reactivity they thought they produced in response to each question. According to the formulation in Matte's Table 1, there should have been significant differences between the perceptions of subjects who received the DLC questions and those who received the probable-lie questions. There were no differences in this regard. All subjects, regardless of guilt and type of comparison question ranked the importance of questions types in the following order: relevant > comparison > neutral. In terms of how they perceived their physiological responses, innocent subjects rated their responses to comparison questions as greater than their responses to relevant questions and guilty subjects rated their responses to relevant questions as being greater than their responses to comparison questions, regardless of the type of the comparison question. This finding directly contradicts Matte's theoretical assertions and suggests that at least some of the speculation presented in Matte's (1998) Table 1 is incorrect. It is interesting that while Matte (1998) did cite the Horowitz et al. (1997) study, he failed to mention the results in that study that directly contradict much of his Table 1.

Conclusions

Matte's analysis of the DLC has been found to be lacking in several respects. Matte's review of the empirical literature was incomplete and misleading. Matte's inclusion of data from the positive control test with data from studies of the DLC question test was a particularly egregious error. Even a cursory analysis of how the two techniques are presented to the subject reveals that they are not comparable. Except for the small and methodologically flawed study conducted by Abrams (1991), the empirical literature is strongly supportive of the validity of the DLC question test. In direct contradiction to Matte's conclusions, all of the high quality studies conducted to date show that tests using the directed-lie comparison questions are as valid as, or more valid than, tests using probable-lie control questions. This fact was reflected in the policy of the U. S. Government when they adopted directed-lie questions as the only comparison questions used in their new Test for Espionage and Sabotage. Notably, the TES is administered in a setting where the cost of a false negative outcome is particularly high. Policy makers with the Government must have been assured that the directed-lie comparisons minimize false negative outcomes, otherwise they would not have proceeded in that direction.

Matte's (1998) assertions about the construct validity were found not to address construct validity but rather to address face validity. Even so, many of Matte's statements are contradicted by data. Matte's analysis of the "construct" validity of the DLC should be treated as nothing more than opinion until data are presented to support his position.

A considerable body of evidence supports the validity of the DLC. It has been adopted by the U. S. Government for use in an extremely sensitive national security setting wherein the cost of false negative outcomes is particularly high. The directed lie test is easily standardized and offers many psychometric advantages to practitioners (see the discussion in Honts, Raskin, & Kircher, 1995). Its use is scientifically supportable and

should be considered by policy makers who are interested in conducting reliable and valid psychophysiological detection of deception tests.

References

Abrams, S. (1991). The directed lie control question. *Polygraph*, 20, 26-31.

Barland, G. H. (1981). *A validity and reliability study of counterintelligence screening tests.* Unpublished manuscript, Security Support Battalion, 902nd Military Intelligence Group, Fort George G. Meade, Maryland

Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the Accuracy of Security Screening Polygraph Examinations.* Department of Defense Polygraph Institute, Fort McClellan, Alabama.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin.

Department of Defense Polygraph Institute Research Division Staff (1997). A comparison of psychophysiological detection of deception accuracy rates obtained using the Counterintelligence Scope Polygraph (CSP) and the Test for Espionage and Sabotage (TES) question formats. *Polygraph*, 26, 79-106.

Department of Defense Polygraph Institute Research Division Staff (1998). Psychophysiological detection of deception accuracy rates obtained using the Test for Espionage and Sabotage (TES). *Polygraph*, 27, 68-73.

Driscoll, L. N., Honts, C. R., & Jones D. (1987). The validity of the positive control physiological detection of deception technique. *Journal of Police Science and Administration*, 15, 46-50.

Fuse, L. S. (1982). *Directed lie control testing technique.* Unpublished manuscript.

Golden, R. I. (1969, August). *The yes-no technique.* Paper presented at the annual meeting of the American Polygraph Association, Houston, Texas.

Graziano, A. M., & Raulin, M. L. (1993). *Research methods: A process of inquiry,* second edition. New York: Harper Collins.

Honts, C. R. (1999). The discussion (stimulation) of comparison questions between list repetitions (charts) is associated with increased test accuracy. Manuscript accepted for publication in *Polygraph.*

Honts, C. R., (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports*, 4, 91-116.

Honts, C. R., (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. *Forensic Reports*, 5, 215-218.

Honts, C. R. (1994). The psychophysiological detection of deception. *Current Directions in Psychological Science*, 3, 77-82.

Honts, C. R., & Driscoll. L. N. (1989). Validity of the positive control polygraph test: Comments on Forman and McCauley. *Polygraph*, 18, 158-167.

Honts, C. R., Kircher, J. C., & Raskin, D. C. (1995). Polygrapher's dilemma or psychologist's chimaera: A reply to Furedy's logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 20, 199-207.

Honts, C. R., & Raskin, D. C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16, 56-61.

Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.

Horvath, F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques, *Journal of Police Science and Administration*, 16, 198-209.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.

Matte, J. A. (1998). An analysis of the psychodynamics of the directed lie control questions in the control question technique. *Polygraph*, 27, 56-67.

McBurney, D. H. (1998). *Research methods*: 4th edition. Pacific Grove, CA: Brooks/Cole.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*: 3rd Edition. New York: McGraw-Hill.

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence*. New York: Springer. (247 - 296).

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126-136.

Raskin, D. C., Honts, C. R., & Kircher, J. C. (1997). The scientific status of research on polygraph techniques: The case for polygraph tests. Chapter in, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) *Modern scientific evidence: The law and science of expert testimony* (pp. 565-582).

Reed, S., (1994). A new psychophysiological detection of deception examination for security screening. *Psychophysiology*, 31, S80. (Abstract)

Rosenfeld, J. P. (1994). Alternative views of Bashore and Rapp's (1993) alternatives to traditional polygraphy: A critique. *Psychological Bulletin*, 117, 159 - .

Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology*, 67, 391-400.

U. S. v. Gilliard (1996). Transcript of the Daubert hearing. Available online: http://truth.idbsu.edu/polygraph/gilliard/abrams.htm

U. S. v. Gilliard (1998). United States Court of Appeals, Eleventh Circuit. No. 96-9459. Available online: http://www.law.emory.edu/11circuit/wpds/jan98/96-9459.man

Situational Sequencing Test

Ryszard Jaworski

Keywords: Case study, experimental technique, forensic examination, Peak of Tension.

In November 1992 the Regional Court in Krakow asked me to carry out a polygraph examination of one of two defendants accused of murder. I was skeptical whether the examination would be productive, as the defendant had been in detention awaiting trial for 21 months. Prolonged detention may influence the emotional state of the examined person in various ways. Some innocent persons may develop an emotional attitude toward the case. During the test it may cause physiological reactions following the questions connected with the event. Identifying real reasons of the reactions would be very difficult. Other persons, especially those convicted before and familiar with the evidence procedure may conclude that prolonged proceedings are caused by low value of the collected evidence, which means high probability of acquittal. Such reasoning of the defendant would be quite justified, but it might cause them to calm down and their emotional state to subside. It is possible that such a person would not display the reactions following the questions connected with the event, even if the examined person were the perpetrator of the crime.

The Regional Court insisted on carrying out the examination irrespective of its usefulness to fulfill a firm demand by one of the defendants. The defendant had pleaded not guilty, and motioned new evidence, mainly regarding the conduct of a polygraph examination. The Court rejected the motions but the defendant complained to the higher instance, which required transferring the files and waiting for the court of higher instance to consider the complaint. Normally the next trial could take place after two or three months, but the Regional Court considered the matter for over a year, and the prospects were that the case could take a few years.

Facts of the Case

According to the indictment, on the 2nd of February, 1991, in Krakow, Wladyslaw Kowalczyk and Emil Pasternak deceitfully enticed Mariusz Kuzaj to the cellar of the building where Pasternak lived. They detained Kuzaj in the cellar, persecuted him and appropriated his car and personal effects. They demanded a ransom from the victim's family, but in vain. They bound him with a rope, forced a soporific agent into his mouth, sealed the mouth with a tape, put a loop around his neck, and tightened the knot. After that, they put M. Kuzaj into a metal drum, filled it with cement mix and poured water into the drum. They carried the drum out of the cellar and transported it to the neighboring street. The body was found after a few weeks. Due to the partial decay of the body, the autopsy did not specify the cause of death. The causes may have been an overdose of the soporific agent, suffocation following sealing of the mouth, strangulation by the rope tied around the neck, or suffocation in the drum after being filled with cement.

Wladyslaw Kowalczyk described his role in a completely different way. Kowalczyk contended that he did not know that Pasternak intended to kill Kuzaj. Rather, Pasternak had told him he wanted to force Kuzaj to pay a debt. Kowalczyk purportedly did not assist Pasternak voluntarily, since Pasternak had a very sharp file tied to a long,

Dr. Ryszard Jaworski is a Professor of Law at the University of Wroclaw, Poland. He conducts criminal polygraph testing in Poland, and has submitted this case study as a demonstration of a novel approach to polygraphy. Dr. Jaworski can be reached at: Uniwersytet Wroclawski, Katedra Kryminalistyk, 50-145 Wroclaw Uniwersytecka 22/26, Poland.

metal pipe. He put the file to Kowalczyk's throat, and demanded Kowalczyk perpetrate various actions against the detained Kuzaj. Kowalczyk admitted that, acting under physical coercion, at the very beginning of the crime he tied Kuzaj's hands and gagged his mouth with a stocking. At the later stage of the crime (after removing the stocking) he poured some unknown type of liquid into the victim's mouth and sealed it with a tape, all under a threat of death from Pasternak. Consequently, Kowalczyk did not feel he was murdering Kuzaj. Kowalczyk claimed that when he realized that Pasternak's activities might result in Kuzaj's death, he escaped from the cellar using a moment of Pasternak's inattention.

After a few days Pasternak met Kowalczyk again, and told him that he had released Kuzaj but had kept his car in return for his debt. Kowalczyk believed that Kuzaj had been released, and he helped Pasternak push Kuzaj's broken car to a parking lot.

Kowalczyk's version of events did not appear credible to the Court. Kowalczyk had been convicted several times of theft and robbery, which made his participation in this crime seem reasonable. Kowalczyk was 40 years old, physically very strong, 185 centimeters tall (6'1") and weighed about 90 kilograms (198 pounds.) Pasternak was 55 years old, 160 centimeters tall (5'3") and weighed less than 60 kilograms (132 pounds.) It is hardly credible that a much weaker man could intimidate a strong and stout man, even if the latter were using a sharp file.

A fair evaluation of Kowalczyk's explanation also required taking into consideration circumstances favorable for him. During the first interrogations after detention he revealed many details and took part in the reconstruction of the event (presenting the version favorable for himself). Many of the indicated details were consistent with other evidence.

The other defendant, Pasternak, was arrested much later. He pleaded not guilty and put the whole responsibility on Kowalczyk. He categorically refused to undergo any polygraph examination. He had also been previously convicted.

Additionally, there was very serious circumstantial evidence indicating that Pasternak had murdered three other persons. Pasternak established relationships with those persons, and when they disappeared under unexplained circumstances, Pasternak appeared to own their possessions (including a car). As the apparent victims lived alone, their absence was reported late and the searches were futile. Because the bodies were never found, Pasternak was never charged with murder.

In accordance with Polish praxis, the examination was to be carried out as an expert appraisal, and the Court ruled to appoint a polygraph expert. The Court insisted on establishing the following:

1. Whether Kowalczyk was either a voluntary or coerced accomplice;

2. Whether Kowalczyk knew the actual final purpose of Pasternak's activities (murdering Kuzaj);

3. Whether Kowalczyk helped Pasternak in activities that might have caused the victim's death, specifically, in tightening the knot on the neck, putting into the drum, and filling the drum with cement and water;

4. Whether Kowalczyk knew at which moment the victim died.

Analyzing the actual state of the case and the information of W. Kowalczyk I decided that there was a chance of carrying out the examination and performing the analysis of recordings according to the Reid technique. Prolonged detention would not necessarily have caused excessive emotional arousal of Kowalczyk, as he was a recidivist. He understood the meaning of evidence and knew that possible doubts had to be ruled to his favor. Moreover, the examined person had to be treated as a person of increased resistance as he had repeatedly been convicted and interrogated in other cases. In his milieu he was considered a strong personality of high psychological resistance.

Thus, it could be assumed that possible unfavorable effects of prolonged detention were to a certain extent balanced by W. Kowalczyk's increased emotional resistance.

There were other problems that had significance for the test and potential reactions of the tested person. The most important, though not all, were as follows.

1. The charge of homicide was very serious and the test result, even if unfavorable for Kowalczyk, could not make his situation worse. The risk to Kowalczyk by demanding the polygraph test was slight. The results could only act in his favor. That was an additional factor that could theoretically decrease his reactions, even if he were the perpetrator of the murder.

2. It was suspected that Kowalczyk had familiarized himself with the polygraph test and that he would attempt to control his physiology in such a way that the size and proportions of the reactions would be favorable to him. His high resistance to stress gave him such potential. During the pretest interview he admitted that he had read much about the polygraph and that "he was not scared of the machine", which confirmed the above suspicions.

3. The circumstances described above spoke in favor of performing the Reid "Yes" test, yet I was, and still am, convinced that judges are not able to evaluate it properly according to the assumptions of the Reid technique. The knowledge of an average judge concerning the polygraph test is still very moderate and many lawyers still see the polygraph as a "lie detector". I am afraid that the reactions of the tested person in the "Yes" test would have been evaluated by the judge and the lawyer of the defense contrary to the assumptions of the Reid technique. For this reason I gave up this version of the test.

4. The crime consisted of several stages, too many to comprise the problems relevant for the court in a Reid test. Additionally, Kowalczyk confirmed his participation in some of the stages, while the Court wanted to know his actual role in all stages. That called for another solution. It is also worth mentioning here that the Lafayette polygraph, model 76058 from 1976 had no amplification of the cardiograph, which limited the duration of the test and number of questions in the test.

5. Kowalczyk committed earlier varied crimes (burglary, robbery, battery), which created problems in the preparation of traditional probable lie control questions.

The Tests

The recordings were made with a Lafayette four-channel polygraph, model 76058. The instrument had neither an event marker, nor the amplifier for the cardiograph channel. Before each test I read the questions to the examinee and discussed them, stressing his role according to the test version.

Taking into consideration the above circumstances, I prepared two tests using the Reid technique. In the first one the critical questions were:

3. Did you help Pasternak voluntarily?

5. Did you cooperate in the homicide?

8. Did you give all the details truthfully?

10. Did you help to put Kuzaj into the drum?

11. Did you lie answering my questions?

The control questions were:

6. Did you ever use a dangerous implement in a fight?

12. Did you cheat your last employer?

See Charts 1-3. Charts 1 and 3 are straight through tests of this first series. Charts 2 is a control stimulation test. The key was #11.

Some critical questions from the second test were concerned with the same problems, but they were phrased with different words and different issues, not mentioned in the first test. I also employed a

nontraditional sequence of critical and control questions. It was aimed at preventing the tested person from controlling his reactions. I also wanted the questions to cover the greatest number of the fragments of the event. The critical questions of this test were:

3. Did you know Pasternak's intentions earlier?

5. Did Pasternak promise you any reward?

6. Did Pasternak share anything with you after the event?

9. Did you help to tie Kuzaj's hands?

10. Did you help to put Kuzaj into the drum?

The control questions were:

7. Did you ever use a dangerous implement in a fight?

11. Did you ever cheat your parents in important matters?

See Charts 4-7. Chart 5 is a straight through of the second series, preceded by another control stimulation test, marked Chart 4. Chart 6 is another control stimulation test and Chart 7 is a mixed question series. Charts 6 and 7 were conducted as the last two tests of this individual. The examinee may have been trying to manipulate some of the Reid tests. The key in both control stimulation tests were #11.

Because of some of the anticipated problems listed in the previous section, I complemented the classic Reid test with other tests approximating the assumptions of the Peak of Tension (POT) test.

I prepared three tests, each with 12 questions. The questions concerned the stages of the event. I assumed that possible reactions to the questions about consecutive stages might indicate actual participation of the defendant in the stages. It would enable me to determine whether the examined person participated only in some initial stages of the crime (according to his evidence), or whether he cooperated with Pasternak in all stages (according to the Prosecution).

In each test the form of questions was varied. In the first test all the questions concerned the hypothetical role of the examined person as "the witness of the event". Hence, all questions began with "Were you present at ...?" The second test employed the same stages of the event, but the questions considered the role of the examined person as "the accomplice". Therefore, these questions all began with "Did you help to ...?" In the third test the questions considered the role of the examined person as "the perpetrator" and were phrased "Did you personally ...?" I expected that the way of constructing the questions would cause differences in general level of agitation between the three versions of the test and that the version where the agitation level would be the highest may correspond to the actual role of the examined person.

The following test questions were used in the "witness" series.

1. Were you present at binding the hands?

2. Were you present at putting the stocking into the mouth?

3. Were you present at sealing the mouth?

4. Were you present at forcing the liquid into the mouth?

5. Were you present at tying the knot on the neck?

6. Were you present at pushing the car?

7. Were you present at binding with the rope?

8. Were you present at putting into the drum?

9. Were you present at filling with cement?

10. Were you present at pouring the water?

11. Were you present at carrying out the drum?

12. Were you present at transporting the drum?

See Chart 8 for the "witness" series.

After this test, Kowalczyk stated that he was unsure whether question 7 pertained to the binding of the victim's hands at the beginning of the event or the binding the body before putting it into the drum. I explained that the latter was the case. During the test the external manifestations of the subject's emotional state could be observed in the form of biting the lower lip and his difficulties with swallowing saliva.

The following test questions were used in the "accomplice" series.

1. Did you help to bind the hands?

2. Did you help to put the stocking into the mouth?

3. Did you help to seal the mouth?

4. Did you help to force the liquid into the mouth?

5. Did you help to tie the knot on the neck?

6. Did you help to push the car?

7. Did you help to bind with the rope?

8. Did you help to put into the drum?

9. Did you help to fill with cement?

10. Did you help to pour the water?

11. Did you help to carry out the drum?

12. Did you help to transport the drum?

See Chart 9 for the "accomplice" series.

The following test questions were used in the "perpetrator" series.

1. Did you personally bind the hands?
2. Did you personally put the stocking into the mouth?

3. Did you personally seal the mouth?

4. Did you personally force the liquid into the mouth?

5. Did you personally tie the knot on the neck?

6. Did you personally push the car?

7. Did you personally bind with the rope?

8. Did you personally put into the drum?

9. Did you personally fill with cement?

10. Did you personally carry the water?

11. Did you personally carry out the drum?

12. Did you personally transport the drum?

See Chart 10 for the "perpetrator" series.

The Evaluation of the POT Data

Pulse and Relative Blood Volume

In all the tests the pulse was 70-80 cycles per minute, relatively slow, and approximating the resting state. There was a very considerable differentiation in blood volume between the recordings. In the recording of the test in the "witness" version the amplitude was high and did not oscillate considerably. In the recording of the "accomplice" version the amplitude was considerably smaller, the smallest following question 5, and increased in the other half of the test. In the recording of the test in the "perpetrator" version the amplitude was also very small but increased considerably following question 8.

As seen in the cardiograph recordings, the contents of the questions in the "accomplice" and "perpetrator" versions agitated the examinee to a considerably higher degree than the "witness" version, as a small amplitude of blood volume tracing indicates greater emotional agitation. The contents of the questions in the other halves

of the tests, approximately from question 7, agitated the examinee considerably less than the questions from the first half. The increased cardiograph amplitudes indicate small emotional agitation. The relatively slow pulse may suggest only moderate agitation of the examinee about the problem phrases in the questions, but it may also have resulted from the high physical efficiency of the examinee's body.

