Journal of Applied Psychology
1994, Vol. 79, No. 2, 252-259

0021-9010/94/$3.00

Mental and Physical Countermeasures Reduce the Accuracy of Polygraph Tests

Charles R. Honts, David C. Raskin, and John C. Kircher

Effects of countermeasures on the control-question polygraph test were examined in an experiment with 120 Ss recruited from the general community. Ss were given polygraph tests by an examiner who used field techniques. Twenty Ss were innocent, and of the 100 guilty Ss, 80 were trained in the use of either a physical countermeasure (biting the tongue or pressing the toes to the floor) or a mental countermeasure (counting backward by 7) to be applied while control questions were being presented during their examinations. The mental and physical countermeasures were equally effective: Each enabled approximately 50% of the Ss to defeat the polygraph test. The strongest countermeasure effects were observed in the cardiovascular measures. Moreover, the countermeasures were difficult to detect either instrumentally or through observation.

Despite a long-standing controversy regarding their validity (e.g., Lykken, 1981; Raskin, 1989), physiological detection of deception (polygraph) tests are widely used in law enforcement (Raskin, 1989), are sometimes used in the courts (Honts & Perry, 1992; Morris, 1989), and play a major and increasing role in the national security systems of a number of countries (Barland, 1988; Honts, 1991). The use of polygraph tests for national security screening in the U.S. Department of Defense rose 586% between 1986 and 1990 (U.S. Department of Defense [DOD], 1987, 1991). Vital decisions concerning who should have access to cryptology, government secrets, and nuclear command and control have largely been based on polygraph test results. Although evidence has indicated that polygraph tests may be ineffective in national security screening systems (Honts, 1991, 1992), current proposals would greatly expand the use of polygraph screening to include all people with top secret security clearances and people involved in the war on drugs (DOD, 1991).

Control question tests (CQTs) are the most commonly used polygraph tests in law enforcement (Raskin, 1986) and they are widely applied within the national security system of the United States (Honts, 1991, 1992). The rationale of the CQT predicts that guilty and innocent subjects will react differentially to relevant and control questions. It is expected that guilty subjects will produce their strongest physiological responses when lying to relevant questions because such questions deal directly with the issues of the examination (e.g., "Did you stab John Doe?" or "Have you given classified material to unauthorized people?"). However, innocent subjects are expected to show their strongest physiological responses to control questions (e.g., "Before 1990, did you ever do anything that was dishonest, illegal, or immoral?"). The control questions are explained and presented to the subject in such a manner that a denial is obtained. It is assumed that both guilty and innocent subjects will be concerned about the veracity of their denials to the control questions. Innocent individuals are expected to produce larger physiological responses to control questions than to relevant questions because they are certain of the veracity of their responses to the relevant questions and, it is assumed, they are either lying or at least uncertain about the veracity of their responses to the control questions.

The possibility that countermeasures might be used to defeat or distort the CQT has raised concerns about its usefulness (Lykken, 1981). However, the spontaneous use of countermeasures by untrained subjects has been found to be ineffective against the CQT (Honts, Raskin, Kircher, & Hodes, 1988). Furthermore, providing detailed information about the nature of the CQT and possible countermeasures does not seem to affect accuracy rates (Rovner, 1986). However, other research has shown that training in simple physical maneuvers, such as biting the tongue or pressing the toes to the floor, can be effective in defeating polygraph tests by enhancing physiological reactions to control questions. Honts, Hodes, and Raskin (1985) reported that 60% of their decisions were incorrect when subjects had been trained to unobtrusively bite their tongues and press their toes to the floor when control questions were presented during the test. Using similar training and stronger incentives to pass the test, Honts, Raskin, and Kircher (1987) failed to correctly classify any subjects who were using countermeasures.

Countermeasure detectors and counter-countermeasures have been suggested as possible solutions to the threat posed by physical countermeasures (Honts, 1987). It appears that the U.S. government is expending considerable resources on the development of countermeasure detectors (DOD, 1988, 1991), but the effectiveness of these efforts is unknown. Regardless of

Charles R. Honts, David C. Raskin, and John C. Kircher, Department of Psychology, University of Utah.

The research reported here served as partial fulfillment of the requirements for Charles R. Honts's PhD in Psychology at the University of Utah.

We acknowledge and thank Sharon Akimoto and Michael Redford for their assistance in training the subjects in this experiment and Mary Devitt for her assistance in the preparation of this article.

Correspondence concerning this article should be addressed to Charles R. Honts, who is now at the Department of Psychology, University of North Dakota, Box 8380, Grand Forks, North Dakota 58202-8380. Electronic mail may be sent to ud195188@VM1. nodak. edu.

the efficacy of detectors that measure physical countermeasures, effective mental countermeasures would pose a serious threat because they are virtually undetectable instrumentally (Honts, 1987).