Respiration

The depth and rate of breathing in all three tests is disturbed, though not to the same degree. It is visible in all three recordings that the irregularity of breathing rate in the first halves of the tests is greater than in the second. The greatest differences between the first and the second half of the recording occur in the "perpetrator" version.

Great irregularity of breathing rate indicates greater emotional agitation. For this reason it may be concluded from the breathing recording that the examinee was connected emotionally with the questions from the first halves of the tests (especially in the "accomplice" and "perpetrator" versions) more strongly than with those from the other halves.

Electrodermal Activity

The trend of the electrodermal activity (EDA) recording drops in all the tests, which may indicate that the examinee's emotional agitation decreases during the test. The upright inclinations of the EDA curve in the "witness" and "perpetrator" versions are caused by the hesitation before answering or answering in a complete sentence. It can not be ruled out that they were the examinee's attempt to hide the emotional agitation caused by the question's content. It may be concluded from the EDA recording that the other test questions do not agitate the examined person emotionally. If they were emotionally relevant to him, the increase of the electric conductivity of the skin should be expected, while actually the opposite was the case.

External Manifestations of Emotions

During the "accomplice" and "perpetrator" tests the examinee's face was very pale, which indicated a high degree of emotional agitation. That did not occur in the "witness" test. It follows from the degree of intensity of external manifestations of emotions that W. Kowalczyk was agitated much more strongly by the hypothesis of his role as an accomplice or a perpetrator than a witness of the event.

In addition, based on the physiological recordings, Kowalczyk is emotionally agitated to a much higher degree by the "accomplice" and "perpetrator" versions of the test than by the "witness" version. The considerable differences in cardiograph amplitudes, and external manifestations of emotions between these tests support this observation. It could be the basis for rejecting the defendant's claim that he was forced against his will to perform various activities. In reality he was free to make decisions and act.

Kowalczyk was more strongly agitated by the initial questions of the tests than by the final ones. The differences were the greatest in the "perpetrator" version. In the other halves of the tests the cardiograph tracing amplitude increased, breathing rate became more regular, and the electric conductivity of the skin quickly decreased. Thus, the claims of the examined person that he did not participate in the final stage of the event (putting the body into the drum, filling with cement, pouring the water, carrying out the drum) should be considered probable.

These conclusions were included in the expert opinion presented to the Court. However, the conclusions were based on the tests discussed above, while the final opinion on the whole examination was not so unequivocal and not conclusive. The following factors are responsible for more cautious formulation of final conclusions.

1. It was the first time I had carried out the discussed tests, treating them as complementary and auxiliary to the Reid tests.

2. The reactions in the classic Reid tests prevented ruling out the examinee as having

participated in certain elements of the crime. The first test was carried out twice, with the control-stimulating test in between, and it concerned the relevant details of the examinee's life. The other test was also done twice, with a mixed question series placed at the end of the examination. It was preceded with another control-stimulating test. During the test I watched the external manifestations of emotions, and evidence of attempts to control physiological reactions. Additionally, external manifestations of emotions and attempts to interfere with the recordings occurred during the Reid tests. In the control-stimulation tests the examinee appeared to attempt to interfere with the recordings.

It can not be ignored that the examined person had prepared for the test and said during the pre-test interview that "he was not scared of such examination" and "he could handle the machine". It also can not be unequivocally stated what influence the prolonged detention had, and how the examined person's increased psychological resistance influenced his reactions.

The examined person had physical predispositions to control his physiology as he had practiced bodybuilding for 20 years and in detention he took physical exercises every day. High physical efficiency quite naturally causes the slowing down of breathing and pulse rates.

The Regional Court passed judgment that Kowalczyk was considered only the accomplice of the crime. The court released Kowalczyk from the charge of assisting in murder, recognizing that he had assisted the perpetrator only in the initial stages of the crime. He was found guilty of assisting in deprivation of liberty, of physical and moral cruelty to the victim and of appropriation of Kuzaj's possessions. The Court rejected the defendant's claim that he had acted under physical coercion when committing the actions against the victim. It was declared that he had the freedom of decision and action. When outlining the reasons for the judgment, the Court stated the sentence was not based on the results of the polygraph expert opinion, but other evidence (testimonies of the witnesses, logical analysis of the facts, etc.)

Kowalczyk was sentenced to 8 years of imprisonment, and Pasternak received 25 years of imprisonment. The court of second instance reduced W. Kowalczyk's penalty to 3 years of imprisonment. The court of higher instance upheld the judgment of the Regional Court.

The Court's judgment concurred with the conclusions from the polygraph examinations. It might be questioned why the Court rejected the polygraph expert opinion, simultaneously passing a sentence consistent with its results. Regarding its failure to accept the polygraph evidence, the Court may have been influenced by the lack of an unequivocal final opinion, based on the inconclusive results of the classic test. It is also probable that the Court diminished the value of this evidence on purpose, as it is not universally accepted. The sophistication of the judges regarding the polygraph examination was very modest, and the theory of criminal proceedings presents a negative opinion in the matter, the result of prejudice and ignorance. Thus, there was a risk of reversal of the judgment if the Court had confirmed that the sentence was based on the results of the polygraph examination.

Discussion

Some may harbor doubts about these experimental tests. For example, the form of questions and the type of answers may facilitate the examinee controlling physiological reactions (agitation or inhibiting natural agitation). However, a similar situation exists in Peak of Tension (POT) tests, and many years of their application do not support such objections. Discovering the attempts of self-stimulation is made possible by watching the examinee's face. Videotaping the examination would make it possible to prove such attempts.

It has been argued by some that the examinee who knows the methodological basis of the tests, and is intelligent enough, may control his physiological functions in such a way that they will indicate the highest

level of agitation in the "witness" version, the test that would support his claims of innocence. Other tests, especially the ones based on the Reid technique, may successfully reveal such an attitude of the examinee, and the intention to mislead the expert. Moreover, the physiological recordings in this report show how diagnostically important the blood pressure amplitude may be. Controlling this physiological parameter is very difficult, and it is doubtful a person could alter this feature, even after special training. Critics may argue that every examinee will always be agitated most strongly with the "perpetrator" version, as it is potentially most threatening. Similar objections were raised with the Reid technique, but praxis showed that they were not justified. Other tests, especially the Reid test may verify the results of the analysis of the tests discussed here.

Conclusions

Methodological assumptions of the experimental tests discussed here were borne out by the concurring judicial outcome. I am of the opinion that as tests complementing the basic investigation techniques, they have strong advantages. They enabled the investigators to determine the actual role of the examinee in the crime. They showed that prolonged detention did not influence the emotional reactions of the examinee during the tests. I am of the opinion that these tests may be very useful in the cases when the roles of persons during a crime are diversified.

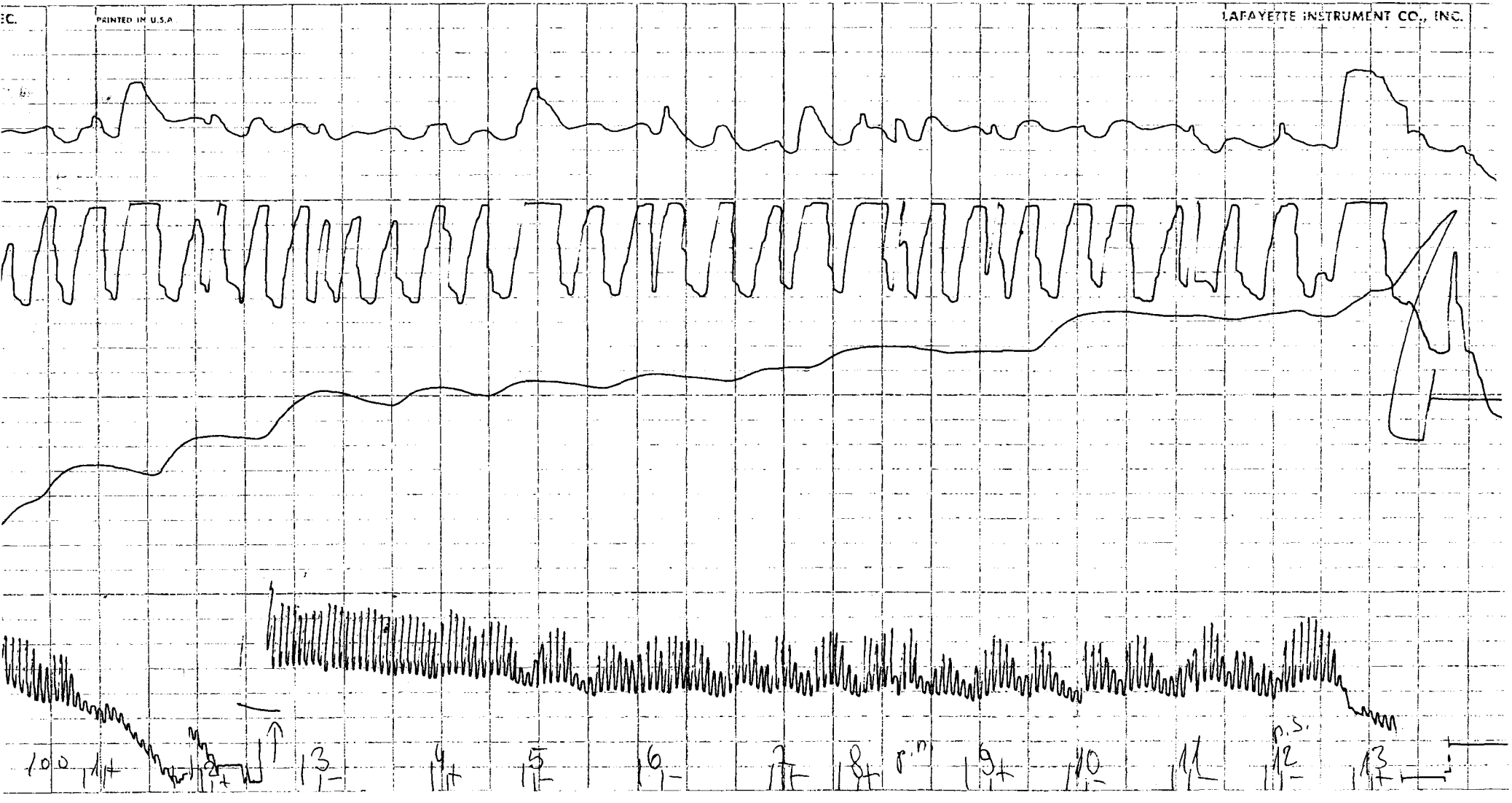


Chart 1: First chart of the Reid test, first series. The notation "p.n." signifies a sniffle, and the "n.s." is a swallow.

Chart 2: Control stimulation test. The key is #11.

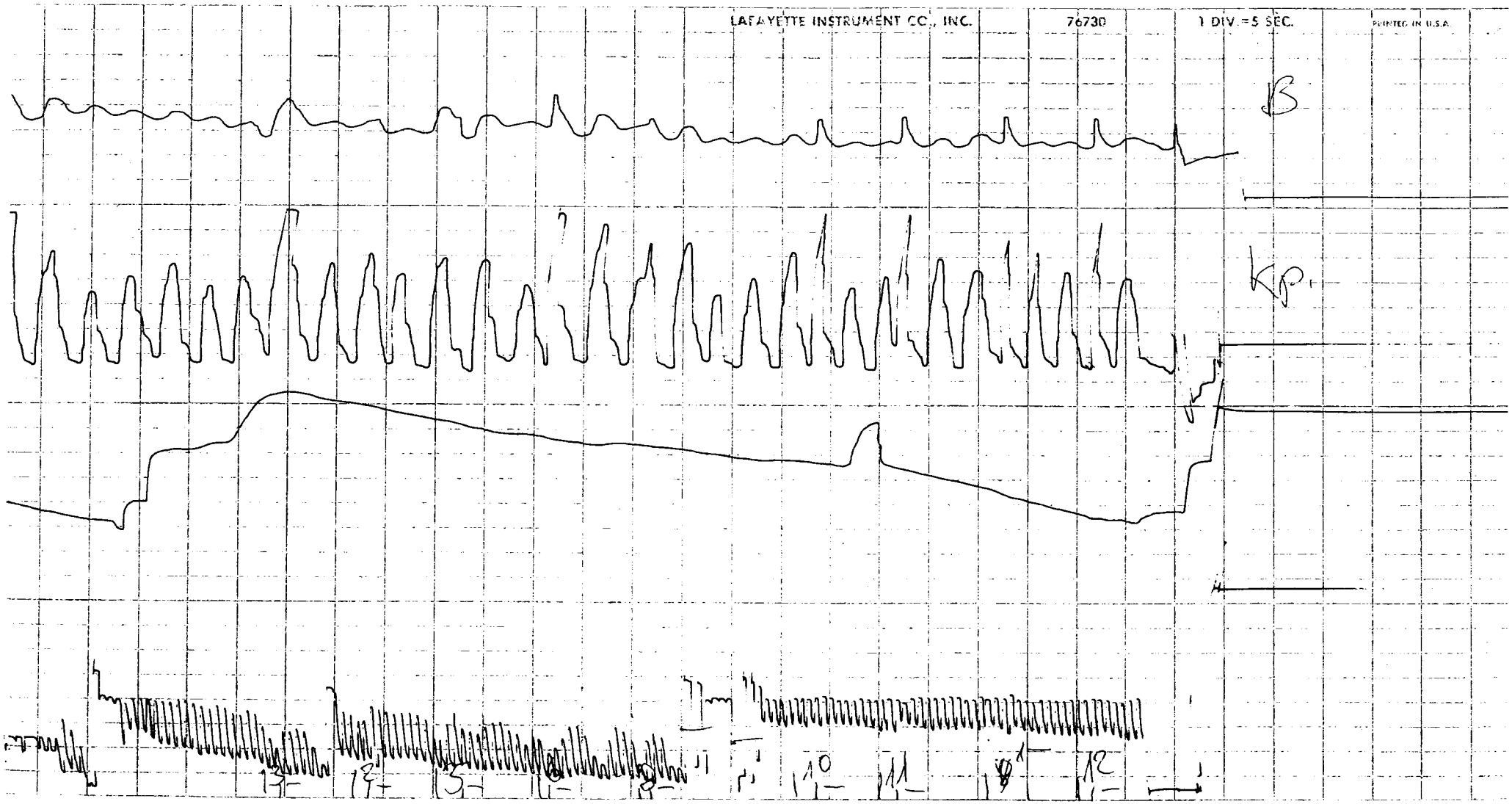
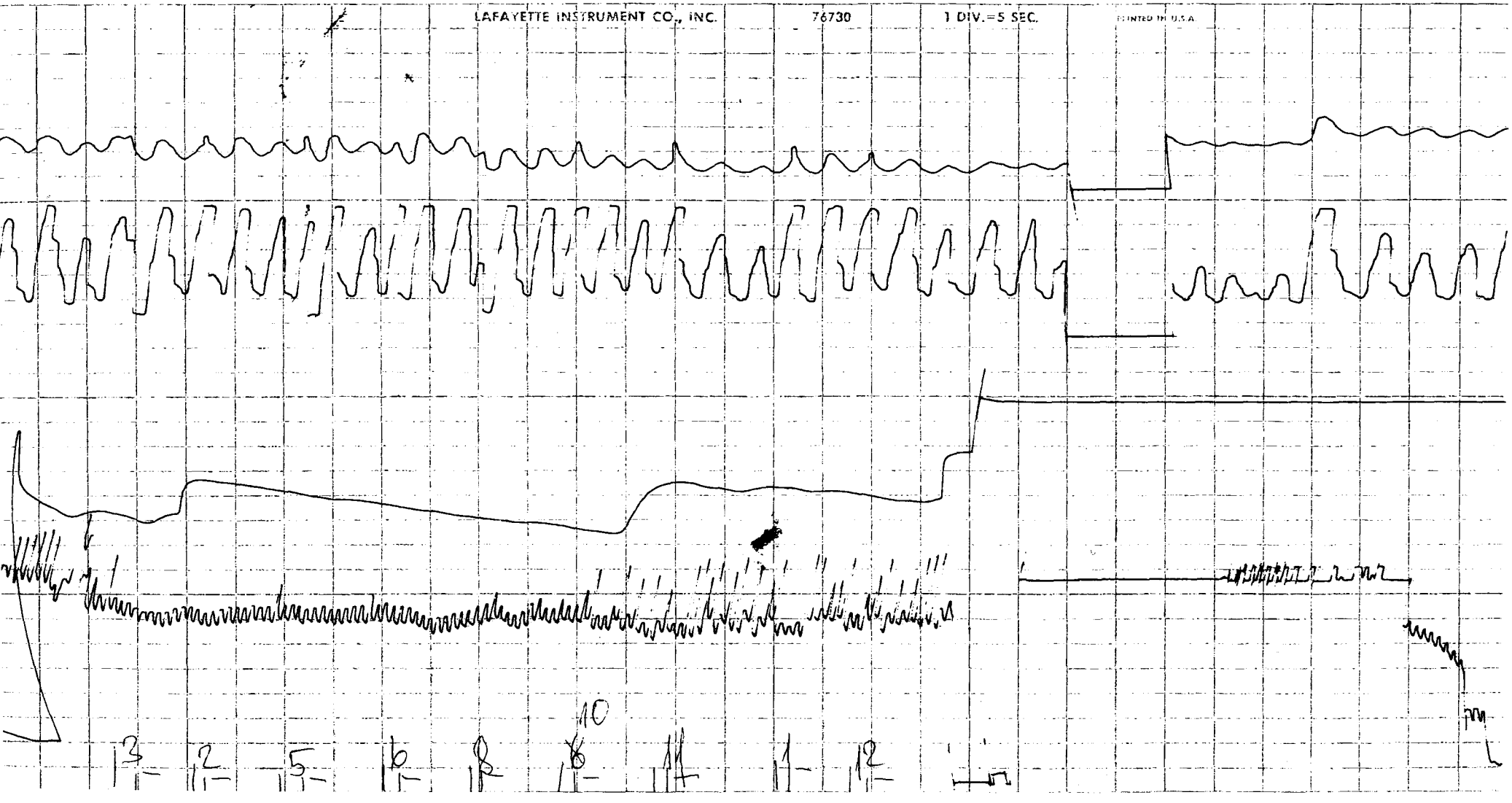


Chart 3: Second chart of the Reid test, first series.

Chart 4: Control stimulation test before second Reid series. The key was #11.



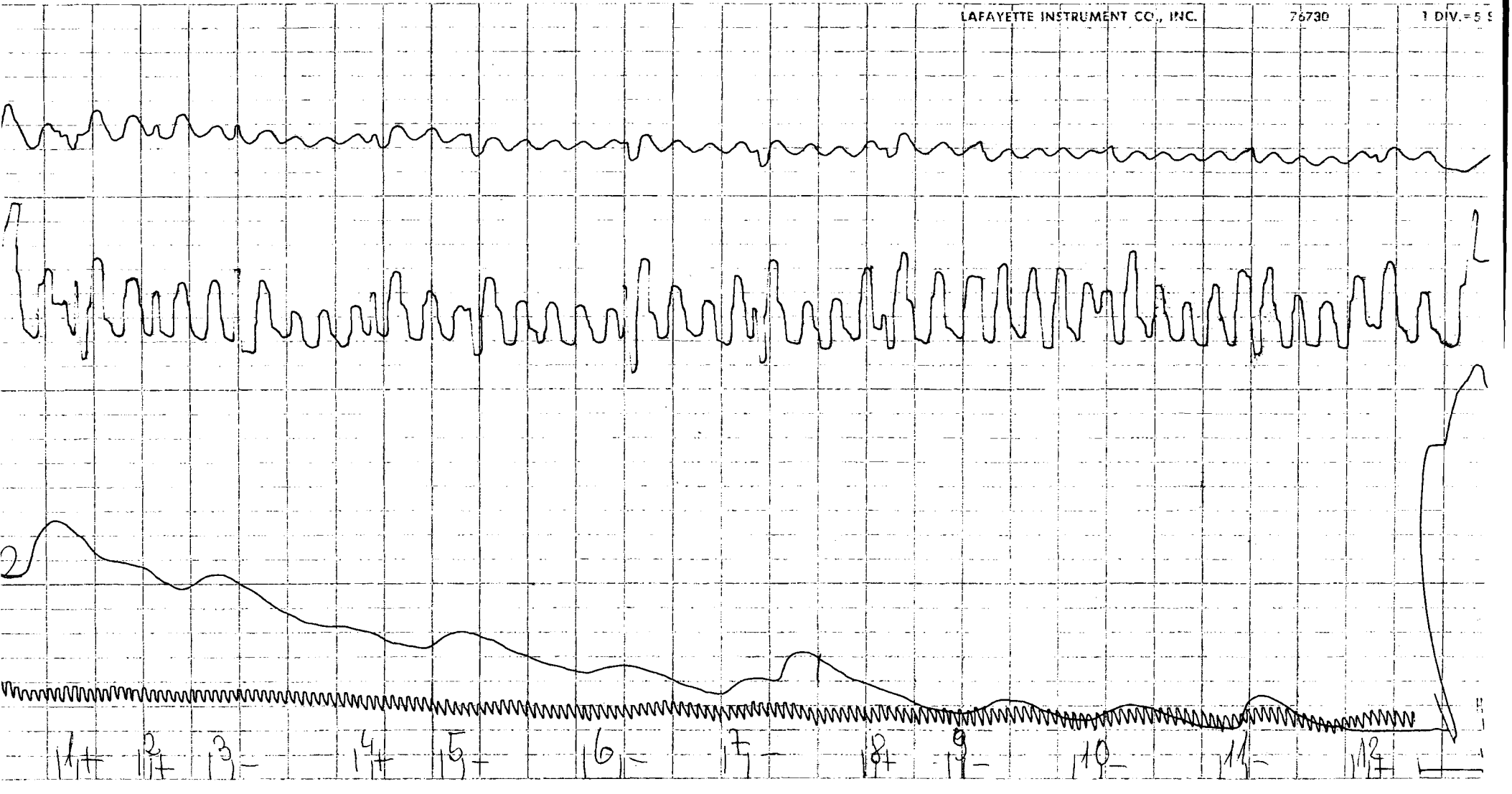
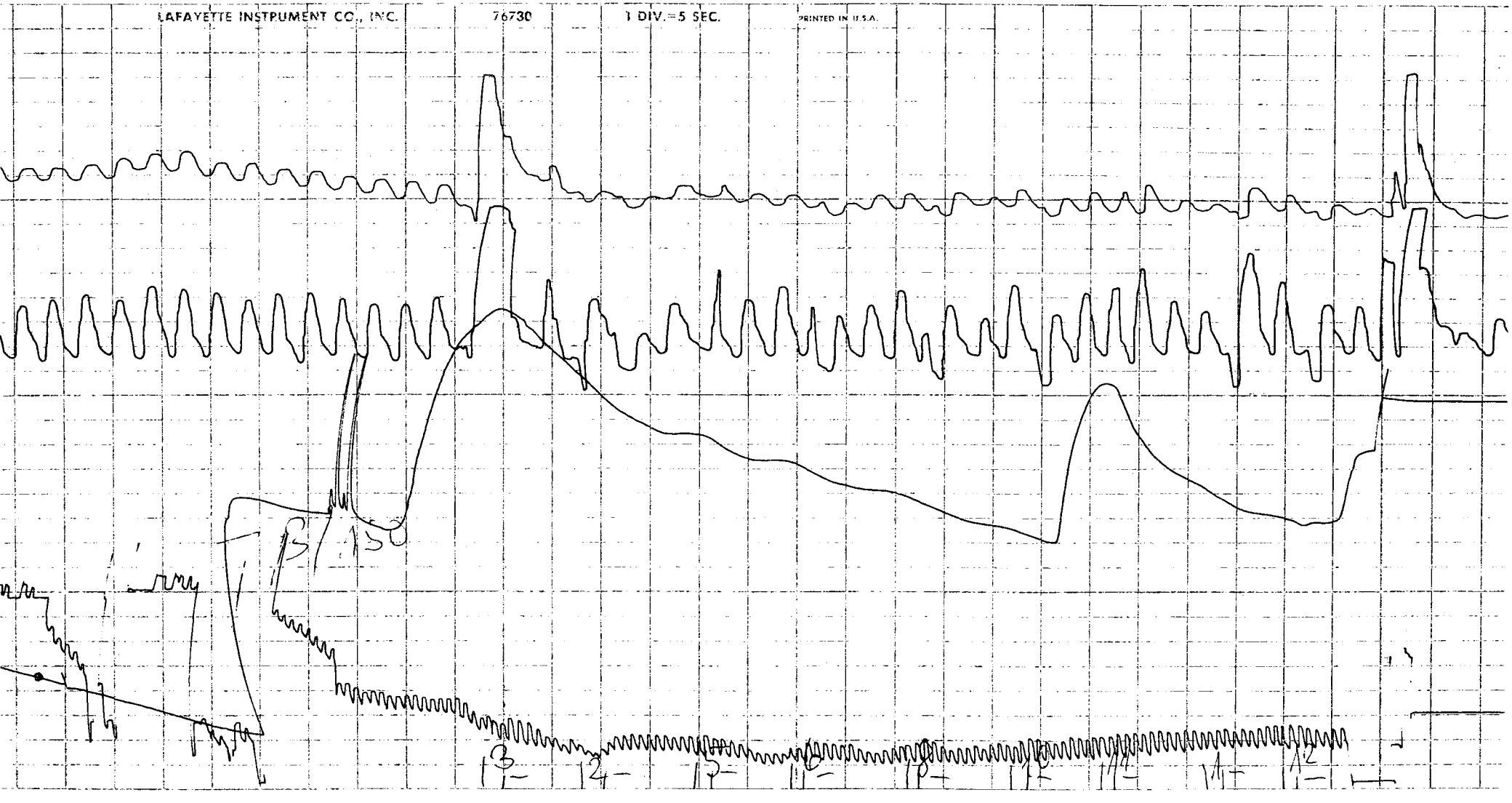


Chart 5: First straight through chart of the Reid test, second series.

Chart 6: Control stimulation test. This test was administered after the POT tests, but before the last Reid mixed question test (Chart 7). The key was #11.



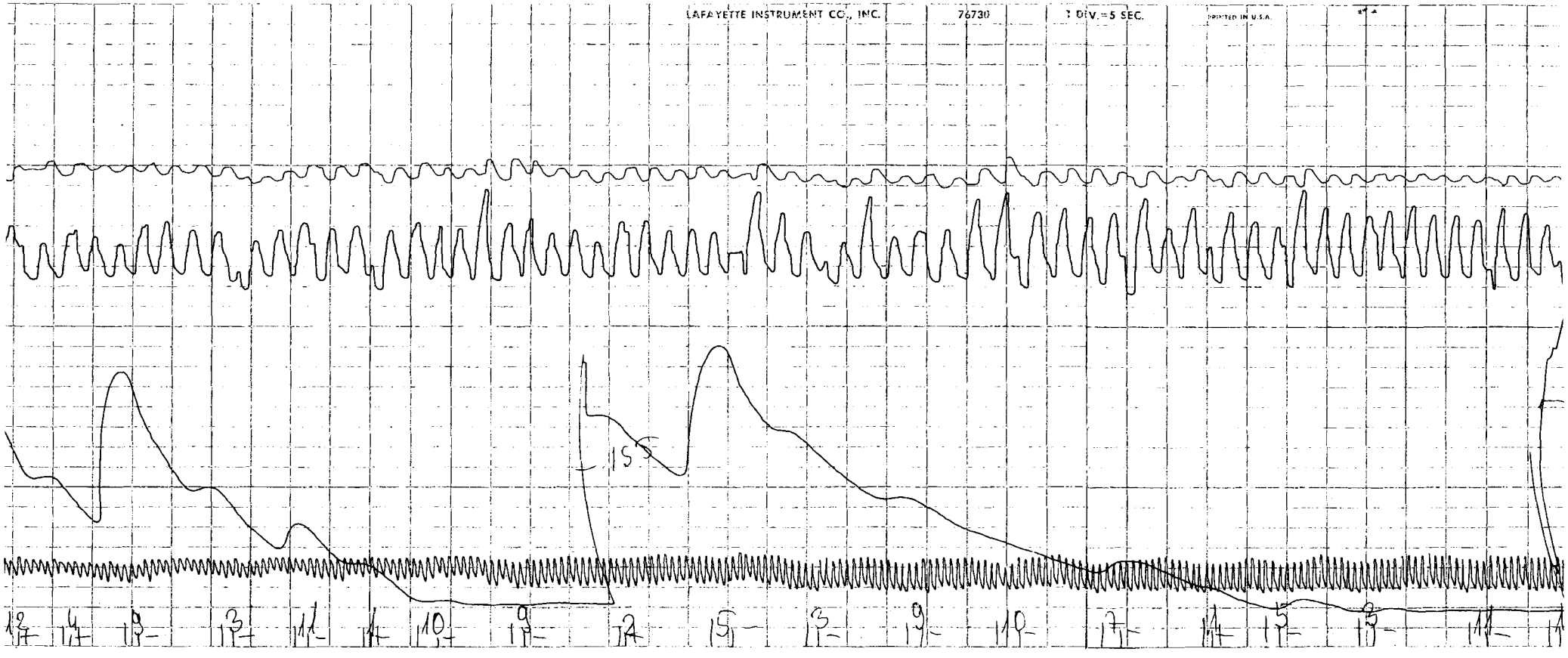
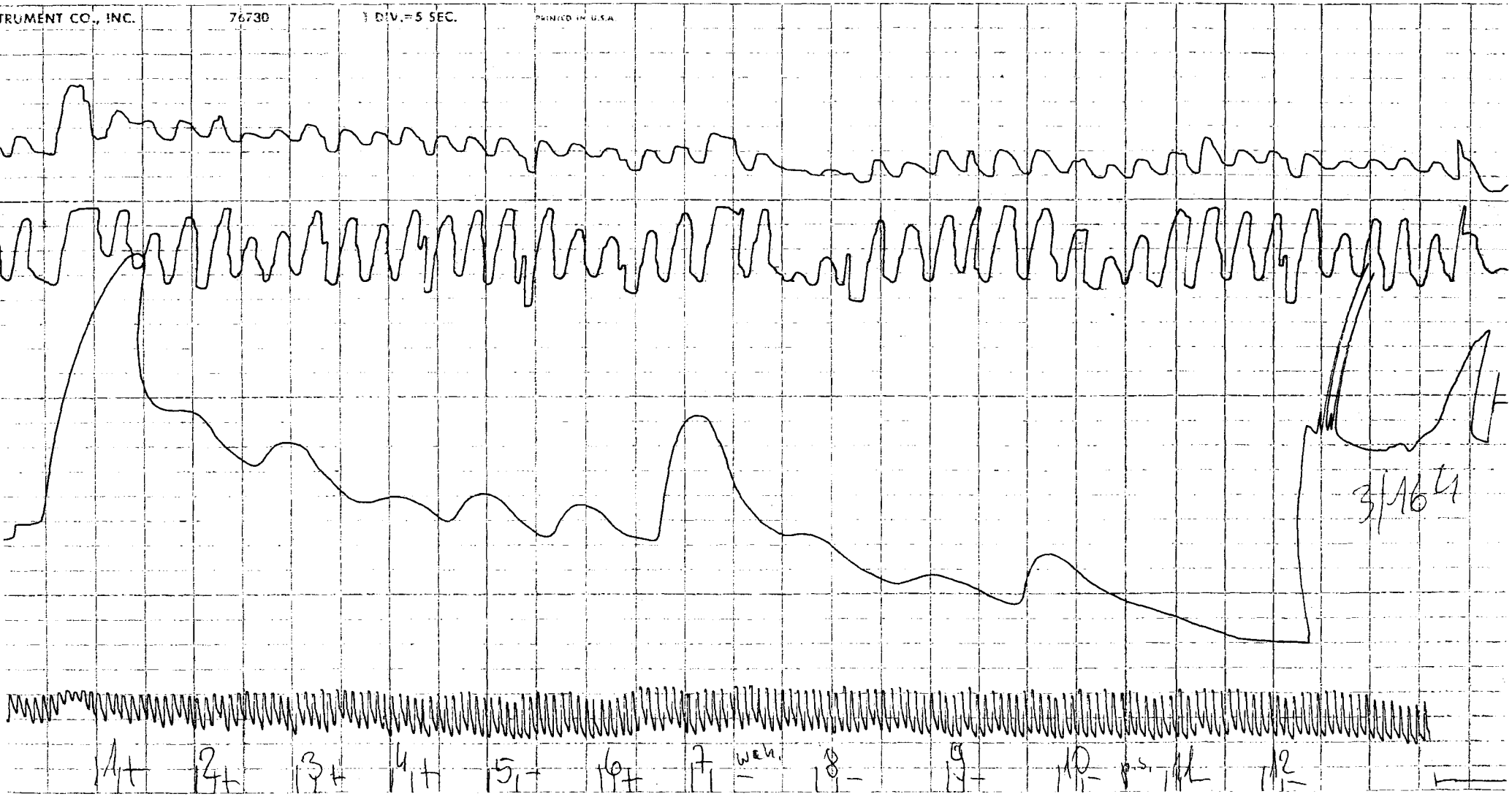


Chart 7. A nontraditional mixed question test of the Reid second series. This was the last test administered to this subject.

Chart 8. POT polygram in the "witness" portion of the examination. Notation "wah." indicates a hesitation to answer by the subject, while "p.s." was a swallow.



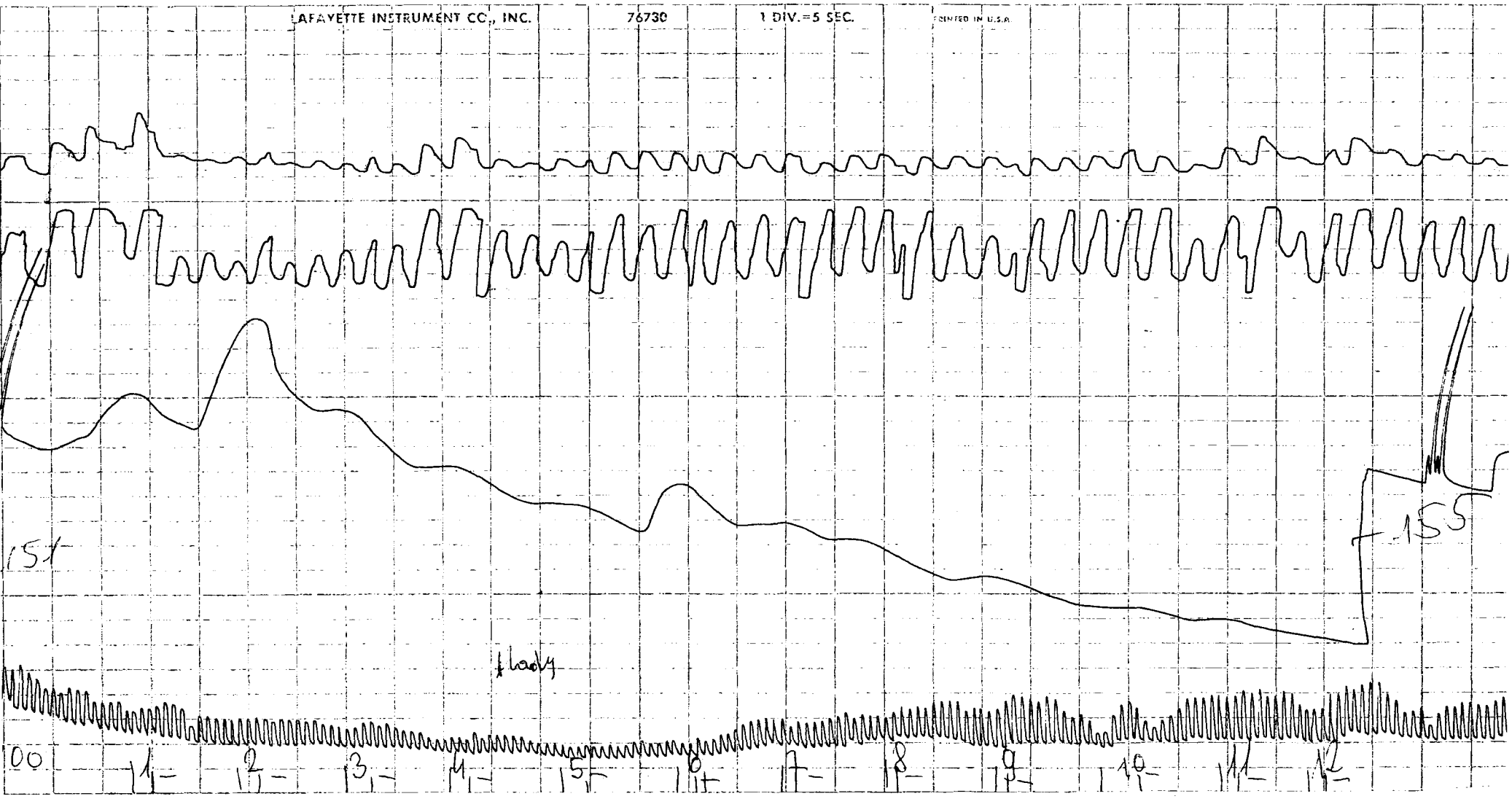
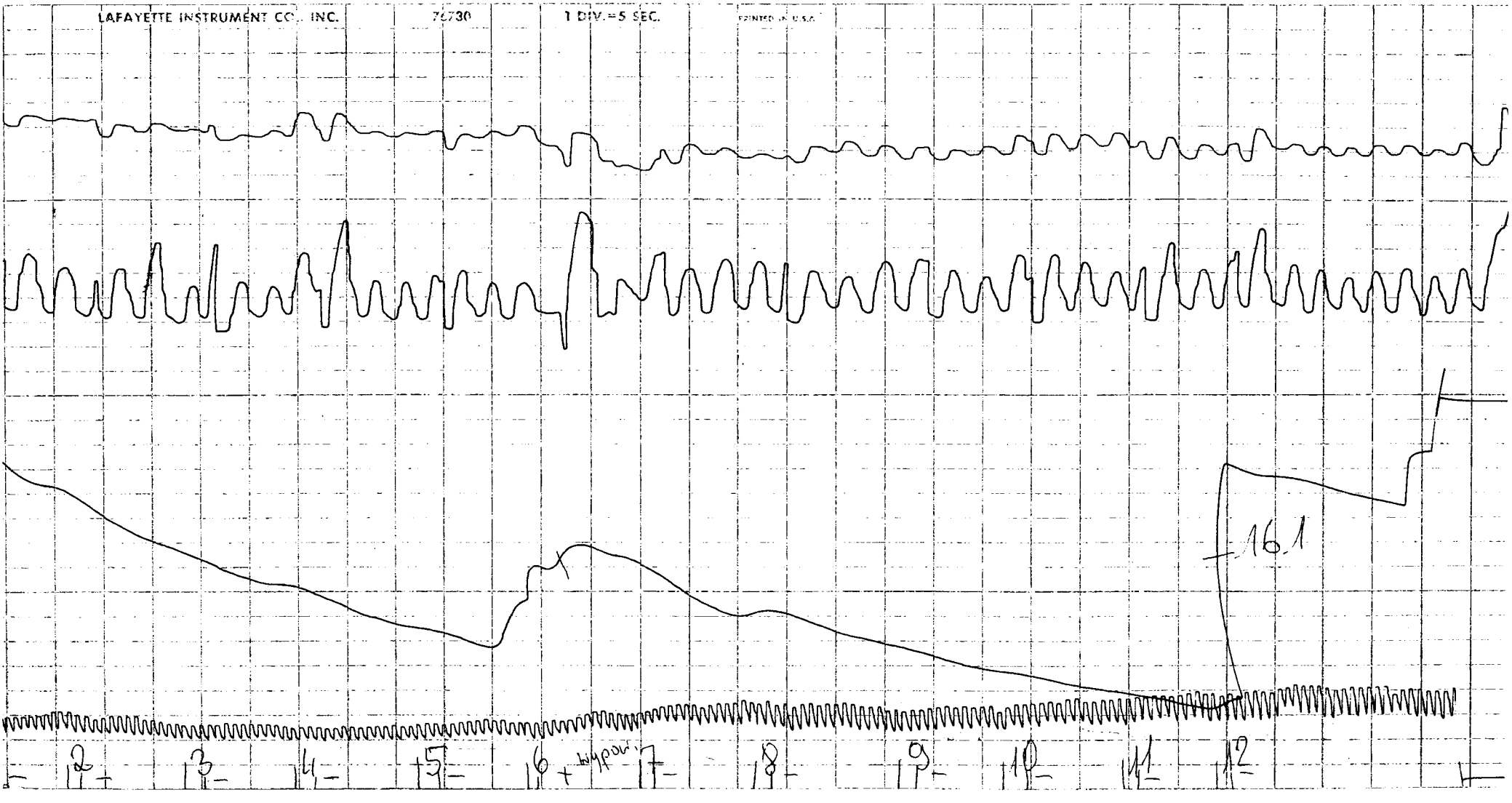


Chart 9: Polygram in the "accomplice" portion of the POT examination. Notation "blady" means pale. During the test the subject's face became visibly pale.