Only one published study has addressed the effectiveness of mental countermeasures against the CQT. Dawson (1980) explored the possibility that method actors could defeat the CQT because of their extensive training in controlling their emotions. Although none of Dawson's method actors defeated the test, Dawson did not inform his subjects that defeating the test required them to produce stronger physiological reactions to the control questions than to the relevant questions. Therefore, it is possible that Dawson's subjects applied their countermeasures inefficiently. The present study was designed to assess the effectiveness of a mental countermeasure used by subjects who received information and training similar to those found to be effective in previous research on physical countermeasures.

Method

Subjects

Subjects were recruited from the local community through a temporary help-wanted ad in the classified section of both daily newspapers in Salt Lake City, Utah. When potential subjects telephoned in response to the ad, they were screened regarding two issues. Those who reported being under the care of a physician for cardiovascular disease or for mental or emotional conditions were not accepted as subjects. Additionally, respondents who reported having taken a polygraph test or having participated in one of our previous studies were not accepted as subjects. Sixty-eight men and 70 women were accepted as subjects, but 18 subjects were later eliminated from the experiment for the following reasons: 3 confessed their guilty status to the polygraph examiner before the examination, 1 failed to follow the countermeasure instructions, 3 declined to participate after hearing the instructions, and 11 failed to complete both sessions of the experiment. Sixty male subjects and 60 female subjects completed the experiment. Their ages ranged from 18 years to 66 years (M = 30.5 years), and their education ranged from 8 years to 20 years (M = 14.0 years). Subjects were paid $10 for their participation, and each subject was promised a $15 bonus if he or she passed the polygraph test.

Apparatus

The physiological data were recorded on a Beckman Type R Dynograph at a chart speed of 2.5 mm per s. Respiration was monitored from Hg strain gauges placed around the chest and abdomen. Skin conductance was recorded without frequency filtering from two Beckman 10-mm Ag/AgCl electrodes filled with .05 molar NaCl electrolyte in a Unibase medium (Fowles & Schneider, 1978). The electrodes were attached to the palmar surface of the middle phalanx of the second and third fingers of the left hand. A finger-pulse measure of vasomotor activity was recorded with a .10-s time constant and an upper frequency cutoff of 30 Hz using a photoelectric plethysmograph attached by a velcro band to the palmar surface of the distal phalanx of the subject's left index finger. A continuous measure of cardiovascular activity (Geddes & Newberg, 1977) was obtained with a Stoelting electrocardio unit from a blood pressure cuff placed on the upper arm and inflated to a pressure of 70 mm Hg.

An instrumental manipulation check on the physical countermeasure subjects was obtained by electromyography (EMG) recordings from the area of the left temporalis (jaw) and gastrocnemius (calf) muscles. EMG recordings were obtained with standard Ag/AgCl electrodes filled with electrode paste. The EMG signals were processed through two optically isolated preamplifiers (Fridlund, Price, & Fowler, 1982) and two contour-following integrators (Fridlund, 1979). The processed-EMG signals were recorded on magnetic tape and were not available to either the polygraph examiner who conducted the examination or the independent numerical evaluator. The EMG data were later recorded on paper by replaying the tapes through the Beckman polygraph.

We used a Terak 8510/8515 computer with specially developed software to control the timing and sequencing of question presentation, the operation of the instrumentation tape recorders, analog to digital conversion, data reduction, storage, quantification, and feature extraction. Details of this software are provided in Kircher and Raskin's (1988) article. Statistical analyses were performed on an IBM compatible PC on which SPSS/PC+ was run (Norusis, 1988a, 1988b).

Design

Twenty subjects were randomly assigned to each of six treatment conditions without regard to sex. There were two control groups: One was innocent of the mock crime, and the other enacted the mock crime but received no special training or instruction about countermeasures. Subjects in the remaining four groups were trained to perform one or more countermeasures during their polygraph tests, as described below.

Procedure

Following the initial telephone screening, subjects were given an appointment to report individually to a room in the psychology building and were told they would be given additional instructions at that time. On arriving for the appointment, a subject found a note addressed to her or him on the door. The note instructed the subject to enter the room, lock the door, go into a closet in the back of the room, close the door, read and sign a consent form, and listen to a tape recording. All subjects were informed that on a later date they would be given a polygraph test concerning their guilt or innocence of a simulated theft. Innocent control subjects were told that a rare coin had been stolen but that they did not steal it. Guilty control and all countermeasure subjects were told to go to another floor of the building, enter an office, search a desk, find a cash box, open the box, find and steal a rare coin, and then return to the room and wait for an assistant. Subjects were warned that because only the experimenter knew that this was an experiment, if they were caught stealing the rare coin by anyone else, they could be arrested.