Chart 10: Polygram in the "perpetrator" portion of the POT examination. Abbreviation "wypow." means statement. During the test the face of the examined person was very pale.



Differences in Electrodermal Response Patterns for Males and Females

Yaacov Friedlander, Eran Gazit, and Efrat Eliahu

Abstract

Thirty Peak of Tension (POT) charts of males and 30 POT charts of females taken from the Maryland Institute of Criminal Justice real case archives served as the raw data in the present research. The main purpose was to determine whether there are sex differences in the pattern of electrodermal responding. Two methods were used in this research to test for differences between electrodermal responses to lies and truthfulness. The first was the mathematical ratio between the size of the electrodermal reaction to the known lie key question in the POT test and the size of the electrodermal reaction to the last known truth question - LTR-L. The second was the mathematical ratio between the size of the electrodermal reaction to the known lie key question in the POT and the size of the electrodermal reaction to the first known truth question on this test - LTR-F. The results of the present research reveal significant differences between females and males in the LTR-L parameter indicating that females show a stronger contrast in response intensities than males between deception and truthfulness. There were no significant differences between males and females in the LTR-F parameter.

Some basic physiological differences exist between males and females. One of those differences appears in their reaction to a threat. Investigators of the brain have tried to attribute those differences to a distinction in the chemical structure of the brain between the sexes. These unique cognitive structures were probably developed because of the different functions of each sex; either biological function, or social evolutionary development.

Desmond Morris (1967) states that the male body of the great apes, as well as human beings, has prepared itself in its evolutionary development to attack, hunt and defend the tribe. Morris attributes some differences in physiological responding to this biosocial role. In his other book *The Human Zoo* (Morris, 1969), he states that the great ape male, as an opportunistic animal, has to look for his food in several places, and has the natural tendency to be always in readiness. The male's brain has a continuously demand for a strong changing stimulation from the environment.

The electrodermal response has been used to investigate psychophysiological differences between the sexes. Reim (1926) demonstrated that female skin goes through a faster depolarization process than male skin. Russel (1988) investigated differences in skin resistance recordings between the sexes, and the correlation between response patterns and gender. Russel's research question was: would males be detected lying more frequently than females, given the previously evidence of sex differences in cognitive styles, and their respective skin resistance differences? Working with the assumption that men would react more strongly and frequently than women, Russel tried to guess the sex of the subjects according to their electrodermal response patterns. The results showed that out of 40 male subjects, he correctly identified 38 males. Out of 40 female subjects he successfully guessed 39 females. His conclusion was that the subject gender could be ascertained by the characteristics of a subject's electrodermal responses.

This is a student paper, and was written as a partial requirement for completion of the polygraph course at the Maryland Institute of Criminal Justice. All writers contributed equally in this project. Comments and requests for reprints should be sent to Mr. Yaacov Friedlander, 48 Leah Street, Haifa, Israel, or via e-mail to fhhi201@uvm.haifa.ac.il.

Gazit and Daie (1980) in Israel tried to compare lying responses between males and females, using the duration of the electrodermal response to a known lie on a "demonstration test". They assumed that males would respond more strongly than females. Twenty males and females were randomly chosen from a population of suspects of minor thefts who agreed to have a polygraph examination. They found that males responded significantly stronger than females.

In this paper we used the electrodermal response to determine whether there were physiological differences between the sexes in their responses to a lie. The electrodermal response is considered as one of the indices that has the highest correlation with the body's emotional state and readiness (Gazit & Daie, 1980). The electrodermal response is considered the most accurate way of detecting a lie, and the most significant one (Bouverman, Klaider, Kobayashi & Vogal, 1965). The purpose of the present research was to investigate whether there are significant differences between males and females in their physiological reactions during truthfulness and deception. Our research was based on 60 real cases (30 males and 30 females), from the archives of the Maryland Institute of Criminal Justice (MICJ). Based on those cases, we used the demonstration test in which we measured the electrodermal response to the known lie question, the assumed weakest truth response (the last question on the demonstration test), and the assumed strongest truth response (the first question on the demonstration test). We chose the electrodermal response as a measurement, relying on previous research that showed the electrodermal response provides the best indicator of deception. The use of real cases made our research field research. By using lie, assumed weakest truth, and assumed strongest truth reactions, we intended to explore two questions:

A. Is there a significant difference between males and females in their responses to lies and their responses to truth?

B. Is the strongest truth response acting as an "orienting response", and would it produce a response similar to the lie response?

Hypotheses of the research

A. Differences will be found between males and females in the size of the ratio between electrodermal response to lie and electrodermal response to truth (Lie Truth Ratio - LTR) in POT test when lie response is defined as the response to the key question and truth response is defined as the response to the last question in the POT test.

B. No differences are expected to be found between males and females in the size of the ratio between electrodermal response to lie and electrodermal response to truth (LTR) in POT test when lie response is defined as the known lie response to the key question and truth response is defined as the response to the first question in the POT.

Method

Definitions and abbreviations

LTR: lie-to-truth ratio. The ratio of the size of the response to the critical item and the size of the response to the noncritical item in the POT test.

LTR-L: lie-to-truth ratio (last). The ratio of the size of response to the critical item and the size of the response to the last noncritical item in POT test.

LTR-F: lie-to-truth ratio (first). The ratio between the size of response to the critical item and the size of the response to the first noncritical item in the P.O.T. test.

MLTR: male lie-to-truth ratio.

FLTR: female lie-to-truth ratio.

MLTR-L: male lie-to-truth ratio (last).

FLTR-L: female lie-to-truth ratio (last).

MLTR-F: male lie-to-truth ratio (first).

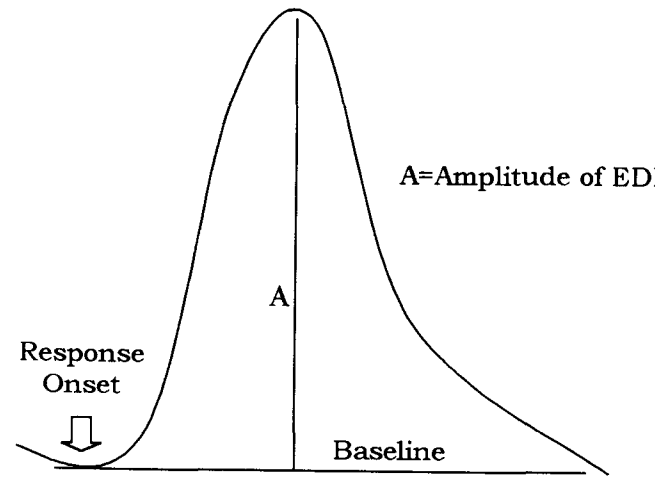
FLTR-F: female lie-to-truth ratio (first).

Subjects and Source of Data

Thirty POT tests for males and 30 POT tests for females were selected randomly from

polygraph subject archives in the MICJ. POT tests were specifically selected because they provided a known lie answer. The last question in the POT was selected as one of the known truth questions. It was assumed that the reaction to that question would be the weakest in the test due to its placement well after the critical item. The first question in the POT was selected as the second known truth. It was assumed that as the first question on the test it serves as to elicit the orienting-response, and render the strongest physiological reaction to a truthful answer. The use of the between-response within-subject ratio controlled for any contaminants such as differences in amplifier gains, location of sensors, or handedness.

Figure 1.



Parameters

The magnitude of the electrodermal response was the dependent measure in this study. The size of the electrodermal response is defined as the height of the phasic response measured from the horizontal baseline of the beginning of the response to the peak of the response in case there is a simple response (Figure 1).

In case of a complex response certain components of the secondary responses (B, C...) were added to the strongest response (A) (Figure 2).

Measurement Device

A simple millimetric ruler used to measure the size of the electrodermal responses.

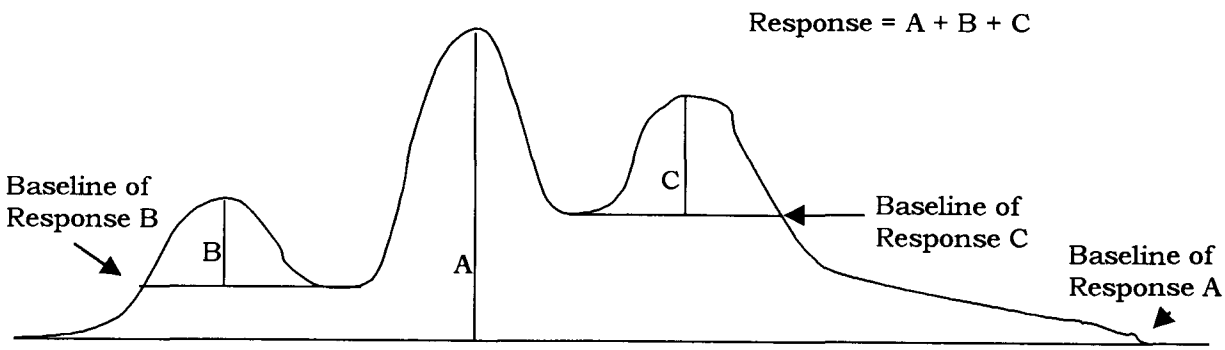
Statistical Analysis

The statistical test selected to analyze differences between males and females in the electrodermal response was the t test for two equal groups. The level of significance to test the research hypothesis is 0.05 (two-tailed).

Results

Appendix 1 is the raw size data of the of the electrodermal responses in millimeters in lie (L), first true (FT) and last true (LT) for both sexes: males (M) and females (F). Also shown are the data of the mathematical ratio between lie response and first true response (LTR-F) and between lie response and last true response (LTR-L) in males and females.

Figure 2.



The value of t_{DIFFER} obtained by the comparison of MLTR-L and FLTR-L was 2.30 ($p<0.05$ in a two-tailed hypothesis). Females show a greater lie-to-truth ratio (last) than males, and the difference reached significance. The data suggest that women are more sensitive than men to the difference between the known lie question in the POT and the last known truth question on this test.

The value of t_{DIFFER} obtained by the comparison of MLTR-F and FLTR-F was 0.36 ($p>0.5$, ns). The fact that the differences between MLTR-F and FLTR-F are not significant was not surprising. Males and females both react strongly to the first question in the POT, which is a known truth question, as well as to the critical item. The ratio between lie response and first truth response was similar in males and females, at about 2.0. This means that the reaction to the known lie question in the POT was, on average, twice the magnitude of the reaction to the first known truth question on that test for both sexes. The values of LTR-L is higher than the values of LTR-F for both sexes, indicating more similarity between the sizes of the responses to known lie question and first known truth "orienting response" than the similarity between the sizes of the responses to the known lie question and the last known truth question in the POT test.

Discussion

We have addressed in this research whether there are gender differences in the electrodermal response patterns during truth and deception. In all the research we have reviewed regarding gender differences the writers dealt with those differences only in the responses during deception. Most of those researches have found that males react stronger than females when telling a lie.

We evaluated within-test responses during both deception and truthfulness for the sexes. Our research found that females and males respond differentially: females show a stronger contrast in response intensities than males between deception and truthfulness. It is too early to interpret these findings as an indication that females are more sensitive than males to the differences between lie and truth but those results are quite interesting, and so, invite replication and extension. Among the possible explanations for the findings are: genetic physiological differences between the sexes, females' general stronger sensitivity to details and social atmosphere, or educational and cultural differences between the sexes.

The second hypothesis was that the responses to the critical item and the first presented item would be similar for both males and females. We did not expect any significant differences between the genders, and our hypothesis was supported by the data.

References

Boverman, D.M., Klaider, E.L., Kobayashi, Y. & Vogel, W. (1968). Roles of activation and inhabitation in sex differences in cognitive abilities. *Psychological Review,* 75, 23-50.

Desmond, M. (1967). *The Naked Ape.* London: Corgi Books.

Desmond, M. (1969). *The Human Zoo.* London: Cape.

Gazit, M., and Daie, T., (1980). *Sex differences in physiology activity,* unpublished dissertation.

Reim, H. (1926). *Handbook of Psychophysiology.*

Russel, S.J. (1988). Electrodermal activity, cognitive script and sex differences in a single blind study. *Polygraph,* 17(2).

Appendix 1

Size of electrodermal responses of males and females, in millimeters, and the ratios of the response magnitudes after lying to those of the first and last truthful responses

MALES					FEMALES				
L	FT	LT	LTR-F	LTR-L	L	FT	LT	LTR-F	LTR-L
6	10	3	0.60	2.00	22	17	18	1.29	1.22
43	21	13	2.05	3.31	41	51	13	0.80	3.15
17	40	43	0.43	0.40	41	19	11	2.16	3.73
27	33	15	0.82	1.80	104	117	23	0.89	4.52
6	1	11	6.00	0.55	30	10	12	3.00	2.50
19	21	45	0.90	0.42	14	11	1	1.27	14.00
5	19	5	0.26	1.00	35	3	10	11.67	3.50
12	4	1	3.00	12.00	25	26	11	0.96	2.27
7	36	21	0.19	0.33	92	67	3	1.37	30.67
80	33	49	2.42	1.63	103	60	23	1.72	4.48
54	16	10	3.38	5.40	61	44	19	1.39	3.21
12	9	2	1.33	6.00	18	40	7	0.45	2.57
18	14	4	1.29	4.50	1	1	1	1.00	1.00
7	1	4	7.00	1.75	5	69	1	0.07	5.00
91	44	82	2.07	1.11	21	40	3	0.53	7.00
33	26	21	1.27	1.57	44	12	15	3.67	2.93
66	14	23	4.71	2.87	5	9	3	0.56	1.67
9	5	19	1.80	0.47	2	3	2	0.67	1.00
8	21	9	0.38	0.89	69	49	48	1.41	1.44
64	33	38	1.94	1.68	101	135	50	0.75	2.02
56	20	83	2.80	0.67	31	10	10	3.10	3.10
31	28	27	1.11	1.15	52	33	5	1.58	10.40
68	45	36	1.51	1.89	56	17	17	3.29	3.29
19	18	13	1.06	1.46	28	25	8	1.12	3.50
24	23	19	1.04	1.26	50	34	20	1.47	2.50
7	8	4	0.88	1.75	59	22	12	2.68	4.92
6	9	6	0.67	1.00	36	41	8	0.88	4.50
4	2	1	2.00	4.00	40	19	11	2.11	3.64
11	5	7	2.20	1.57	31	14	28	2.21	1.11
6	7	4	0.86	1.50	8	1	1	8.00	8.00

The Irrelevant Question: A Descriptive Review

Norman Ansley

Abstract

This paper reviews the literature on irrelevant questions, using excerpts to provide definitions, describe their functions, placement in formats, and research.

Keywords: Control/comparison tests, definitions, irrelevant questions, norms, neutral questions, Relevant-Irrelevant Tests, test formats

Definitions

Krapohl and Sturm (1997) in their terminology reference define the irrelevant question as follows:

> A question designed to be non-emotion provoking. Irrelevant questions are most often placed in the first position of a question list, since an orienting response of no diagnostic value usually follows the presentation of the first question. In CQT formats it is also used after a relevant or control/ comparison question that has elicited a strong response so as to permit physiologic arousal levels to return to baseline before presenting another diagnostic question. Irrelevant questions are used in every type of PDD test. Also called norms or neutrals.

The Department of Defense Polygraph Institute describes the irrelevant question in a 1993 handout on test question construction as:

> 1. An irrelevant question is the first question asked during a polygraph examination. It is designed to allow the orienting response to habituate. It is a neutral question unrelated to the testing issue. It is worded so the examinee answers "Yes."
>
> 2. Irrelevant questions can be used to establish a norm when some type of distortion is occurring on the chart.

Stan Abrams, in his *Complete Polygraph Handbook* (1989), in the glossary, describes an irrelevant question as:

> A neutral question developed to bring the subject's level of reactivity down after arousal or placed in a position such as first on a test when a reaction would occur because of its position rather than the question itself.

James Allan Matte has a section defining terms in his textbook (1996):

> Neutral (Irrelevant) Question: It is of a non-stimulating nature. In a Zone Comparison test format, this type of question is usually used as the first question in the test, to absorb an examinee's orienting response and reduce general nervous tension. The examinee's place of birth, last name or first name is usually used for that purpose. In other techniques it is also used to create a "norm' pattern at the beginning and end of each chart. It is also used between relevant questions where necessary to terminate lingering reactions due to extraneous stimuli or extended thought process.

The author is a Life Member of the American Polygraph Association and president of Forensic Research, Inc. Comments and reprint requests should be sent to the author at 35 Cedar Road, Severna Park, MD 21146.