After returning to the room where they had received their tape-recorded instructions, countermeasure subjects were met by an assistant and taken to another room to receive their countermeasure training. The four countermeasure conditions used training and methods based on procedures used in previous research (Honts, Hodes, & Raskin, 1985; Honts, Raskin, & Kircher, 1987). Countermeasure subjects were informed about the general nature of CQTs and were told how CQTs were evaluated. They were then given training in a countermeasure and were told that it should enable them to beat a CQT.

Subjects in the muscle tension, pain, and pain and muscle countermeasure conditions were instructed to press their toes to the floor, to bite their tongue, or to simultaneously press their toes and bite their tongue, respectively. Mental arithmetic subjects were instructed to count backward by 7s from a number larger than 200 when the control questions were asked. Each countermeasure subject was instructed to begin the countermeasure as soon as he or she recognized a control question, stop just long enough to answer, and then continue the countermeasure until the next question began. Each countermeasure subject was then read a set of questions from a typical CQT and was coached in using his or her countermeasure unobtrusively so that it would not be detected by the polygraph examiner during the subsequent test. None of the questions used in this practice test were used in the actual polygraph

Table 1
Typical Question Series Used in the Polygraph Tests for a 35-Year-Old Subject

Type of question	Question
Introductory	Do you understand that I will ask only the questions we have discussed? (yes)
Introductory	Regarding the theft of the rare coin, do you intend to answer all of the questions truthfully? (yes)
Neutral	Do you live in the United States? (yes)
Control	During the first 34 years of your life, did you ever take something that did not belong to you? (no)
Relevant	Did you take the rare coin? (no)
Neutral	Is your first name [John]? (yes)
Control	Prior to 1984, did you ever deceive someone? (no)
Relevant	Did you take the rare coin from the desk? (no)
Countermeasure	Are you doing anything in an effort to defeat or distort this test? (no)
Neutral	Were you born in the month of [September]? (yes)
Control	Between the ages of 18 and 34, did you ever do something dishonest, illegal, or immoral? (no)
Relevant	Regarding the rare coin that was reported missing, did you take it? (no)

Note. Expected responses are given in parentheses after each question.

examinations, and subjects were not informed of the order of questions during the examination. The countermeasure training required a maximum of 30 min.

Approximately 1 week after their initial appointments, all subjects were administered a CQT polygraph examination by an experienced polygraph examiner who was unaware of the subjects' guilt, innocence, or countermeasure training. Following standard field practice, the examiner administered a stimulation test before the CQT (Raskin, 1989). The examiner used field techniques to administer a CQT that contained three pairs of relevant and control questions and five filler questions. An additional question was included that addressed the use of countermeasures. All questions were reviewed with the subject before the actual test. If the subject made significant admissions in response to control questions, then those questions were reworded to begin with "other than what you told me about." In accordance with standard field practice, the examiner actively discouraged admissions to control questions. A typical question series used in this study is shown in Table 1. The questions were repeated 5 times with approximately 25 s between onsets of successive questions, and there was a short break and brief discussion between each repetition of the question series (see Raskin, 1989).

The physiological data were evaluated by the original examiner and an independent evaluator who used numerical chart analysis techniques developed at the University of Utah (Kircher & Raskin, 1988). Comparisons were made for each of the physiological systems for each of the relevant and control pairs, and scores were assigned for each comparison. Thus, 60 scores were assigned for each subject. The following 11 physiological changes were evaluated as response criteria: decrease in respiratory cycle amplitude; increase in respiratory cycle time; increase in respiratory baseline; increase in skin conductance amplitude, duration, and complexity; increase in cardiovascular response amplitude and duration; decrease in finger-pulse amplitude, and increase in the duration of finger-pulse response. The use of these criteria involved some objective measurement and some subjective judgment by the examiner and has been described in detail elsewhere (Kircher & Raskin, 1988; Raskin & Hare, 1978).

Each comparison yielded a score on a 7-point scale that ranged from −3 to 3. Positive scores were assigned when the control question evoked the stronger physiological response, and negative scores were assigned when the relevant question evoked the stronger response. A score of 0 was assigned when there was little or no difference between the reactions to the two questions. The absolute value of the scores indicated the magnitude of the difference in reaction to the two questions. The scores were summed across all pairs of relevant and control questions and across all of the response channels. Total scores of 6 or greater yielded a decision of *truthful,* total scores of −6 or less yielded a decision of *deceptive,* and total scores between 6 and −6 were considered *inconclusive* (no decision). This scoring system has been found to be very reliable, with interrater correlations of total numerical scores consistently exceeding correlations of .90 (see Raskin, 1986).