In their 1977 text, *Truth and Deception*, second edition, John Reid and Fred Inbau say:

> The primary purpose of the irrelevant questions is to ascertain the subject's normal reactions - his "norm" - under test conditions. Additional reasons for using irrelevant questions are:
>
> 1. To terminate a "lingering" type of reaction...
>
> 2. To nullify or terminate shock reactions due to noises occurring outside the examining room and heard by the subject during the test.
>
> 3. To provide an outlet for a relief response after relevant questions...
>
> 4. To separate the reactions to pertinent questions by using an irrelevant question in between.
>
> 5. To accentuate a deceptive response to a relevant question by the subject's failure to answer or by giving a delayed answer to the following irrelevant question...
>
> 6. To invite the scheming type subject to cause false or fraudulent reactions on irrelevant questions for the purpose of misleading the examiner into believing these reactions are greater, by comparison, than the reactions to the relevant test questions.

Fred Inbau, writing on the lie detector technique in 1942, before his adoption of control questions, wrote:

> The first two or three questions should be irrelevant to the matter under investigation. Questions such as "Is your name John Smith?" "Do you live in Chicago?" are helpful in ascertaining the nature of the subject's reactions to the test situation alone, irrespective of possible deception...

Fred Inbau gave the same definition and examples in his 1948 book, which included the control question test called the "Revised Questioning Technique." In the third edition of Fred Inbau's book in 1953, now listing John E. Reid as co-author, they state:

> The primary purpose of irrelevant questions ... is to ascertain the subject's norm under test conditions.

In a 1976 lesson plan on the relevant-irrelevant General Question Test the Army said about the irrelevants that they are used to determine the examinee's norm plus excitement level with a verbal stimulus.

Weir (1974), writing about the Relevant-Irrelevant Test, had a paragraph on terminology. He described the irrelevant question as:

> A polygraph question, of supposedly neutral impact, which does not relate to the matter under inquiry. Frequently called 'norm' questions or 'neutral' questions.

Harrelson (1964), writing about the relevant-irrelevant Keeler Technique, defines the irrelevant question as "A question formulated from information about which the subject would not normally lie, which does not pertain to the matter under investigation, and to which the examiner knows the answer or can reasonably infer same."

Among the items in the descriptions we note that an irrelevant question is:

1. Not emotion provoking;

2. Not related to the matter under inquiry; and

3. Is not a question the subject would normally lie to.

We note that some of the purposes of an irrelevant question are to:

1. Allow the orienting response to subside;

2. Allow other reactions to return to baseline;

3. Reduce general nervous tension;

4. Establish a norm or baseline; or norm plus excitement level;

5. Provide an outlet for relief from a response;

6. Separate reactions to relevant questions; and

7. Invite countermeasures.

Formats and Usage

Of the widely used techniques, the following employ an irrelevant question as the first question: Backster Zone Comparison (Backster, 1969), DoDPI Zone Comparison (DoDPI, 1991), DoDPI MGQT (DoDPI, 1989), Reid CQT (Reid & Inbau 1977), Relevant-Irrelevant (Harrelson, 1964; Minor, 1989; Weir, 1974), PCQT (Forman & McCauley, 1986), Arther CQT (Arther, 1987), and Marcy CQT (Matte 1996). In 1939 the Fordham Law Review published a paper by the late Reverend Walter G. Summers. His technique, used in research and criminal cases, paired his emotional standards questions (control/ comparison) with significant (relevant) questions, three of each on each chart, with the emotional standard in front of the significant in each pair. Summers wrote "These are interspersed among a larger number of non-significant questions, as, Are you wearing a black coat? and Did you eat breakfast this morning?" (Summers, 1939). His technique was used for many years by the New York State Troopers (Kirwan, 1952).

The Summers test format was a CQT, essentially a zone comparison. The number and placement of irrelevants was at the examiner's discretion. The relevant-irrelevant Keeler test also called for irrelevant questions as needed. Consider this question sequence by Leonarde Keeler who was testing Alfred de Marigny, acquitted of murdering Sir Harry Oakes.

1. "Is your name Alfred de Marigny?" "Yes."

2. "Do you live in Nassau?" "Yes."

3. "Do you know who killed Sir Harry Oakes?" "No."

4. "Have you had something to eat today?" "Yes."

5. "Did you kill Sir Harry Oakes yourself?" "No."

6. "Were you born in Nassau?" "No."

7. "Did you put your hand on that screen?" "No."

Keeler said there was no evidence of lying. Four of the seven questions were irrelevant, including the opening two. Note that irrelevant question number 6 was answered "No." This account was published in a book by Marshal Houts (1972), a friend of Keeler and no stranger to polygraph testing. A British author adds three relevant questions, lengthens the last one, but does not mention any irrelevant questions (Lessor, 1983).

Unlike the Summers and Keeler tests where irrelevants are placed as needed, fixed format tests put irrelevant questions where they expect they will be needed. An example is the DoDPI MGQT (1989):

1. Irrelevant
2. Irrelevant
3. Relevant
4. Irrelevant
5. Relevant
6. Control
7. Irrelevant
8. Relevant
9. Relevant
10. Control

The mixed series for the MGQT third chart is: 4-1-5-6-3-10-9-6-8-10. The MGQT test question sequence is based on the Reid technique. While three irrelevants are in the first two charts, only two are in the mixed series, both at the beginning.

The U.S. Air Force OSI MGQT differs from the DoDPI format in several ways. In fact, the OSI format looks more like a zone

comparison than a MGQT. When Michael Koll (1987) lectured on the OSI, his handout showed several samples from cases, and each one had only one irrelevant question, and that in the first position. Each of his examples was "Is your name_____ ?"

Prominent among the standard polygraph test formats is the Backster Zone Comparison Test. In his 1969 notepack, Backster lists the following three irrelevant questions for use in his you phase test. The Backster Exploratory Test lists only 13 (j) as the first question.

1. 14 (j) "Were you born in the United States?"

2. 13 (j) last name

3. 13 (k) first name

In 1993 Backster appeared on a panel on question formulation at an APA seminar (Smith, 1993). The prepared examples in the handout used one of the following three to begin his you Phase test:

1. 13. (a) "Is your first name _____?"

2. 13. (b) "Is your middle initial_____?"

3. 13. (c) "Is your last name_____?"

One of these three irrelevant questions was at the beginning of each Backster Zone Comparison S-K-Y test and each Backster Zone Comparison Exploratory Test.

Richard O. Arther has usually had an irrelevant question in the number one position. However, in 1961 he experimented with removal of the irrelevant question. He said he soon realized that was a mistake and that there must be an irrelevant as the first question. Although he comments that several irrelevant questions is an unnecessary feature, in 1985 he added the irrelevant question "Do you live in Canada?" as the last question, and it is answered "No" while the first question "Do you live in the United States?" is answered "Yes." (Arther, 1987). Matte (1996), in describing the Marcy Control Question Test format of 13 questions, notes irrelevants in positions 1, 4, 8, and 13, and like Arther, the last irrelevant is answered "No."

William M. Marston (1938) in describing a continuous type of test, had a most unusual placement of the irrelevant questions. After a short test in which apparently no questions were asked, a second record was made to "establish a record of b.p. behavior during irrelevant questions and answers." Marston said the examinee may be asked "Have you been in this room more than one minute?" "Have you drunk any liquor since you came into this room?" and so on. The relevant questions were all in the next chart. Marston added that some experts prefer to mix innocent and crucial questions, interspersing them in unpredictable order: "Have you smoked since coming into this room?" "Did you break into Mrs. Trimwell's apartment?" "What is your first name?" "Did you steal the diamond ring?" and so on. Note that one of Marston's irrelevant questions asked the examinee to answer with his first name.

The Canadian Police College student polygraph manual lists three irrelevant questions for use in their control question test. They are: "Is your name_____?" "Do you live in_____ ?" and "Were you born in_____?" In their ten-question zone format, the irrelevant questions are placed at questions one and eight (Koppang, 1985).

Harrelson (1964), writing about the relevant-irrelevant Keeler technique, ascribes four uses for irrelevant questions: to reduce the excitement level; to assist in returning the tracing to or toward the proper baseline in preference to a mechanical adjustment; to serve as an aid to interpretation of specific reactions; and to conserve the subject's ability to react. Harrelson gives ten examples of irrelevant questions:

1. "Is your first name_____?"

2. "Do you live in_____?"

3. "Do you drink_____ ?"

4. "Are you married?"

5. "Have you had anything to eat today?"

6. "Do you smoke?"

7. "Are you a citizen of the United States?"

8. "Were you born in_____?"

9. "Are you wearing_____?"

10. "Are you_____years old?"

This is a mix of identification and obvious irrelevant questions. Harrelson has some observations about what might cause specific reactions to these questions. He cautions not to use the topics of race, religion, politics, any topic subject to opinion or interpretation, or a condition subject to change like the weather. Harrelson encourages examiners to determine the cause of reactions to irrelevant questions.

Clarence D. Lee, an authority on relevant-irrelevant testing, wrote two texts on the topic (1943, 1953), but said little about irrelevant questions. His sample formats in both publications show irrelevant questions only as the first and second questions, and not elsewhere. They were worded:

1. "Is your name_____?"

2. "Do you live in_____?"

Weir (1974), writing about the relevant-irrelevant technique, gives four examples of suitable irrelevant questions that relate to identification and appear relevant to the examinee:

1. "Is your first name_____?"

2. "Is your middle name_____?"

3. "Were you born in the month of February?"

4. "Do you live in the City of Boston?"

Consistent, significant reactions to one or more of the irrelevant questions results in interrogation, the reactions making the question relevant. Obvious irrelevants such as "Do you smoke?" or "Are you now in the City of St. Louis?" are not used. Weir observes that these questions appear ridiculous, seem like a game, and do not pose a threat to the subject.

Paul K. Minor, writing about a modified relevant-irrelevant technique in 1989, offered examples from some of his cases. In a 12-question test he opened with two irrelevants, and had a third one at position 11. In a 13-question test he had irrelevants in positions 1, 2, 5, 8, and 12. In a 12-question test Minor had irrelevant questions in positions 1, 2, 4, 8 and 11. Some of his irrelevant questions were identification types about citizenship, residence, age, name and employment. Others were obvious, such as "Are you now in _____?" or "Is today Tuesday?" (Minor, 1989).

The Army General Question Test (GQT) began in 1951 as an RI format. It remained as such for many years, but was supplemented then replaced by control question tests, notably zone and MGQT. An Army lesson plan of 1976 states the correct terminology for this technique as any of the following: 1. Relevant-Irrelevant Technique, 2. General Question Technique, or, 3. the Keeler Technique. A Master Question List gave ten irrelevant questions, of which four were of the obvious type. The number and location of irrelevant questions in the RI format was not given. On a date unknown, the Army changed the GQT to a control question test employing disguised controls in positions 3 and 9, and irrelevant questions in positions 1, 2, 6, and 10. (USAMPS 1976, Crowe, Chimarys & Schwartz, 1995; Matte, 1996)

Formats and Usage Summary

Among the items presented in Formats and Usage, we note that:

1. All standard control question and relevant-irrelevant test formats open with an irrelevant question.

2. Some formats open with two irrelevant questions.

3. Some techniques leave the placement of irrelevants to the examiner.

4. Some fixed-sequence formats anticipate the need in placing irrelevants.

5. Two formats have irrelevants in the first and last position, and the irrelevant questions in the last position are answered "No."

6. Most irrelevant questions are designed to be answered "Yes."

7. There are two types of irrelevant questions, identity and obvious.

8. Indications of deception to identity questions result in an interrogation.

9. In some relevant-irrelevant formats the irrelevants are comparison questions.

Research

Kircher and Raskin (1986) were interested in how examinees perceived their reactions to control, relevant, and irrelevant questions. One hundred men served as subjects, with 50 stealing a ring in a mock theft. All examinees were promised $25.00 if they could produce a truthful outcome. After the test, each subject was presented with a sequence of all possible pairings of the questions and asked to choose from each pair the question he felt had produced the largest physiological response. Programmed guilty subjects reported having reacted most strongly to relevant questions; programmed innocent subjects reported that the control questions produced their strongest reactions; and both groups reported that irrelevant questions produced their weakest reactions.

Bob Roy Frisby (1979) was interested in whether the obvious or the identity irrelevants produced the least reactions. His subjects were 24 men and 24 women in police classes at Washington State University. He asked the same questions on each of two charts but with a different sequence. Each series contained half obvious and half identity questions. The first series asked about: first name, city of birth, smoking, year of birth, driving a car, breakfast today, campus address, and color of hair. The identity irrelevant questions produced fewer responses, 604, than did the obvious irrelevant questions, which produced 732.

Research Summary

From these two research projects we note that:

1. Examinees perceived their reactions to irrelevant questions were weaker than their reactions to control and relevant questions.

2. Identity irrelevant questions produce fewer responses than obvious irrelevant questions.

Conclusion

An irrelevant question is the opening question on each chart in all standard control/comparison and relevant-irrelevant test formats. It is there because the reaction caused by the opening question is not scored. Some formats include irrelevant questions within the pre-ordered list of questions. Other techniques put them in as needed. Within the test format they separate reactions to relevant questions, provide relief from relevant and control questions, and reestablish baselines. In some techniques irrelevant questions are also comparison questions. The answers to obvious irrelevant questions are almost always true, and the base rate of lying to identity irrelevants is low. The identity irrelevants are introduced as relevant, and if there are consistent significant reactions, the questions become relevant. Most irrelevant questions are worded to be truthfully answered "Yes," but some are worded to be truthfully answered "No." The significance of the difference, if any, is unknown.

With Weir's argument that all questions on a test should appear relevant, and the result of Frisby's research showing fewer reactions to identity irrelevant questions, the use of identity irrelevants in place of obvious irrelevants may be justified. Given the important functions of the irrelevant question, it is surprising that it has received such little serious attention.

References

Abrams, S. (1989). *The complete polygraph handbook.* Lexington, MA: Lexington Books.

Abrams, S. (1977). *A polygraph handbook for attorneys.* Lexington, MA: Lexington Books.

Arther, R. 0. (1987). Irrelevant questions. *Journal of Police Science,* 22(2) 1-4.

Backster, C. (1969). *Standardization notepack and technique guide, Backster Zone Comparison guide,* Rev. ed., San Diego, CA: Backster Associates.

Crowe, M. J., Chimarys, M. & Schwartz, J. (1995). The GQT polygraph test: Scoring and validity. *Polygraph,* 24(3) 234-236

Department of Defense Polygraph Institute (1989). *Modified General Question Technique Summary.*

Department of Defense Polygraph Institute (1993). Test Question Construction.

Department of Defense Polygraph Institute. (1991). Zone Comparison Technique lesson plan.

Forman, R. F. & McCauley, C. (1986). Validity of the positive control polygraph test using the field practice model. *Journal of Applied Psychology,* 71(4) 691-698. Reprinted in *Polygraph,* 16 (2) 145-160.

Harrelson, L. H. (1964). Keeler Polygraph Institute training guide. Chicago, IL: The Keeler Polygraph Institute.

Houts, M. (1972). *King's X, common law and the death of Sir Harry Oakes.* New York, NY: William Morrow & Co.

Inbau, F. E. (1942). *Lie detection and criminal interrogation.* Baltimore, MD: Williams & Wilkins.

Inbau, F. E. (1948). *Lie detection and criminal interrogation,* second edition, rev., Baltimore, MD: Williams & Wilkins.

Inbau, F. E. (1953). *Lie detection and criminal interrogation,* third edition, rev., Baltimore, MD: Williams & Wilkins.

Kirwan, W. E. (1952), Letter to N. Ansley, October 16, on the letterhead of the New York Troopers Scientific Laboratory.

Koll, M. (1987). Question formulation. Handout at a meeting of the Wisconsin Polygraph Association.

Koppang, C. E. (1985). Polygraph examination student manual. Ottawa: Canadian Police College.

Krapohl, D. & Sturm, S. (1997). *Terminology reference for the science of psychophysiological detection of deception.* Chattanooga, TN: American Polygraph Association.

Leasor, J. (1983). *Who killed Sir Harry Oakes?,* Boston, MA: Houghton Mifflin Co.

Lee, C. D. (1943). *Instruction manual for the Berkeley psychograph.* Sacramento, CA: Lee & Sons.

Lee, C. D. (1953). *The instrumental detection of deception, the lie test.* Springfield, IL: Charles C. Thomas.

Marston, W. M. (1938). *The lie detector test.* New York, NY: Richard R. Smith. Reprinted by the American Polygraph Association, 1989.

Matte, J. A. (1996). Forensic psychophysiology using a polygraph. Williamsville, NY: J.A.M Publications.

Minor, P. K. (1989) "The relevant-irrelevant technique," in S. Abrams *The complete polygraph handbook.* Lexington, MA: Lexington Books.

Reid, J. E. & Inbau, F. E. (1966). *Truth and deception: The polygraph ("lie detector") technique.* Baltimore, MD: Williams & Wilkins.

Reid, J. E. & Inbau, F. E. (1977). *Truth and deception: The polygraph ("lie detector") technique,* second edition, rev. Baltimore, MD: Williams & Wilkins.

Smith, M. L. (moderator) (1993). "Polygraph test question formulation...". A panel discussion. Handout at the Seminar of the American Polygraph Association, Newport Beach, CA.

Summers, W. G. (1939). Science can get the confession. *Fordham Law Review,* 8, 335-354.

United States Army Military Police School (July 1976). Lesson plan "General Question Test (GQT) Construction."

Weir, R. J., Jr. (1974). In defense of the relevant-irrelevant polygraph test. *Polygraph,* 3(2) 119-166.

The Utility of Polygraph Exams in Unknown Paternity TANF Cases

Michael T. Hanna & Deborah Welter

Abstract

This article summarizes a program involving 96 cases of Temporary Assistance to Needy Families (TANF) benefits claimed by recipient mothers involving alleged or unknown fathers. Polygraph exams were offered to these women and the results demonstrate the effectiveness not only of polygraph as a technique but more so the utility of polygraph as a screening device. A majority of the recipient mothers, 51 of 96, offered information that included the identity of the fathers' of their children, as well as other specific information that led to the identification of these previously unknown fathers, well before an examination was ever scheduled. In all, when the examination results and all other aspects of the process are included, but pending cases excluded, 93% of all cases were resolved or closed.

Keywords: Investigations, paternity, utility

Most polygraph examiners are well aware of the value of the polygraph technique. Polygraph techniques are implemented in a variety of situations and for many diverse reasons. The technique is used for criminal and background investigations, as a screening tool in sensitive intelligence cases, and most recently as a monitoring/assessment tool for convicted sex offenders (Abrams & Abrams, 1993).

There has always been controversy surrounding the polygraph technique. Most critics point to the lack of admissibility in most state and federal courts as evidence of its unreliability. The validity and reliability of polygraph is not the subject of this article. Recent studies, however, have demonstrated both the reliability and validity of the technique. The official publication of the American Polygraph Association, *Polygraph*, summarized many of the most respected and reputable studies conducted in recent years (Ansley, 1997).

The utility of the polygraph technique is this article's frame of reference. In other words, what does the technique achieve as measured both before and after the examination, as well as what it achieves as part of the investigative process? How does the mere presence or possibility of a polygraph affect the resolution of these type of cases?

Most polygraph examiners have been aware of the screening effect of polygraph for many years. It is not discussed formally as part of the training process, nor does one find it in the polygraph literature. This is primarily due to the fact that the effect is beyond the scope of normal training, as well as the scope of most polygraph research. We are not evaluating polygraph as a device or technique itself, but utility it has on the solution of these cases.

Before the passage of the Employee Polygraph Protection Act (EPPA, 1988), as the General Manager of a contract security

Michael T. Hanna J.D. is an Assistant Professor, Criminal Justice & Legal Studies at Missouri Western State College, and a practicing polygraph examiner. Deborah Welter B.S. is a Paralegal/Supervisor with the Buchanan County Prosecuting Attorney's Office. Comments and request for reprints should be sent to Michael T. Hanna, Department of Criminal Justice and Legal Studies, Missouri Western State College, 4525 Downs Dr., St. Joseph, MO 64507-2294.

company the first author supervised the selection and screening of qualified applicants for various security positions. We utilized polygraphs as part of the pre-employment selection process. The process that we used involved asking the applicant to complete a written questionnaire concerning personal background information. The applicant was asked about prior theft, drug usage, dismissals by former employers and other information that an employer may have an interest in. The individuals completed these questionnaires the day they were scheduled for the examinations. Fewer than half of the applicants would actually take the test. The value of this process is that we saved a vast amount of time and money by training and hiring people who were more likely to be better candidates to become security guards. The refusal to take the examination, as well as the admissions and confessions obtained before the examination is administered, shall be referred to here as the polygraph "screening effect".

In Buchanan County, Missouri; a mixed urban-rural area in Northwest Missouri, we saw similar results as those discussed above, regarding the screening effect of polygraph. This program was initiated by an aggressive Child Support Enforcement Office headed by the Buchanan County Prosecutor, Dwight Scroggins. Scroggins has spearheaded an effort that has seen child support collections triple to over $7.2 million in 1997, from $2.0 million in 1991.

Changes in the Federal Statutes have also initiated this program of increased efforts to identify reported unknown fathers. On August 22, 1996, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRA) became law. Most states have replaced their Aid to Families with Dependent Children (AFDC) programs with a new program called Temporary Assistance for Needy Families (TANF). States now have a great deal of latitude to determine who will be eligible for and under what conditions assistance will be provided (Roberts, 1996).

Under the old law, recipients were required to cooperate by providing the name of the absent parent, and any other information about him known, or reasonably available to the mother. The primary method for the recipient to meet this requirement was for her to sign an attestation that she has revealed all known information. The state had to accept this attestation as full cooperation, unless the state had credible evidence that she was lying (Roberts, 1996)

Both state agencies and mothers had difficulty with this system. The recipient mothers found that the scope of the questionnaires went well beyond paternity issues. State support personnel objected to the attestation as the sole measure, in many cases, of the recipient mother's cooperation. Their primary objection was the potential for lying on the attestation. Under the PRA, states are now free to enhance the traditional federal definition of "cooperation".

Of significance to Missouri and many other states is that under Section 103 of the PRA the states are under pressure to increase their paternity establishment rates, or face a reduction in their TANF block grant funds. States that do not enforce non-cooperation sanctions can lose as much as 5% of the TANF block grant in the next fiscal year (US Congress, 1996). Also, the national paternity establishment percentage (PEP) was increased from 75% to 90%. Those states between 75% and 89% must show an improvement of 2% annually to avoid any sanctions.

On a national level, the efforts to identify these unknown fathers is a part of the government's main collection effort, the Child Support Enforcement Program. Established in 1975, the program had grown to the point that in fiscal year 1993 collections were received in 873,000 AFDC cases. The collected money raised the income level of 242,000 families, enough to remove them from the welfare roles. It saved 12% of all AFDC payments nationally, a total of over $1.6 billion (US Congress, 1994). Clearly, there is an economic necessity as well as a statutory mandate to increase collection efforts targeting unknown father cases.

Given this background, let us examine how a program in Buchanan County assisted the state in meeting these economic goals and statutory mandates. The Child Support Enforcement Unit selected 96 cases that were listed in their case files as questionable. The main criterion for a questionable designation,

in addition to listing the father as "unknown", was the lack of information provided by the recipient mother. Other factors that led the caseworker to suspect that the recipient mother may be withholding full cooperation, were also taken into consideration.

All 96 recipient mothers were informed that polygraph examinations may be requested of them to verify the "unknown" status of the father of their child receiving benefits. Within the first few months, 51 of the 96 cases were resolved or closed after the notification. The "resolved" status indicated that the mother provided the name and other identifying information relative to the previously unknown father. The "closed" cases are mothers that refused to receive any further benefits. In the next phase, a pre-polygraph interview and attestation was initiated with 27 cases (See Table A.) The 18 cases listed as pending were listed in this manner for a variety of reasons, primarily because the cases were not in an investigative status at the time of this study.

Table A.
Questionable Cases

Number	Status
22	Closed
29	Resolved
27	Pre-Polygraph Interview
18	Pending
96	Total

The recipient mother was contacted and interviewed about the identity of the alleged unknown father in 27 cases. During this interview she was asked to complete a simple 12-question questionnaire, and sign an attached agreement to submit to a polygraph examination concerning the identity of the father of the listed child. Both documents were attested and notarized.

Of these 27 cases, four were closed as a result of the recipients' refusal of any further benefits. Another nine recipients named or clearly identified the father during this interview. Six of the recipients refused to submit to a polygraph. These six were summoned into court and asked about the identity of the unknown fathers under oath. Of these six, five provided the name and identifying information for the previously unknown father. The one who did not was dismissed on other grounds. A total of eight were given polygraph examinations. There was an even split of four NDI (No Deception Indicated) calls and four DI (Deception Indicated) decisions for the eight examinees. Of the four who were deemed NDI, one did so by providing the name of the previously unknown father before the exam. Of the three who were called DI, three offered the name of the father or other identifying information relative to another possible father after the examination. (Table B)

Table B.
Cases of Polygraph Interviews and Exams

Number	Status
4	Closed
9	Identified father prior to test
6	Refused exam
5	Named father in court
1	Dismissed on other grounds
8	Polygraph examinations
4	NDI (1 pretest confession)
4	DI (3 posttest confessions)
27	Total

Table C provides a summary of the total cases that were resolved or closed as a percentage of the total cases. The resolution or closed rate was 93% of the total when the pending cases are excluded.

Table C.
Cases Resolved or Closed

Cases	Status
51	Resolved prior to second interview
13	Prior to polygraph
5	During testimony
4	During polygraph session
73 of 96	76% of total closed/resolved
73 of 78	93% of total closed/resolved (pending cases excluded)

Conclusion

This article does not seek to address the validity and reliability of the polygraph technique. The point here illustrated is that polygraph as a tool or a utility has a great deal of value. The greatest value in our project was as an initial screening tool where 51 of the total cases were resolved by merely suggesting polygraph to the mothers. The authors do not attempt to know the underlying motivations as to the reasons individuals feel compelled to be honest, and when. Clearly when one is pressed or tested as to one's honesty, most people appear prepared to provide the withheld information. The mere offering of the exam is not enough, although tempting for time and convenience sake. If these recipient mothers discovered that exams were not being conducted they would quickly realize that the threat of a polygraph exam had no meaning.

The constitutional and ethical questions that this technique pose are beyond the scope of this article. In this program, we offered these exams as a voluntary process and all of the individuals signed a consent form prior to the examination. The EPPA addresses only private employers for pre-employment exams as well as private employers for theft and other internal investigations. Governmental entities are specifically excluded (EPPA, 1988). One must also be aware that the receipt of TANF funds under false or misleading circumstances is potentially a criminal act. Criminal investigative activity is an appropriate forum for polygraph examinations.

Each case that is solved in this manner saves the state and federal government thousands of dollars in benefits. It places the financial burden upon the missing parent who should be financially responsible for the child. The cost of conducting a relatively small number of exams is greatly outweighed by the savings in time and money to the government. The saving would be even greater if the state patrol or local law enforcement examiners can be persuaded to conduct these exams. However, most law enforcement examiners are already spoken for. Private examiners are available in most jurisdictions. The state polygraph associations will provide a list of members and contact points to assist locating a private examiner. Like most professions, the best examiners are to be found through word of mouth.

In conclusion, the polygraph technique proved to be an effective, inexpensive utility in the resolution of a significant percentage, 93%, of unknown father TANF cases in Buchanan County, Missouri.

References

Abrams, Stanley and Abrams, Jared (1993) Polygraph Testing of the Pedophile, Ryan Gwinner Press, Portland, OR

Ansley, Norman, "The Validity and Reliability of Polygraph Testing", Polygraph, Vol. 26, No. 4, 1997, pp 215-239.

29 CFR Section 801-880, Employee Polygraph Protection Act, 29 U.S.C. 2001-2009, 1988.

Roberts, Paula, "A Guide to Implementing the Child Support Cooperation Policies: Provisions of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996," Washington D.C.: Center for Law and Social Policy, November 1996.

U.S. Congress, House Committee on Ways and Means, Overview of Entitlement Programs, 1994 Green Book, Washington, D.C., U.S. Government Printing Office, 1994.

U.S. Congress, House Committee on Ways and Means, Overview of Entitlement Programs, 1996 Green Book, Washington, D.C., U.S. Government Printing Office, 1996.

The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiners' Decisions

Eitan Elaad[1], Avital Ginton[1] and Gershon Ben-Shakhar[2]

Abstract

The present study deals with the question of whether judgments made by experts working in familiar contexts are affected by prior expectations and beliefs. Two experiments in which prior expectations were manipulated were designed to determine whether and to what extent polygraph examiners are affected by their prior expectations when analyzing and interpreting polygraph charts. Prior expectations affected the examiners' judgments when the polygraph charts did not include clear indications of guilt or innocence, but when the objective physiological evidence included strong indications which clearly contradicted the examiner's expectations, judgments were not affected by these expectations. Theoretical and practical implications of these results are discussed.

Keywords: Control question technique, hindsight bias, hypothesis-confirmation bias, outcome knowledge, polygraph, prior expectations

Detecting deception and discriminating between truthful and deceptive individuals are extremely important goals in modern society. They are important both from the criminal justice perspective (e.g. to find out whether a given person is deceptive regarding a specific, usually criminal, event) and for personnel selection purposes (e.g. to discriminate between honest and dishonest individuals among a group of job applicants). It is, thus, not surprising that scientists and forensic experts have been attempting for many years to develop instruments and methods for the purpose of detecting deception.

One approach to this problem, which has received considerable attention, is psychophysiological. Various methods of psychophysiological detection of deception have been developed since the beginning of this century (e.g. Ben-Shakhar and Furedy, 1990; Lykken, 1981; Raskin, 1989; Reid and Inbau, 1977), but the one method which is clearly the most common in field practice is the Control Question Technique (CQT).

Detailed descriptions of this method can be found in various sources (e.g. Reid and Inbau, 1977; Raskin, 1989), and therefore we shall provide here only a brief description of the CQT, as typically used in the criminal investigation context. The CQT is administered in several stages: First, the examiner becomes familiar with the facts of the case by reading the written report and by speaking directly to the police investigator who ordered the examination. Typically, relevant background information, such as the suspect's past criminal record, is made available to the examiner. During the next stage the examiner conducts an extensive pre-test interview in which the examinee is given the opportunity to talk about the offense and to present his or her version of the case. The series of

[1]Division of Identification and Forensic Science, Israel National Policy Headquarters
[2]The Hebrew University of Jerusalem, Israel

Acknowledgements

We wish to thank the polygraph examiners of the Israel Police for their assistance in the execution of this study. In addition, we are grateful to Maya Bar-Hillel, Gideon Keren, Benjamin Kleinmuntz, and an anonymous reviewer for their helpful comments on an earlier version of this manuscript.

This article was originally published in the *Journal of Behavioral Decision Making*, Vol 7 (1994), pp 279-292, and is reprinted here with permission of Dr. Elaad and the publisher, John Wiley & Sons Limited.

questions, to be asked later in the actual examination stage of the polygraph test, is formulated during this pre-test interview through an interaction between the examiner and the examinee. The examiner discusses the formulation of the questions with the examinee and ensures that he or she understands them and can give a direct 'yes' or 'no' answer to each question. The examiner explains the testing procedure and informs the examinee that the examination is voluntary. The next stage is the actual examination stage during which the examinee is attached to the polygraph.

During the examination stage a series of questions is presented to the examinee while continuously measuring the various physiological reactions. The questions are of the following three general types:

(1) Relevant questions - directly crime-relevant questions of the 'Did you do it?' type (e.g. 'Did you break into Mr. Jones's apartment last Friday night?').

(2) Control questions - focusing on general, non-specific misconducts, of a nature as similar as possible to the issue under investigation (e.g. 'Have you ever taken something that did not belong to you?').

(3) Irrelevant questions - focusing on completely neutral issues (e.g. 'Are you sitting on a chair?').

These are intended to absorb the initial orienting response evoked by any opening question, and to enable rest periods between the more loaded questions. Typically, the whole question series is repeated three or four times. The inference rule underlying the CQT is based on a comparison of the responses evoked by the relevant and the control questions. Deceptive individuals are expected to show more pronounced responses to the relevant questions, whereas truthful individuals are expected to show the opposite pattern of responsivity (i.e.; more pronounced responses to the control questions).

The CQT raised a major controversy, revolving around its rationale and inference rule, as well as around the empirical question of its validity (e.g. Ben-Shakhar and Furedy, 1990; Furedy and Heslegrave, 1989; Lykken, 1974, 1978; Raskin, 1982, 1989). In addition, some researchers have argued that the polygraph examiner knows a great deal more than what is revealed through the physiological data gathered during the examination (e.g. Ben-Shakhar et al., 1986; Ben-Shakhar and Furedy, 1990). From our description of the typical CQT examination it is clear that CQT polygraph examiners are exposed to a great deal of non-physiological information, such as information provided to them by the investigator and impressions formed during the pre-test interview and during the test itself. It is impossible to differentiate between the impression formed by this prior information and those gained from the purely physiological data obtained during the test phase of the polygraph examination procedure. This feature, which has been labeled by Ben-Shakhar et al., (1986) as 'contamination', implies that judgments and conclusions derived from the physiological information are contaminated with various kinds of non-physiological information. Contamination is inherent to the CQT, because this procedure is not limited to the psychophysiological data, but rather relies on the whole examiner-examinee interaction, including the pre-test interview. Furthermore, CQT polygraphers believe that it is essential that the same examiner construct the questions during the pre-test interview and administer them during the test phase of the examination. Typically, the same person evaluates the polygraph charts and draws the final conclusion on the basis of all the available information, although in some polygraph agencies (including the Israeli Police) the procedure includes an additional, 'blind', evaluation of the polygraph charts. One implication of the contamination feature of the CQT is that the weight of the strictly physiological information in the polygraph examiner's conclusion is not known, and in principle can be very small.

It should be noted that contamination does not necessarily lead to an increased error rate and a decreased validity. In principle, the non-physiological information can be accurate and valid, and polygraph examiners are trained to utilize all the information they can in an attempt to improve the accuracy of their conclusions. However, contamination may

introduce various biases, because the prior information may affect the formulation of the questions and the way they are presented to the suspects. For example, when examiners believe that a given examinee is deceptive, they may present the relevant questions in a way that may affect the results in the expected direction. On the other hand, when examiners are under the impression that their suspect is truthful, the control questions might be overemphasized. In addition, the prior information may affect the judgments of polygraph examiners even when they evaluate the polygraph charts in themselves. This is a particularly likely possibility because the procedure of chart evaluation is often subjective, and precise quantification rules are not available in many polygraph agencies.

Ben-Shakhar (1991) raised the possibility that judgments of CQT examiners are affected by prior expectations. He hypothesized that polygraph examiners often generate a hunch regarding the veracity of their examinee, on the basis of the non-physiological information that was available to them (e.g. the pre-test interview, and information gathered from previous, non-polygraphic interrogations). The test-specific information is then used to test this prior hypothesis, but the hypothesis-testing process is influenced by a confirmation bias or by a primacy effect (i.e. when the judge is supposedly considering the 'objective' data neutrally for the purpose of diagnosis, or evaluation, he or she is in fact searching it for confirmation of the initial impression or the prior hypothesis which the judge entertains before looking at the data). Some data sources are sufficiently simple or well defined that they hardly lend themselves to variable interpretation. However, in the case of CQT-polygraphy, where the rich and complex physiological data are subjectively evaluated by examiners, rather than mechanically quantified, it can readily be distorted if the search process is biased and precommitted. Moreover, under these circumstances a biased search is likely to produce supportive findings especially if it is untempered by critical attempts to falsify the initial hypothesis or to pit it against some competing alternative. The richness and vagueness of the information increase the likelihood of finding some confirmatory evidence. Indeed, it is possible that most polygraph charts contain some confirmatory information for almost any possible hypothesis.

The literature on human judgment and decision making describes several biases that might be relevant to judgments made by CQT polygraphers. For example, Evans (1989) mentions belief bias (i.e. the effect of prior beliefs and attitudes on reasoning and judgment), hindsight bias (i.e. the effect of outcome knowledge on estimates of a priori probabilities of events), and confirmation bias (i.e. the tendency to seek information consistent with one's current belief or hypothesis). The hypothesis formulated by Ben-Shakhar (1991) is based on a combination of the belief bias and the confirmation bias, because it postulates that polygraph examiners typically develop a belief or hypothesis based on information gathered before the administration of the CQT, and that the interpretation of the charts is influenced by that prior belief through a focus on physiological information which is consistent with it.

As noted by Klayman and Ha (1987), the term 'confirmation bias' appears in the literature with different meanings. Our use of this term differs from its original use by Wason (e.g., 1960, 1968) and his followers. In the context of polygraph chart interpretation, all the physiological information is available to the examiner, and the optimal use of this information is by proper weighting of all the cues, whether confirming the initial hypothesis or not. The confirmation bias was demonstrated mainly in the context of social perception and interaction (e.g. Darley and Gross, 1983; Snyder and Swann, 1978a,b). These demonstrations utilized artificial experimental procedures with university students serving as subjects, and therefore it is not clear whether their results would generalize to situations where experts (e.g. polygraph examiners) are performing familiar tasks in realistic situations.

There are some demonstrations showing that judgments of experts may he affected by certain types of biases, For example, Detmer et al. (1978), showed that judgments of surgeons are affected by the hindsight (i.e. outcome knowledge), and Arkes

et al (1981) reported a similar bias with physicians. Stenson, Kleinmuntz and Scott (1975) demonstrated that experts were biased by prior instructions when sorting MMPI profiles. On the other hand, Smith and Kida (1991) suggested that biases in human judgment may have been overgeneralized, and people can be effective judges when operating in natural, familiar contexts. Christensen-Szalanski and Bushyhead (1981) produced mixed results and demonstrated that physicians overestimated the patients' probability of pneumonia but were sensitive to relative differences in the predictive value of symptoms, and appeared to use base-rate information correctly when making clinical judgments.

The main purpose of the present study is to determine whether and to what extent prior expectations affect judgments and decisions made by professional CQT-polygraph examiners working in natural, realistic setups. Two experiments in which prior expectations were manipulated by providing some examiners with outcome knowledge (e.g. telling them that the suspect had eventually confessed) were conducted. In these experiments, which focused on the effect of outcome knowledge on chart evaluation and interpretation, examiners were asked to score and assess polygraph charts from previous examinations. These two experiments are similar in many ways to studies that dealt with the hindsight bias (e.g. Fischhoff, 1975; Slovic and Fischhoff, 1977), and demonstrated that reporting the outcome of a historical event or a scientific experiment increased the perceived likelihood of that outcome.

Experiment 1

Method

Subjects

Subjects were ten polygraph examiners employed by the Israel National Police. Six of them had at least 3 years of experience in scoring polygraph records and the other four had about one.