At the time of the numerical scoring, the original examiner also made a forced-choice decision regarding countermeasure use by the subject. This decision could be based either on the examiner's observation of the subject's overt behavior during the examination, on abnormalities in the charts, or on both.

The EMG data were evaluated independently of the standard numerical evaluation. Procedures developed and validated in earlier research (Honts et al., 1987) were used for numerical scoring of the EMG data. Scores were assigned on a 7-point scale ranging from −3 to 3 such that EMG responses of greater strength and duration to the control questions resulted in positive scores.

Computerized measurements of the physiological responses were also obtained. For all relevant and control questions, we extracted the following seven features from the physiological waveforms for analysis, using algorithms described previously (Kircher & Raskin, 1988): skin conductance amplitude and duration, cardiovascular response amplitude and duration, vasomotor response amplitude and duration, and length of the respiration tracing for 10 s following question onset. All of the features were then converted to the same metric with the following range correction transformation:

$$\text{percentage of range} = 100[(x - \text{minimum})/(\text{maximum} - \text{minimum})],$$

where x is an observed measurement and the *minimum* and *maximum* represent the extreme values observed for all relevant and control questions for that measure for a given subject. For each physiological measure, the percentage of range scores were averaged within a question type to produce a mean for relevant questions and a mean for control questions. The mean for relevant questions was then subtracted from the mean for control questions to produce a difference score for each physiological component. Because reactions in the respiratory system result in a decrease in the respiratory length measure, respiration scores were multiplied by −1. Thus, for all features analyzed, positive difference scores indicated that control questions evoked stronger physiological reactions, and negative difference scores indicated that relevant questions evoked stronger physiological reactions. For each subject, a single difference score was obtained for each of the seven physiological measures to represent the differential reactivity to all control questions and relevant questions across the five repetitions of the question sequence.

Manipulation Check

Following the collection of physiological data and before the examiner informed subjects of his decision, subjects were given a questionnaire about their use of countermeasures. In this questionnaire, subjects were asked to describe the countermeasures that they used (if any were used) and when they used them, and they were instructed not to return the questionnaire to the examiner until he had rendered his opinion. After informing subjects of his decision, the examiner inspected their questionnaires and discussed the contents with them. The subjects were then debriefed, and any questions they asked about the experiment were answered.

Results

Human Evaluations

We evaluated the interrater reliability of numerical scoring by correlating the total numerical scores assigned by the original examiner and the independent evaluator. The observed correlation ($r = .92$, $p < .01$) indicated that the numerical scoring was very reliable. Only the original examiner's results are presented here because both sets of results were very similar and the original examiner's opinion is normally used in the field.

The decisions of the examiner are presented in Table 2. For analysis, the examiner's decisions were coded as follows: 1 for *truthful,* 2 for *inconclusive,* and 3 for *deceptive.* This coding scheme created an ordinal scale of decisions based on the underlying interval scale for numerical scoring. A Kruskal–Wallis nonparametric analysis of variance (ANOVA) was then used to test for differences in classifications across the three physical countermeasure conditions, and no differences were found ($H = 0.13$, *ns*). Similarly, we used the Kruskal–Wallis test to test for differences in classifications between the physical countermeasure groups and the mental countermeasure groups, and no differences were found ($H = 0.52$, *ns*). Because there were no differences among the countermeasure conditions, data from the four countermeasure conditions were collapsed for additional analysis.

Differences in classification between the innocent control and the guilty control subjects were found to be significant ($H = 14.72$, $p < .01$). The examiner correctly classified 72.5% of the innocent control and the guilty control subjects, with 12.5% inconclusive outcomes and 15% incorrect outcomes. However, the examiner correctly classified only 41.2% of the countermeasure subjects; 11.3% of the countermeasure subjects were called *inconclusive* and 47.5% were incorrectly classified. A Kruskal–Wallis analysis indicated that classification of countermeasure subjects was significantly poorer than classification of guilty control subjects ($H = 5.65$, $p < .05$). A Kruskal–Wallis analysis also indicated that the classification of countermeasure subjects was significantly different from the classification of innocent control subjects ($H = 10.23$, $p < .01$).

Table 2
Percentage of Classifications of Subjects by the Original Examiner

Condition	Classification: Truthful	Inconclusive	Deceptive
Control			
Innocent	75	15	10
Guilty	20	10	70
Countermeasure			
Pain & muscle	55	5	40
Muscle	50	10	40
Pain	45	15	40
Mental	40	15	45

Note. $n = 20$ for all conditions.

Manipulation Checks

According to subjects' posttest reports, 76% of the countermeasure subjects reported exact compliance with their instructions. There was no significant difference across the countermeasure conditions in the number of subjects who reported using their countermeasures as instructed. Only 26% of the subjects who reported compliance with their countermeasure instructions failed the CQT. However, 68% of the subjects who did not correctly follow their countermeasure instructions failed the test. None of the innocent control subjects reported the spontaneous use of a countermeasure, but 40% of the guilty control subjects reported spontaneous use of a countermeasure. However, no guilty control subject who spontaneously used countermeasures was able to pass the CQT.

The mean scores from the evaluation of the EMG data are shown in Table 3. The data from four subjects were lost because of equipment problems. A one-way ANOVA indicated significant differences among the six groups in the scores assigned to the EMG data obtained from the gastrocnemius, $F(5, 110) = 12.77$, $p < .01$, and from the temporalis, $F(5, 110) = 3.71$, $p < .01$. Post hoc tests using the LSDMOD procedure revealed that subjects in the muscle condition and in the pain and muscle condition produced gastrocnemius EMG scores that were

Table 3
Mean Electromyography (EMG) Scores Across the Six Conditions

		Muscle site for EMG			
		Gastrocnemius		Temporalis	
Condition	*n*	*M*	*SD*	*M*	*SD*
Control					
Innocent	18	0.5	1.0	2.0	3.9
Guilty	20	0.3	1.0	2.3	2.8
Countermeasure					
Pain & muscle	20	8.2	13.6	4.0	4.3
Muscle	18	15.7	12.3	2.8	4.7
Pain	20	0.9	2.3	6.6	5.9
Mental	20	0.9	2.2	1.2	4.6

Table 4
Means and Standard Deviations for the Seven Range-Corrected Difference Scores for Physiological Features Across the Six Conditions

	Control condition				Countermeasure condition							
	Innocent		Guilty		Pain & muscle		Muscle		Pain		Mental	
Feature	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Skin conductance amplitude	11.1	18.4	−14.4	18.8	−6.9	20.5	−3.2	18.7	4.0	20.9	−12.4	17.2
Skin conductance duration	13.1	22.5	−8.3	18.3	−3.3	22.2	−3.3	22.2	1.8	25.8	−10.2	13.1
Cardiovascular response amplitude	9.4	12.7	3.3	18.1	20.1	14.8	18.6	23.6	13.1	21.1	16.0	17.0
Cardiovascular response duration	8.9	19.4	7.1	23.3	15.8	23.7	24.2	28.3	12.7	20.4	22.5	28.2
Vasomotor response amplitude	7.7	17.2	−7.1	18.3	7.5	10.8	9.0	18.1	6.4	15.8	2.7	13.4
Vasomotor response duration	8.4	19.4	−0.8	18.7	5.0	15.2	7.2	19.1	4.8	16.8	1.3	14.4
Respiration length	3.3	13.2	−4.6	11.0	−1.7	14.2	1.9	17.0	−6.1	13.2	−0.1	13.5

larger than the scores produced by the other four groups (range = 4.24, $p < .05$). Post hoc tests also revealed that subjects in the pain condition produced larger temporalis scores than did subjects in the innocent control, guilty control, and mental countermeasure groups (range = 4.24, $p < .05$). No other comparisons were found to be significant.

On the basis of the subjects' posttest reports of countermeasure use, we analyzed the numerical scores of the standard physiological measures from the four countermeasure conditions with a Group (4) × Correct Use of Countermeasures (2) ANOVA. No significant effect related to the countermeasures groups was found, but the main effect of correct use of the countermeasures was significant, $F(1, 77) = 10.47$, $p < .01$. Subjects who reported the correct use of their countermeasures produced a mean numerical score of 3.30, whereas subjects who incorrectly applied their countermeasures produced a mean numerical score of −4.14.

Detection of Countermeasures

The original examiner's subjective decisions of countermeasure use were correct for only 12% of the physical countermeasure subjects. None of the mental countermeasure subjects produced behavior or physiological responses that the examiner considered to be indicative of countermeasure use. None of the spontaneous countermeasure users in the guilty control condition were detected, but the original examiner did falsely accuse 15% of the innocent control subjects of using countermeasures when they had not.

Physiological Measures

Table 4 shows the mean difference scores for seven physiological features of the six groups. We conducted a multivariate analysis of variance (MANOVA) with a between-subjects (group) factor and a within-subjects (feature) factor on the four countermeasures conditions. Neither the main effect of group nor the Group × Feature interaction were significant. Because this analysis indicated that all the countermeasures manipulations produced essentially the same effects, we combined all of the manipulations into a single countermeasure condition for additional analysis.