Stimulus material

The first three charts (i.e. three repetitions of the question list) of 14 real-life criminal polygraph records were selected to be used in the present study. All records resulted from CQT polygraph examinations conducted by field-trained examiners who used Lafayette field model polygraphs, which recorded the following physiological measures:

(1) Dual respiration (thoracic and abdominal), recorded by pneumatic tubes positioned around the thoracic area and abdomen.

(2) Skin resistance responses (SRR), recorded by two stainless steel electrodes attached to the volar side of the index and fourth fingers of the subject's left hand.

(3) Cardiovascular activity, recorded by a pneumatic blood pressure cuff positioned around the upper right arm, and inflated to a pressure between 40-50 mm Hg.

All polygraph records chosen for this study indicated inconclusive results.[1] To

[1] The process of evaluating polygraph charts at the Polygraph Unit of the Israeli Police Force follows the numerical scoring procedure which was proposed originally by Backster (1963). According to this procedure, two or three pairs of Relevant-Control questions are identified in each polygraph chart, and numbers (-3, -2. - 1, 0, 1, 2, 3) are assigned to each pair for each physiological measure. The absolute value of the assigned number reflects the magnitude of the difference between the responses evoked by the two questions within the pair (e.g. -3 or +3 reflect a very large difference, - 1 or + 1 reflect a small difference and 0 reflects no difference), and the sign of the assigned number reflects the direction of the difference such that positive numbers are associated with a pattern of larger physiological reactivity to the control question, and negative numbers reflect the opposite pattern. These numbers are then summed up across question pairs, across physiological measures and across polygraph charts to yield a global score. Thus, if for example a polygraph examination is based on three charts and three physiological measures and if two pairs of Relevant-Control questions are identified for each chart, then the global score ranges between -54 and +54. Typically the following classification rule is used. If the global score exceeds +5, an NDI (no deception indicated) classification is reached; if the global score is less than -5, the polygraph record is classified as DI (deception indicated); and if the global score ranges between -5 and +5, the record is classified as inconclusive.

determine the inconclusive nature of the records, they were scored blindly by three independent examiners, who were not scheduled to take part in the experimental phase. Each of these examiners had 15 years' experience in scoring polygraph records, and therefore their scoring may serve for comparative purposes as a control for the experimental conditions. The global scores assigned by these control examiners were averaged, and it turned out that all 14 averages were within the 'inconclusive zone' (i.e. between -5 and +5), although in two cases one of the three scorers assigned a global score larger than +5. To estimate the interscorer reliability, a Pearson correlation coefficient was computed between the global scores assigned by each pair of independent scorers. The three coefficients were 0.66, 0.66, and 0.78. Typically, reliabilities among polygraph examiners using numerical scoring procedures are much higher (e.g. Raskin, 1989). However, in this experiment interrater correlations were computed on the basis of a restricted range because only inconclusive charts were used, and this might account for the attenuation in the observed correlations.

Procedure

The 14 records were arbitrarily divided into two sets of seven records each. Each set was distributed to the ten polygraph examiners for a blind numerical scoring. The examiners were requested to score the charts at their own pace. To manipulate the examiners' expectations, each group of records was accompanied by different outcome information: In the Guilt-expectation condition the examiners were told that the examinee ultimately confessed to being responsible for the crime, while in the Innocence-expectation condition they were informed that another person had confessed to that crime. Five polygraph examiners scored one set of seven records under the Guilt-expectation condition, while the other five examiners scored the same set of records under the Innocence-expectation condition. The order of the two conditions was counterbalanced across examiners.

To increase the credibility of the instructions, all the examiners participating in this experiment were told that some of the records they were asked to score were mistakenly interpreted by the original examiner. An analysis of these mistakes revealed that some of them might have been prevented had the original examiner considered only the first three charts of each examinee and avoided the fourth chart. The examiners were further told that the purpose of the study was to determine whether reducing the records to three charts would increase the accuracy of blind post-test interpretations.

The scoring was blind with respect to the nature of the case, to the background information about the case, to the specific questions presented in the polygraph test, and to the test's results. The only information provided to the examiners was about the type of each question (i.e. relevant, control, or irrelevant). However, as part of the experimental manipulation, examiners were provided with information about the guilt or innocence of the examinee (i.e. whether the examinee had confessed, or whether he or she was exonerated through the confession of another person suspected of committing the same crime).

Results

Each polygraph record was analyzed by each of the ten examiners using the numerical scoring procedure. The global scores assigned to the polygraph records were classified into three categories: DI (deception indicated), if the global score was less than or equal to -6; NDI (no deception indicated), if the global score was greater than or equal to +6; and inconclusive if the global score was between -5 and + 5. Exhibit 1 presents the frequencies of the three classifications made by each examiner under each experimental condition, as well as the classifications made by the three control examiners under the control condition of no outcome knowledge.

Exhibit 1 reveals that under the Innocence-expectation condition 21% (15 out of 70) of the polygraph records were classified as NDI, whereas none was classified as DI. Exhibit 1 reveals that under the Innocence-expectation condition 21% (15 out of 70) of the polygraph records were classified as NDI, whereas none was classified as DI. In comparison, the frequency of NDI

Exhibit 1
Frequencies of the classifications made by each examiner under the two experimental conditions and the control condition of Experiment 1

Experimental condition	Guilt expectation			Innocence expectation			Control		
Examiner	NDI	Inconclusive	DI	NDI	Inconclusive	DI	NDI	Inconclusive	DI
1*	0	6	1	4	3	0			
2*	0	7	0	1	6	0			
3	0	7	0	2	5	0			
4	1	5	1	1	6	0			
5	1	6	0	0	7	0			
6*	1	6	0	0	7	0			
7*	0	7	0	1	6	0			
8	1	6	0	2	5	0			
9	1	5	1	2	5	0			
10	1	6	0	2	5	0			
11							1	13	0
12							1	13	0
13							0	14	0
Across examiners	6	61	3	15	55	0	2	40	0

NDI = no deception indicated; DI = deception indicated.
*Less experienced examiners.

classifications under the control condition was only 5% (two out of 42). Under the Guilt expectation condition, 8.6% and 4.3% of the records were classified as NDI and DI, respectively. The distribution of these classifications does not allow for statistical tests (the vast majority of the cases were classified as inconclusive, and the frequencies of the other categories are much too small). Therefore we conducted statistical tests on the continuous scale of the global scores. Because different examiners might differ in their scoring tendencies, a within-examiner standardization of the global scores was used (Ben-Shakhar, 1985). Each global score was transformed into a standard score relative to the mean and the standard deviation of the individual examiner's global score distribution.

The average Z scores across examiners within each experimental condition was computed for each polygraph record. Exhibit 2 displays the mean Z scores for each polygraph record under each of the two experimental conditions and under the control condition. For 13 out of the 14 records the difference between the average Z scores was in the expected direction (i.e. more positive scores under the Innocence-expectation condition than under the Guilt-expectation condition). A matched-group t-test was conducted to compare the mean Z scores obtained under the Innocence- expectation condition with those obtained under the Guilt-expectation condition. This comparison yielded a statistically significant difference[2] ($t_{(13)} = 5.04$).

A similar analysis was conducted for the examiners. A mean Z score was computed for each of the ten examiners, under each experimental condition across the seven polygraph records scored by the examiner. These means, which were computed separately

[2] A significance level of 0.05 was used in all statistical analyses reported in this study.

Exhibit 2
Mean standard scores computed for each polygraph record across examiners under the two experimental conditions and the control condition of Experiment 1

Experimental condition	Guilt expectation	Innocence expectations	Control
Record			
1	-0.51	-0.00	+0.25
2	-1.04	-0.62	-0.80
3	-0.93	-0.15	-0.73
4	+0.66	+1.25	+0.85
5	-0.17	+1.19	+1.39
6	-0-86	+0.24	-0.83
7	-0.37	+0.02	-1.30
8	+0.14	+0.91	+0.52
9	-0.43	-0.06	+0.36
10	+0.73	+1.08	+1.52
11	-1.37	+0.72	-0.95
12	-0.46	+0.52	-0.74
13	+0.49	+0.28	+0.99
14	-1.03	-0.22	+0.54
Marginal mean	-0.368	+0.369	+0.076

for the experienced and for the less experienced examiners, are presented in Exhibit 3. The Z scores obtained under the two experimental conditions were compared separately for the experienced and inexperienced examiners using matched-group t-tests. The differences were statistically significant for both the experienced examiners ($t_{(5)} = 3.05$), and for the inexperienced examiners ($t_{(3)} = 3.90$).

A closer inspection of Exhibit 3 reveals that Examiner 7, a less experienced examiner, was highly influenced by outcome knowledge (there was no overlap between the scores assigned by this examiner under the two experimental conditions), whereas Examiner 4, a more experienced examiner, was not affected by the expectations at all. It is possible that examiners differ in their vulnerability to this 'hindsight bias' (e.g. Fischhoff, 1982). However, excluding these two extreme cases, variations among examiners seem to be negligible, and the data suggest that experience cannot account for the expectation effect observed in this experiment.

Discussion

The results of this experiment revealed that outcome knowledge affected the examiners in their scoring of CQT records. Although the effect was reliably demonstrated only with the continuous numerical scale, and not with the classification into broad categories, it seems to be a rather strong effect, as it reached accepted levels of significance with a relatively small sample size. This means that polygraph examiners are affected by prior beliefs when they analyze and interpret polygraph charts. Moreover, prior expectations affected both experienced and inexperienced examiners. This is consistent with Wood (1978), who found that the most knowledgeable subjects were no less prone to hindsight bias than less knowledgeable ones in a task involving general-knowledge questions. Fischhoff (1982) concluded that people normally approach hindsight tasks in a manner that does not use their knowledge or inferential skills at an optimal level.

In the first experiment, only inconclusive records were used, and this

Exhibit 3
Mean standard scores computed for experienced and less experienced polygraph examiners under the two experimental conditions of Experiment 1

Experimental condition	Guilt expectation	Innocence expectations
Experienced examiners		
1	-0.44	+0.44
2	-0.44	+0.43
3	-0.12	+0.13
4	+0.09	-0.09
5	-0.44	+0.44
6	-0.27	+0.27
Less experienced examiners		
7	-0.92	+0.92
8	-0.39	+0.39
9	-0.44	+0.44
10	-0.32	+0.32
Marginal mean	-0.369	+0.369

raises a question regarding the generalizability of the present results to situations where the actual physiological responses clearly indicate deception or innocence. In other words, it is possible that prior information and prior expectations affect polygraph examiners' judgment only when the specific information provided by the polygraph charts is unclear or ambiguous. It is thus important to determine whether outcome knowledge affects examiners' judgment when it stands in complete contrast to the physiological information. It was therefore decided to replicate the first experiment using conclusive polygraph records.

Experiment 2

The second experiment was similar to the first in every respect, with one exception: it utilized conclusive rather than inconclusive polygraph records.

Method

The same ten examiners were asked for a second opinion about several records used in the first experiment for reliability purposes. The examiners were further told that several new records would also be presented. In fact, all records were new records. Six of them were selected because they clearly indicated the innocence of the subjects (a mean global score of +6 or more was assigned by two independent and experienced scorers who were not scheduled to participate in the experiment). The other six records clearly indicated guilt (a mean global score of -6 or less was assigned to them by these two experienced examiners). The interscorer correlation coefficient computed for the global score was 0.88.

After completion of this task and another task (not reported here), the polygraph examiners were debriefed and the real purpose of the study was disclosed. Their reaction was of surprise, but no one expressed anger, disapproval, or other negative feelings. The examiners acknowledged the importance of the study for achieving a better understanding of their profession, and all of them understood that it would have been impossible to conduct the study without deception[3].

[3] There are no formal ethical guidelines for conducting research in the Israel Police. However, as research psychologists we are committed to the ethical guidelines of APA, according to which deception should be avoided unless it is scientifically essential to the study, and it is estimated that the subjects will not be distressed by the deception when debriefed. In this case deception was absolutely necessary and the examiners expressed surprise but not anger or distress when debriefed.

Exhibit 4
Frequencies of classifications made by ten polygraph examiners under two experimental conditions and the control condition of Experiment 2

Experimental condition	Innocence expectation		Guilt expectation		Control	
Actual Outcome	NDI	DI	NDI	DI	NDI	DI
Classification						
NDI	28	0	29	0	12	0
Inconclusive	2	9	1	7	0	2
DI	0	21	0	23	0	10
	30	30	30	30	12	12

NDI = no deception indicated; DI=deception indicated.

Results

As in Experiment 1, all polygraph records were classified into the three categories of DI, NDI, and Inconclusive, based on the global scores assigned to them by the ten examiners. Exhibit 4 presents the frequencies of the three categories made under the two experimental conditions, as well as under the control condition of no outcome knowledge.

Exhibit 4 indicates that the classifications based on the global scores assigned by the ten examiners reflect the actual outcomes of the records and are unrelated to the experimental condition.

The global scores were standardized within examiners as in Experiment 1, and the average Z score across examiners within each experimental condition was computed for each polygraph record. Exhibit 5 displays the mean Z scores for each polygraph record under each experimental condition. For seven out of the 12 records the difference between the average Z scores was in the expected direction (i.e. larger mean Z scores under the Innocence-expectation condition than under the Guilt expectation condition). The mean difference across all records was in the expected direction, but a matched-group t-test revealed that it was not statistically significant ($t_{(11)}$ = 1.42).

A similar analysis was conducted for the examiners. A mean Z score was computed for each of the ten examiners, under each experimental condition, across the six polygraph records scored by the examiner. These means are presented in Exhibit 6. The differences between the Z scores obtained under the two experimental conditions were in the expected direction for eight of the ten examiners. However, these differences were small and a matched-group t-test indicated that they were not statistically significant ($t_{(9)}$ = 1.25).

Discussion

Unlike the previous experiment, the results of Experiment 2 failed to produce a statistically significant effect for the examiners' expectations, although the differences were in the predicted direction. It seems that when prior expectations clearly contradict the specific, physiological information provided by the polygraph charts, examiners tend to rely primarily upon the latter and pay little attention to their expectations. Thus, the results of Experiment 2 suggest that the effect of prior expectations on polygraph charts' reading is limited to inconclusive records. It may be argued along the hypothesis suggested by Ben-Shakhar (1991) that when analyzing polygraph charts, the examiner does search for confirmatory information, but when no confirmatory signs exist, the hypothesis generated from the prior expectations is

Exhibit 5
Mean standard scores computed for each polygraph record across examiners under the two experimental conditions and the control condition of Experiment 2

Experimental condition	Guilt expectation	Innocence expectation	Control
Record			
1	+1.16	+1.30	+1.09
2	+0.62	+0.60	+0.68
3	+1.16	+1.03	+1.09
4	-0.83	-0.35	-0.66
5	-1.13	-1.05	-1.12
6	-1.12	-0.90	-1.12
7	+0.83	+1.31	+1.25
8	+0.73	+0.54	+0.68
9	+1.17	+0.83	+0.99
10	-0.98	-0.34	-1.12
11	-0.99	-1.19	-0.65
12	-1.38	-1.01	-1.01
Marginal mean	-0.063	+0.064	+0.008

Exhibit 6
Mean standard scores computed for experienced and less experienced polygraph examiners under the two experimental conditions of Experiment 2

Experimental condition	Guilt expectation	Innocence expectations
Experienced examiners		
1	-0.15	+0.15
2	-0.02	+0.02
3	+0.13	-0.13
4	-0.07	+0.07
5	-0.05	+0.05
6	-0.08	+0.08
Less experienced examiners		
7	-0.14	+0.14
8	-0.08	+0.08
9	-0.02	+0.02
10	+0.11	-0.11
Marginal mean	-0.037	+0.037

rejected, and the final judgment reflects the physiological information. It should be noted that polygraph examiners at the Israel Police are trained to score polygraph records using the semi-objective technique suggested by Backster (1963). It is possible that examiners trained by more traditional schools, which emphasize subjective methods of global evaluation of polygraph records, would be more vulnerable to biases of prior beliefs and expectations, and would be affected by expectations even when scoring polygraph records that include contradictory signs.

It cannot be completely ruled out that the experimental manipulation was not sufficiently strong, in which case the relatively large discrepancies between the charts and the prior expectations might have raised some suspicion regarding the credibility of the alleged confessions.

General Discussion

The present study was conducted to determine the impact of prior expectations on the judgments of polygraph examiners when analyzing polygraph charts obtained from a CQT polygraph examination. The influence of prior expectations on human judgment has been demonstrated mainly in artificial experimental conditions (e.g. Fischhoff, 1982; Snyder and Swann, 1978a,b), but there are some demonstrations that it can affect professionals working in more realistic and familiar situations (e.g. Arkes et al., 198 1; Detmer et al., 1978). Ben-Shakhar (1991) argued that judgments of CQT polygraph examiners may be particularly vulnerable to a bias created by prior expectations and beliefs, because the CQT is based to a large extent on the examiner-examinee interactions, rather than on objective inference rules and quantification schemes. He further argued that the interpretation of the complex physiological information is infected by a confirmation bias (i.e. larger weight is given to physiological data consistent with the prior expectations, while contradictory evidence is largely ignored). Furthermore, it was argued that prior expectations can affect not only the evaluation of polygraph charts but also the choice of the control questions and the manner in which they are administered to the examinees.

The results of the present experiments supply only partial support to the hypotheses raised by Ben-Shakhar (1991). First, prior expectations had an impact on the examiners' judgments when they analyzed previously obtained polygraph charts. The impact of prior expectations on polygraph examiners conducting the entire CQT test and having access not only to the physiological responses but also to the subjects' behavior during the various phases of the CQT test will have to be examined in future studies. To use Darley and Gross (1983) terminology, only the 'cognition confirmation effect' (i.e. expectancy-confirmation effects that occur in the absence of any interaction between the perceiver and the target person) was demonstrated in this study, but not the 'behavioral confirmation effect' (i.e. where expectations affect the behavior toward a target individual, such that expectancy-confirming behaviors are elicited from this individual).

Second, the impact of prior expectations on the interpretation of polygraph charts was observed only when these charts did not include clear indications of either guilt or innocence. In other words, it seems that when the specific physiological information clearly contradicts prior expectations, examiners tend to ignore their expectations and make their judgments on the basis of the physiological information. This finding is consistent with the interpretation of hindsight effects in terms of availability (i.e. failure to consider alternative outcomes). Slovic and Fischhoff (1977) suggested that victims of hindsight bias focus their attention on the reported outcome and fail to consider alternatives. Indeed, they demonstrated that once subjects were encouraged to search for possible explanations for the alternative outcome (the outcome that did not happen), the hindsight effect was reduced. A similar account was provided by Koriat et al. (1980) to explain overconfidence in human judgment. They suggested that overconfidence occurs because subjects are attempting to justify their chosen answer, and in this process they focus on evidence which supports this answer, and disregard contradictory evidence. Koriat et al. (1980) demonstrated that inducing subjects to list contradicting reasons reduced the overconfidence effect. The results of Experiment 2 suggest that providing examiners with physiological information which clearly contradicts their prior expectations drastically reduced the bias. These clear records are bound to bring the alternative possibility to the attention of the polygraph examiners, and thus reduce the effect of outcome knowledge, and possibly eliminate it.