We conducted a second MANOVA on the innocent control, guilty control, and combined countermeasures groups. This analysis indicated a significant main effect of group, $F(2, 117) = 5.96$, $p < .01$, and of feature, $F(6, 112) = 6.70$, $p < .01$, and a significant Group × Feature interaction, $F(12, 224) = 2.90$, $p < .01$. To explore the differences among groups for individual features, we conducted post hoc tests using Scheffé's procedure. The results of these tests are shown in Table 5. The occurrence of significant differences in skin conductance amplitude and duration between innocent control and guilty control subjects and between innocent control and countermeasure subjects, but not between guilty control and countermeasure subjects, suggested that skin conductance responses were relatively unaffected by the countermeasures manipulation. However, a different picture emerged for the cardiovascular measures. Significant differences between guilty control and countermeasure conditions were found for cardiovascular response amplitude and finger pulse amplitude, but none of the vascular measures produced a significant difference between the innocent control and countermeasure subjects. There was no difference across conditions in the respiration length measure.

We used discriminant analysis to determine if an optimal classification based on the physiological features would outperform the human evaluator. A stepwise discriminant analysis between innocent control and guilty control subjects produced a significant four-variable discriminant function (canonical $r = .69$, Wilks's $\lambda = .52$, $p < .01$). The four variables in that function

Table 5
Mean Differences Between Control and Relevant Questions in Range-Corrected Measurements of the Physiological Features

Feature	Control condition: Innocent	Control condition: Guilty	Countermeasure condition
Skin conductance amplitude	11.1_a	-14.4_b	-4.6_b
Skin conductance duration	13.1_a	-8.3_b	-3.7_b
Cardiovascular response amplitude	9.4_{ab}	3.3_a	17.0_b
Cardiovascular response duration	8.9_a	7.1_a	18.8_a
Vasomotor response amplitude	7.7_a	-7.1_b	6.4_a
Vasomotor response duration	8.4_a	-0.8_a	4.6_a

Note. Common subscripts indicate homogeneous subsets within features by Scheffé's procedure when $\alpha > .05$.

and their standardized discriminant function coefficients were as follows: skin conductance amplitude, 0.95; cardiovascular response amplitude, 0.69; cardiovascular response duration, −1.04; and respiration length, 0.58. This discriminant function correctly classified 75% of the innocent control subjects and 85% of the guilty control subjects. It also correctly classified 59% of the countermeasure subjects. Over all groups, the discriminant analysis correctly classified 65.8% of the subjects. We calculated a detection-efficiency coefficient (Kircher, Horowitz, & Raskin, 1988) by coding innocent subjects as 0 and all guilty control and countermeasure subjects as 1. Truthful outcomes were coded as 0, and deceptive outcomes were coded as 2. The two coded vectors were then correlated. The resulting detection-efficiency coefficient was significant ($r = .29, p < .01$).

Because the discriminant classification procedure does not have an inconclusive zone, it was difficult to compare the results of this analysis to the human evaluator's performance. To make such a comparison, we reclassified subjects on the basis of the examiner's numerical scores so that scores less than zero were considered deceptive and scores greater than zero were considered truthful; there were no zero scores. Without an inconclusive zone, the examiner correctly classified 85% of the innocent subjects and 75% of the deceptive subjects but only 44% of the countermeasure subjects. Over all groups, the examiner was correct 55.8% of the time, and the detection-efficiency coefficient was significant ($r = .26, p < .01$).

We conducted a second discriminant analysis to see if a discriminant solution between the innocent control and countermeasure subjects might improve classification of countermeasure subjects. This analysis also produced a significant four-variable discriminant function (canonical $r = .44$, Wilks's $\lambda = .80, p < .01$). The four variables in this function and their standardized discriminant function coefficients were as follows: skin conductance amplitude, 0.53; cardiovascular response amplitude, −0.70; skin conductance duration, 0.47; and respiration length, 0.43. This discriminant function correctly classified 70% of the innocent control subjects, 78% of the countermeasure subjects, and 75% of the guilty control subjects. Over all groups, this discriminant analysis correctly classified 75.8% of the subjects and produced a significant detection-efficiency coefficient ($r = .38, p < .01$). This represented a 20% improvement in classification performance in comparison with the examiner and more than a doubling of the amount of criterion variance accounted for by the classifications.

Countermeasure Question

To examine whether the countermeasure question (see Table 1) differentiated countermeasure subjects from innocent control subjects, we conducted planned pairwise comparisons on the physiological waveforms, using the procedures described by Kirk (1968). We conducted two series of contrasts between the innocent control group and the countermeasure groups. In the first series, we searched for differential reactivity between the countermeasure question and its adjacent neutral question. In the second series, we searched for differential physiological reactivity between the countermeasure question and its adjacent relevant questions. None of these analyses produced statistically significant results.