The demonstration of a prior-expectations effect on professional polygraph examiners may have some practical implications. Even if such an effect operates only for some polygraph examiners and only

when the physiological information is not conclusive, it might increase the error rate of classifications made by polygraph examiners. Several measures can be suggested in order to minimize and possibly eliminate the effects of prior expectations in CQT polygraph examinations. First, a procedure of blind scoring of the charts (i.e. scoring by an examiner who is unaware of the details of the case under investigation, and is unfamiliar with the suspect) should be routinely implemented. Such a procedure has been adopted already by the Israel Police, and when there are disagreements between the original examiner and the blind scorer, the case is further discussed. Ideally, charts should be scored mechanically by a computer, using various techniques for measuring physiological reactions that are available in experimental psychophysiology (e.g. Kircher and Raskin, 1988). In principle, even a complete computerization of physiological measurement may be insufficient to eliminate all possible sources of bias, because prior beliefs and expectations may affect not only chart interpretation but also the whole manner in which the CQT is administered. Ben-Shakhar et al. (1986) listed several steps that should be taken to achieve a complete decontamination of polygraph examinations, although they doubted whether such changes would be acceptable to CQT examiners who emphasize the examiner-examinee interaction. Additional measures that can reduce the impact of prior expectations are the use of a larger range of inconclusive classifications, and a greater emphasis on the dangers of relying upon prior information in the training of polygraph examiners.

Finally, it should be noted that effects of prior expectations on human judgment may extend to other situations where experts are required to make judgments and predictions on the basis of various tools. Ben-Shakhar (1991) argued that whenever expert judgments are based on subjective tools with no well-defined quantification schemes and inference rules, and when the expert is exposed to a great deal of prior information, the judgments allegedly made on the basis of a scientific instrument might be seriously distorted by other sources of information and might reflect to a large extent the prior beliefs of the judge. For example, many of the tools commonly used for personnel selection and diagnosis (e.g., projective techniques, personal interviews) are based on subjective judgment, rather than on objective and quantified inference rules. Ben-Shakhar (1991) suggested that such tools provide the users with an impression of validity (personal validity) created through the process of confirming initial hypotheses (a test which so often confirms the user's prior expectations acquires a sense of validity). Another example is the preference for intuitive (clinical) judgment over the use of statistical models for prediction. The advantage of statistical over clinical prediction has been documented repeatedly (e.g. Dawes, 1979), yet personnel decisions are typically made at least partly on the basis of intuitive judgment. Future research should focus on these issues and determine which type of tools and testing situations are vulnerable to both the cognitive and the behavioral confirmation effects.

References

Arkes, H. R., Wortman, R. C., Saville, P. D. and Harkness, A. R. 'Hindsight bias among physicians weighing the likelihood of diagnosis', *Journal of Applied Psychology*, 66 (1981), 252-254.

Backster, C. 'Polygraph professionalization through technique standardization', *Law and Order*, 11 (1963), 63-64.

Ben-Shakhar, G. 'Standardization within individuals: A simple method to neutralize individual differences in psychophysiological responsivity', *Psychophysiology*, 22 (1985), 292-299.

Ben-Shakhar, G. 'Clinical judgment and decision making in CQT polygraphy: A comparison with other pseudoscientific applications in psychology', *Integrative Physiological and Behavioral Science*, 26 (1991), 232 240.

Ben-Shakhar, G., Bar-Hillel, M. and Lieblich, I. 'Trial by polygraph: Scientific and juridical issues in lie detection', *Behavioral Science and the Law*, 4 (1986), 459-479.

Ben-Shakhar, G. and Furedy, J. J. *Theories and Applications in the Detection of Deception: A psychophysiological and international perspective*, New York: Springer-Veriag, 1990.

Christensen-Szalanski, J. J. J. and Bushyhead, J. B. 'Physicians' use of probabilistic information in a real clinical setting', *Journal of Experimental Psychology. Human Perception and Performance*, 7 (1981), 928-935.

Darley, J. M. and Gross, P. H. 'A hypothesis-confirmation bias in labeling effects', *Journal of Personality, and Social Psychology*, 44 (1983), 20-33.

Dawes, R. M. 'The robust beauty of improper linear models in decision making', *American Psychologist*, 34 (1979), 571-582.

Detmer, D. E., Fryback, D. G. and Gassner, K. 'Heuristics and biases in medical decision-making', *Journal of Medical Education*, 53 (1978), 682-683.

Evans, J. St B. T. Bias in Human Reasoning: Causes and consequences, London: Lawrence Erlbaum, 1989. Fischhoff, B. 'Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty', *Journal of Experimental Psychology: Human Perception and Performance*, 1 (1975), 288-299

Fischhoff, B. 'Debiasing', in Kahneman, D., Slavic, P. and Tversky, A. (eds), Judgment under Uncertainty: Heuristics and biases, pp. 422, Cambridge: Cambridge University Press, 1982.

Furedy, J. J. and Heslegrave, R. J. 'The forensic use of the polygraph: A psychophysiological analysis of current trends and future prospects', in Jennings, J. R., Ackles, P. K. and Coles, M. G. H. (eds), *Advances in Psychophysiology*, 4, Greenwich, CT: JAI Press, 1989.

Kircher, J. C. and Raskin, D. C. 'Human versus computerized evaluations of polygraph data in laboratory setting', *Journal of Applied Psychology*, 73 (1988), 291-302.

Klayman, J. and Ha, Y.-W. 'Confirmation, disconfirmation, and information in hypothesis testing', *Psychological Review*, 94 (1987), 211-228.

Koriat, A., Lichtenstein, S. and Fischhoff, B. 'Reasons for confidence', *Journal of Experimental Psychology: Human Learning and Memory*, 6 (1980), 107-118.

Lykken, D. T. 'Psychology and the lie detection industry', *American Psychologist*, 29 (1974), 725-739.

Lykken, D. T. 'Uses and abuses of the polygraph', in Pick, H. L. (ed.), *Psychology: From Research to Practice*, New York: Plenum Press, 1978.

Lykken, D. T. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, New York: McGraw-Hill, 1981.

Raskin, D. C. 'The scientific basis of polygraph techniques and their uses in the Judicial process', in Trankell, A. (ed.), *Reconstructing the Past: The role of psychologists in the criminal trial*, Stockholm: Norsted & Soners, 1982.

Raskin, D. C. 'Polygraph techniques for the detection of deception', in Raskin, D. C. (ed.), *Psychological Methods in Criminal Investigation and Evidence*, New York: Springer-Veriag, 1989.

Reid, J. E. and Inbau, F. E. *Truth and Deception: The Polygraph ('Lie Detection) Technique*, Baltimore, MD: Williams and Wilkins, 1977.

Slovic, P. and Fischhoff, B. 'On the psychology of experimental surprise', *Journal of Experimental Psychology: Human Perception and Performance*, 3 (1977), 544-551.

Smith, J. F. and Kida, T. 'Heuristics and biases: Expertise and task realism in auditing', *Psychological Bulletin*, 109 (1991),472-489.

Snyder, M. and Swann, W. B., Jr. 'Behavioral confirmation in social interaction: From social perception to social reality', *Journal of Experimental Psychology*, 14 (1978a), 148-162.

Snyder, M. and Swann, W. B., Jr. 'Hypothesis-testing processes in social interaction', *Journal of Personality and Social Psychology*, 36 (1978b), 1202-1212.

Stenson, H., Kleinmuntz, B. and Scott, B. 'Personality assessment as a signal detection task', *Journal of Consulting and Clinical Psychology*, 43 (1975), 794-799.

Wason, P. C. 'On the failure to eliminate hypotheses in a conceptual task', *Quarterly Journal of Experimental Psychology*, 12 (1960), 129-140.

Wason, P. C. 'On the failure to eliminate hypotheses - a second look', in Wason, P. C. and Johnson-Laird, P. N. (eds), *Thinking and Reasoning*, Penguin, Harmondworth, 1968.

Wood, G. 'The knew-it-all-along effect', *Journal of Experimental Psychology,: Human Perception and Performance*, 4 (1978), 345-353.

Authors' Biographies

Eitan Elaad is the research officer of the Scientific Interrogation Unit. Israel Police. He is also a lecturer at the Department of Criminology, Bar-Ilan University. He received his Ph.D. in Psychology from the Hebrew University Jerusalem in 1988. Research interests include psychophysiological detection of concealed knowledge, the social psychology of deception, and decision-making processes.

Avital Ginton is a commander in the National Police of Israel and Head of Scientific Interrogation and Behavior Section. He received a Ph.D. in Psychology from Tel-Aviv University in 1978, and an MPA from Harvard University in 1993. For many years he lectured in the Department of Psychology at Tel-Aviv and Criminology at Bar-Ilan Universities.

Gershon Ben-Shakhar (Ph.D., Hebrew University of Jerusalem) is a Professor of Psychology at the Hebrew University of Jerusalem. His main research is focused on human psychophysiology (orientation and habituation processes), and psychophysiological detection of information. He also studied psychological testing and its applications to personnel decisions.

Authors' Addresses

Eitan Elaad and Avital Ginton, Behavior Section, Division of Identification and Forensic Science, Israel National Police Headquarters, Jerusalem, 91906 Israel.

Gershon Ben-Shakhar, Department of Psychology, The Hebrew University of Jerusalem, Israel.

Louder and Longer: A Review of the second edition of Lykken's *A Tremor in the Blood : Uses and Abuses of the Lie Detector*

By

Charles R. Honts

A Tremor in the Blood: Uses and Abuses of the Lie Detector, by David T. Lykken, New York, Plenum Trade, 1998, 333 pages. Indexed. $28.95

In 21 chapters and a prolog divided into four parts Lykken addresses the various polygraph techniques, integrity testing, confessions, and the use of expert testimony in American courts of law. Lykken also includes many anecdotes about his involvement in providing beat the polygraph information to imprisoned felons and people trying to get through the national security screening system. Unfortunately, while the first edition provided a relatively clear statement of Lykken's arguments against the polygraph, the second edition has little to recommend it. In his review of the first edition Barland (1982) opened his evaluative review with the following sentence, "Lykken's concept of the polygraph technique is quite distorted." (258). To apply that statement to the second edition would be to would be to grossly understate the case. The second edition of A *Tremor in the Blood* represents Lykken's is nothing more or less than a partisan polemic. Unfortunately, this volume is likely to mislead the hapless reader who relies on Lykken's considerable bona fides, and who comes to A *Tremor in the Blood* expecting science rather than diatribe.

There is an old saying that if you tell a lie loud enough and keep repeating it long enough, then people will come to believe that it is true. After reading the second edition of *Tremor in the Blood* (*TITB*) it would be easy to form the opinion that David Lykken has adopted this old saw as the core to his approach to the polygraph. The disinformation and misinformation in *TITB* is so extensive that it would take treatise well beyond the limits of a book review to begin to address all of the incidences. The errors in *TITB* range from relatively minor (e. g., Dr. John Kircher and Dr. Steven Horowitz are described as practicing polygraph examiners [p. 51]. Neither has ever been a practicing polygraph examiner. Both are tenured academics at accredited institutions of higher learning who have published research on the polygraph, to more serious violations of omission, and on some occasions what appear to be deliberate distortions. A few examples may help to illustrate the quality of scholarship in *TITB.*

One clear example of Lykken's propensity to argue with significant omissions of relevant data comes in his discussion of psychopaths and the polygraph (pp 267-271). In this section Lykken describes only one published study (Raskin & Hare, 1978), and he is critical of the methods used in that study. Lykken leaves the impression that the results of the Raskin and Hare study cannot be trusted. However, what Lykken fails to mention in his book is that his former student and now colleague, William Iacono replicated (Patrick & Iacono, 1989) the findings of Raskin and Hare in a study that directly addressed Lykken's criticisms of Raskin and Hare. Patrick and Iacono (1989) conclude, "Guilty psychopaths were detected just as easily as guilty non-psychopaths and the majority of guilty subjects (87%, excluding inconclusives) were correctly identified. (347). Lykken also fails to mention the study reported by Honts, Raskin, and Kircher (1985) in the *Journal of Research in Personality,* who reached the same conclusion. The reasons why Lykken failed to mention a study published by his former student, of whom Lykken speaks of glowingly in *TITB,* is open to speculation, but there don't seem to be too many options from which to choose.

Lykken's predilection for misinformation is in at least one case much worse

than misleading by omission. In describing a study that I recently published (Honts, 1996) Lykken moves well beyond simple omission. Honts (1996) was a field validity study conducted with the Royal Canadian Mounted Police and the Canadian Police College. In describing that study Lykken makes the following statement, "A polygraphy instructor at the college sent Honts charts from tests administered to seven suspects who had confessed after failing the CQT and also charts of six suspects confirmed to be innocent by these confessions of alternative suspects in the same crimes." (134-135). Nothing in the Honts (1996) publication supports this assertion, and it is in fact false. Lykken is describing a subsample of the study as if those 13 subjects were the only ones considered. It is true that there were 7 guilty subjects confirmed by their own confessions, but none of the innocent subjects were confirmed by confessions that followed polygraph examinations. Then Lykken states, "Using the same scoring rules (and also knowing which suspects were in fact guilty), Honts of course managed to scores all seven as deceptive also." (135). This is also never stated in the Honts (1996) manuscript and Lykken's statement is in fact false. The Honts (1996) manuscript is states "The original examiners and an independent evaluator preformed numerical evaluations according to the techniques taught at the Canadian Police College (CPC). A second set of independent numerical evaluations was made on the data collected in the second wave." (315-316). The key word here is "independent". The word independent in its scientific use in this setting is understood to mean blind or naïve. The Lykken statement quoted above essentially calls me a liar and a fraud, because the information he includes in *TITB* is not what was published in the original source. Before the charts in Honts (1996) were scored all identifying information was removed from them by an assistant and they were coded with random identification numbers. When I scored them, I had no idea which set of charts went with which case. The same was true for S/Sgt. John Kaster of the CPC who did the second independent scoring. Lykken's assertions to the contrary are pure fantasy.

Finally, in describing the Honts (1996) study, Lykken makes the following statement, "How it came about that scientific peer review managed to allow this report to be published in an archival scientific journal is a mystery. Since the author, Honts, and the editor of the journal, Garvin Chastain, are colleagues in the psychology department of Boise State University, it is a mystery they might be able to solve." (135). Lykken's ad hominem attack on the peer reviewers and the editor at *The Journal of General Psychology* is scurrilous and disgusting. Anyone, who knows Professor Chastain, will know that any suggestion that he engaged in under the table dealing with regard to a publication is laughable. The process of publishing Honts (1996) was as follows. The manuscript was submitted to *The Journal of General Psychology*'s office in Washington D.C. The office sent the article out for peer review. Two peer reviews were obtained, one of which was signed by a very prominent psychophysiologist, who is not part of the polygraph community. Both reviews were positive but requested some revisions. The peer reviews and the manuscript were only then forwarded to Professor Chastain, who also evaluated the manuscript and then asked for a revision. The revision was provided as requested and the article was accepted and published. Nothing out of the ordinary was done.

Many, many other examples of the poor quality of scholarship in *TITB* could be provided, but space prohibits. However, given the blatant examples cited above, one has to seriously question the entire content of *TITB*, because, at best, the source is highly biased. Readers interested in reading both sides of the polygraph controversy should seek out the volume by Faigman, Kaye, Saks and Sanders (1997) and its 1998 supplement. I cannot recommend that anyone buy *TITB*, but if you go looking for it, I can only hope that most book sellers place it section where it belongs; with the bad science fiction.

References

Barland, G. H., (1982). On the accuracy of the polygraph: An evaluative review of Lykken's *Tremor in the Blood. Polygraph*, 11, 258-272.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J., (Eds., 1997). *Modern scientific evidence: The law and science of expert testimony*, Vol. 1. St. Paul, MN: West.

Honts, C. R. (1996). Criterion development and validity of the control question test in field application. *The Journal of General Psychology*, 123, 309-324.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1985). Effects of socialization on the physiological detection of deception. *Journal of Research in Personality*, 19, 373-385.

Patrick, C. J., & Iacono, W. G., (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology*, 74, 347-355.

Raskin, D. C., & Hare R. D., (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126-136.

Nonverbal Communication
An Investigator's Guide

by

Albert D. Snyder

Book Review

by

Helen A. Montgomery

Throughout their careers law enforcement and other investigative personnel learn a variety of techniques for obtaining cooperation and information from victims, witnesses, and suspects. Many of these techniques take considerable experience to learn. In this guide the author elaborates on the significance of nonverbal cues during an interview. This book discusses eye contact and facial expressions, gestures, posture, distance, and touching.

The text begins with a summary of the significance of nonverbal communication. Snyder cautions that although the nonverbal cues discussed in this book may be indicators of deception, no single nonverbal cue is indicative of a deceptive person. Snyder emphasizes that rather than link one particular nonverbal cue to a particular meaning or significance, investigators should analyze gestures and behaviors in clusters.

The author divides interpretation of nonverbal cues into four major areas: the head, the mid section, the lower section, and general body movements. Because the majority of nonverbal communication occurs as a result of facial contortions and head movements, more than 40 percent of the text is dedicated to nonverbal cues associated with the head and facial area of the body.

Early in the book Snyder provides a guide for setting up the ideal interview room. He then divides the book into seven categories for review. Each category is replete with examples and illustrations of nonverbal cues and their commonly accepted meanings. Snyder is careful to note that investigators should be aware that several factors can affect the interpretation of nonverbal cues including cultural differences. After each section or category for review, Snyder provides an opportunity to practice the skills learned or reviewed.

In many cases the spoken word is the greatest source of investigative evidence. Proficient interviewing assures a high degree of reliability in the final results of an investigation. In addition, proficient interviewing prevents surprises from arising later in the investigation or adjudication processes. Proficient interviewing is a skill that must be learned through specialized training and experience.

In general, this book provides a good introduction and review of the science of interviewing. Novice investigators will find this book a useful and practical guide for interpreting and understanding the significance of nonverbal behavior of interviewees. Experienced investigators will find this guide an excellent refresher and a handy reference. Both new and experienced investigators will also find an added bonus in the extensive bibliography prepared by Norman Ansley and Brenda Knill.

Instruction to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript submission

Manuscripts should be in English, and three copies submitted to Editor, American Polygraph Association, PO Box 4085, Anniston, Alabama 36206 (USA). Authors may also submit manuscripts on computer disk (3½" IBM PC compatible) in text format, or via e-mail to dkrapohl@aol.com. The cover letter should include a telephone number, return address, and e-mail address if possible. As a condition for publication, authors shall be required to sign a statement that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere.

Manuscript organization and style

Manuscripts must be double-spaced and on 8½ X 11 inch paper. The first page is a title page, which must list the author(s), date of submission, address to request reprints, and a running head. All manuscripts must be complete, balanced, and accurate. All authors are encouraged to follow guidelines in the *Publication Manual of the American Psychological Association* (4th edition). The manual can be found in most public and university libraries, and can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Writers may exercise some freedom of style, but they will be held to a standard of clarity, organization, and accuracy. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psychophysiological Detection of Deception* (1997) which is available from the national office of the American Polygraph Association. Legal case citations should follow the *West* system.

Figures

Each figure, table, graph, and illustration should bear on the back a number, the author's name, and the title of the paper. An arrow on the back to indicate which side is up is also required. *Polygraph* does not use color in its publication, and authors should verify that it is not needed to interpret these materials. Care should be taken to ensure that lettering and detail are legible after the size reduction for the publication.

Manuscript review

Papers will be handled by a single Associate Editor, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 12 weeks of receipt of manuscript.

Copy-editing

The Editor reserves the right to copy-edit manuscripts. All changes will be coordinated with the principal author.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be approved, if the senior author concurs.

Professional copies

The senior author will receive ten (10) copies of the journal issue in which the article appears.