Sex

Sex was not planned as a variable in the original design, and subjects were assigned to groups without reference to their sex. However, random assignment resulted in nearly equal numbers of men and women in each condition (two groups were imbalanced 11 to 9, one each way). Therefore, we conducted a series of analyses to explore the possible effects of sex. Sex was included as a between-subjects factor in an ANOVA of the numerical scores and a MANOVA of the physiological-feature difference scores. No significant main effects of sex or interactions with sex were found.

Discussion

The results of the present study strongly suggest that control-question polygraph tests may be defeated by guilty subjects trained in the use of physical or mental countermeasures. Accuracy of decisions for subjects who received no more than 30 min of countermeasure training was significantly lower than accuracy obtained for guilty control subjects. Decision accuracy for subjects who had been trained to perform physical countermeasures versus mental countermeasures was comparable. Approximately one half of the subjects in each countermeasure condition passed their polygraph tests. Analyses of the physiological measures indicated that the countermeasures principally affected the cardiovascular responses. Moreover, the original examiner detected only 12% of the physical countermeasure users, and none of the mental countermeasure subjects produced behavior or physiological responses that the examiner considered to be suspicious. All of these findings are consistent with prior research demonstrating that physical countermeasures are effective and are difficult to detect (Honts & Hodes, 1983; Honts et al., 1985, 1987).

The lack of difference between the physiological effects of mental and physical countermeasures suggests that mental and physical countermeasures may be mediated by the same psychophysiological mechanisms. Because the mental countermeasure required no muscular activity, the mediating mechanism is likely to have been psychological in nature. Increases in attention and mental load required to correctly perform the countermeasures appear to have made the control questions more salient than the relevant questions. These results suggest that any maneuver that sufficiently augments the task demands signaled by the occurrence of control questions might be an effective countermeasure. Clearly, activity of major muscle groups is not necessary to produce the effects. However, the manipulation checks suggested that correct performance of the countermeasure is critical for its effectiveness. Furthermore, guilty control subjects who spontaneously attempted countermeasures were not successful in defeating the CQT. These results replicate and extend earlier findings by Honts et al. (1988) by indicating not only that proper training is necessary but also that training must be implemented correctly.

The finding that mental countermeasures were just as effective as physical countermeasures has serious implications for field polygraph applications. Specifically, countermeasure detectors that depend on the measurement of muscular activity are unlikely to be effective against mental countermeasures. This hypothesis was supported in our study by the failure of EMG measures to detect the mental countermeasures. These results indicate that the movement sensors popular with the polygraph profession are likely to be inadequate for detecting subjects who are practicing mental countermeasures, and they suggest that overreliance on such technology may be risky. Moreover, the countermeasure question that we included in the question series failed to discriminate countermeasure users from innocent subjects, suggesting that when this precaution is used by field examiners it is not very helpful.

Although we demonstrated significant effects of physical and mental countermeasures on the detection of guilty subjects in this study, the use of laboratory simulations to estimate the accuracy of polygraph tests in the field has been controversial (Kircher et al., 1988). However, it is unclear how countermeasures can be studied systematically in the field because successful use of countermeasures would be nearly impossible to identify in the context of most field examinations (Honts, 1987). Nevertheless, anecdotal reports of countermeasure use in the field have suggested that countermeasures are a real threat. Raskin (1990) described a celebrated double-homicide case in which the confessed murderer described his use of biofeedback and hypnosis to defeat a CQT. Furthermore, in federal national security screening programs, some hostile intelligence agents appear to have beaten repeated polygraph tests over a number of years (Gutman, 1992; Safire, 1989; Wines & Ostrow, 1987).

The use of computers and statistical classification techniques offers potential solutions to the countermeasure problem. Consistent with prior research (Kircher & Raskin, 1988), our results revealed that decisions made based on discriminant analysis were as accurate as decisions made by human evaluators in discriminating between innocent control subjects and guilty control subjects. Moreover, despite the generally adverse effects of countermeasures on the physiological measures, we developed a discriminant function that correctly classified 70% of the innocent control subjects, 75% of the guilty control subjects, and 78% of the guilty subjects who had received countermeasure training.

These findings suggest that a statistical model may be developed to minimize the risk of false-negative errors caused by effective countermeasures. Such a model might include physiological measures not examined in the present study, and it could be applied only when testing of a subject has produced a truthful or inconclusive outcome. However, before any statistical decision model is offered as an alternative to current field practice, it should be tested and validated on independent samples of cases that are specifically representative of those that are likely to be encountered in the field.

Even if successful statistical solutions are found, their adoption by field polygraph examiners is uncertain. The polygraph community has been severely criticized for being poorly trained and technically unsophisticated (Honts, 1991; Raskin, 1986). Moreover, polygraph examiners have generally been slow to adopt computerized scoring systems, although such systems have consistently been shown to perform as well as, or better than, the best human evaluators (Honts & Perry, 1992).

In summary, those who must use polygraph test results should be advised of the dangers posed by countermeasures. Triers of fact in courts of law could then use such information to assign weight to any polygraph result by considering the likelihood that a subject could have received effective countermeasure training. As Honts and Perry (1992) have suggested, effective countermeasures clearly present a problem, but perhaps not a daunting one, for criminal justice uses of polygraph tests.

In contrast, countermeasures may pose a more serious problem for the national security system. The effective use of countermeasures may be one of several factors that contribute to the high rate of false-negative errors obtained in some studies of national security polygraph-screening programs (Barland, Honts, & Barger, 1989; Honts, 1991, 1992). Until strong evidence is developed indicating that countermeasures may be rendered ineffective by techniques that have been validated and can be practically implemented in the field, it seems prudent to limit rather than expand the use of polygraph techniques in government security screening programs.

References

Barland, G. H. (1988). The polygraph test in the U.S. and elsewhere. In A. Gale (Ed.), *The polygraph test: Lies, truth, and science* (pp. 73–95). Newbury Park, CA: Sage.

Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the accuracy of security screening polygraph examinations.* Fort McClellan, AL: Department of Defense Polygraph Institute.

Dawson, M. E. (1980). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology, 17,* 8–17.

Fowles, D. C., & Schneider, R. E. (1978). Electrolyte medium effects on measurements of palmar skin potential. *Psychophysiology, 15,* 474–482.

Fridlund, A. J. (1979). Contour-following integrator for dynamic tracking of electromyographic data. *Psychophysiology, 16,* 491–493.

Fridlund, A. J., Price, A. W., & Fowler, S. C. (1982). Low-noise optically isolated electromyographic preamplifier. *Psychophysiology, 19,* 701–705.

Geddes, L. A., & Newberg, D. C. (1977). Cuff pressure oscillations in the measurement of relative blood pressure. *Psychophysiology, 14,* 198–202.

Gutman, R. (1992, June 16). In one of its worst blunders, CIA recruited Stasi agents. *Salt Lake Tribune*, p. A9.

Honts, C. R. (1987). Interpreting research on polygraph countermeasures. *Journal of Police Science and Administration, 15*, 204–209.

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., & Hodes, R. L. (1983). The detection of physical countermeasures. *Polygraph, 12*, 7–17.

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70*, 177–187.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior, 16*, 357–379.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology, 1*, 241–247.

Honts, C. R., Raskin, D. C., Kircher, J. C., & Hodes, R. L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. *Journal of Police Science and Administration, 16*, 91–94.

Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12*, 79–90.

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.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences.* Monterey, CA: Brooks/Cole.

Lykken, D. T. (1981). *Tremor in the blood: Uses and abuses of the lie detector.* New York: McGraw-Hill.

Morris, R. A. (1989). The admissibility of evidence derived from hypnosis and polygraphy. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 333–376). New York: Springer.

Norusis, M. J. (1988a). *SPSS/PC+ advanced statistics V2.0.* Chicago: SPSS.

Norusis, M. J. (1988b). *SPSS/PC+ V2.0 base manual.* Chicago: SPSS.

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding application and acceptance of polygraph evidence. *Utah Law Review, 1986*, 29–74.

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

Raskin, D. C. (1990). Hoffman, hypnosis, and the polygraph. *Utah Bar Journal, 3*, 7–10.

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

Rovner, L. I. (1986). The accuracy of physiological detection of deception for subjects with prior knowledge. *Polygraph, 15*, 1–39.

Safire, W. (1989, February 2). The polygraph virus. *The New York Times*, p. 25.

U.S. Department of Defense. (1987). *Department of Defense Polygraph Program: Report to Congress for fiscal year 1986.* Washington, DC: Author.

U.S. Department of Defense. (1988). *Department of Defense Polygraph Program: Report to Congress for fiscal year 1987.* Washington, DC: Author.

U.S. Department of Defense. (1991). *Department of Defense Polygraph Program: Report to Congress for fiscal year 1990.* Washington, DC: Author.

Wines, M., & Ostrow, R. J. (1987, August 12). Cuban defector claims double-agents duped U.S. *The Los Angeles Times*, pp. 8, 14.

Received August 21, 1992
Revision received July 19, 1993
Accepted July 20, 1993 ■