inScience Press

Psychology Applications & Developments VII

Edited by Clara Pracana & Michael Wang

Advances in Psychology and Psychological Trends

Psychology Applications & Developments VII
Advances in Psychology and Psychological Trends Series

Edited by: Prof. Dr. Clara Pracana and Prof. Dr. Michael Wang

inScience Press

Edited by:

Prof. Dr. Clara Pracana
Portuguese Association of Psychoanalysis and Psychoanalytic Psychotherapy
Portugal

Prof. Dr. Michael Wang
Emeritus Professor of Clinical Psychology, University of Leicester
United Kingdom

Published and distributed by:

Rua Tomás Ribeiro, 45, 1ºD, 1050-225 Lisboa, Portugal
www.insciencepress.org

Printed by:

GIMA - GESTÃO DE IMAGEM EMPRESARIAL, LDA.
CET - Centro Empresarial Tejo, Rua de Xabregas Nº 6 - Lote B
1900-440 Lisboa, Portugal

Printed on acid-free paper

ISSN of Collection: 2183-2854
e-ISSN of Collection: 2184-2019
ISBN of this Volume: 978-989-54815-9-0

Legal Deposit: 449794/18

All Rights Reserved
© 2021 inScience Press
This work is under inScience Press Open Access License. This publication may be read, downloaded, printed, copied, distributed, displayed, reproduced and performed, but only for non-commercial purposes, provided acknowledgement of the original source and its author(s) is made, with a link to inScience Press.

This publication will be available online in http://insciencepress.org/ and limited hard copies can be ordered from:

InScience Press
Rua Tomás Ribeiro, 45, 1ºD
1050-225 Lisboa, Portugal

Chapter #8

HYBRID POLYGRAPH AND OCULAR-MOTOR DECEPTION TESTS FOR SCREENING AND SPECIFIC-INCIDENT INVESTIGATIONS[i]

Mark Handler[1] & Monika Nacházelová[2]
[1]*Converus, Inc., Lehi, Utah, USA*
[2]*Department of Psychology, Masaryk University, Czech Republic*

ABSTRACT
We describe two experiments combining polygraph and ocular-motor methods to detect deception. The first evaluated a test covering four issues consisting of an automated polygraph and an ocular-motor deception format. 180 participants were randomly assigned to one of three conditions. One group stole $20 from a secretary's purse and lied about it. Another group stole the $20 and a ring from a desk and lied about both crimes. The third group was innocent answering all questions truthfully. Logistic regression combined features extracted to compute the probability of deception. The probability of deception was used to classify participants as guilty or innocent. On cross-validation, classifications were 92.2% and 90.0% correct for guilty and innocent participants, respectively. The second experiment evaluated a directed-lie protocol. 120 participants were randomly assigned to guilty (steal $20) or innocent conditions. All took an automated polygraph and ocular-motor version of the test. On cross-validation, decision accuracy was 87.1% for the innocent and 85.5% for the guilty. Both experiments assessed an indirect measure of blood pressure known as pulse transit time which was diagnostic, making significant contributions to the logistic regression models. Polygraph signals contributed significantly to the decision models and produced modest improvements in classification accuracy.

Keywords: ocular-motor deception test, automated polygraph test, lie detection.

1. INTRODUCTION

Lying in interpersonal communication is a common behavior, and unfortunately, humans are good liars and/or poor lie-catchers. Research suggests about 54% of inter-personal credibility assessments are correct (Hartwig & Bond, 2011;2014; Vrij, 2008). While often trivial, lying can sometimes occur in high-stakes, important milieus. National security, criminal investigations, courtroom testimony, employment applications, relationships and political, settings are areas where lying can have serious consequences (Granhag & Stromwell, 2004).

One of the oldest tests used to improve veracity assessment is the polygraph, which is widely used (Honts, Thurber, & Handler, 2021). However, the polygraph has a long and controversial history. In the United States, federal, state, and local government agencies conduct polygraph tests to screen job applicants, test existing employees with security clearances, and conduct criminal investigations. Current estimates for U.S. federal government screening use are in the range of 70,000 examinations per year (Taylor, 2013). There are several polygraph interrogation techniques, and researchers debate their merits and limitations (Honts & Thurber, 2019; Iacono & Ben-Shakhar, 2019). The American Polygraph Association (APA) and the American Society for Testing and Materials (ASTM) attempt to set standards for what constitutes a validated polygraph technique. And while there are many techniques available, there is no international standard for what constitutes a polygraph technique.

Additionally, there is no consensus on a single theory that explains the relationship between deception and observed effects on physiological measures (National Research Council, 2003). Physiological measures include electrodermal activity, blood pressure, heart rate, peripheral vasomotor activity, and respiration. Electrodermal reactions are most diagnostic in laboratory and field settings, followed by cardiovascular and respiration reactions (Kircher, Kristjansson, Gardner, & Webb, 2012; Raskin & Kircher, 2014). The estimated percentage of correct decisions for polygraph tests is approximately 89% for specific-incident criminal investigations and 85% for screening applications (American Polygraph Association, 2011; Honts, Thurber, & Handler, 2021).

2. BACKGROUND

The Ocular-motor Deception Test (ODT) is another psychophysiological technique to detect deception (Cook, Hacker, Webb, Osher, Kristjansson, Woltz, & Kircher, 2012; Hacker, Kuhlman, Kircher, Cook, & Woltz, 2014; Kircher, 2018). It offers the benefit of a less intrusive and likely largely independent assessment of veracity compared to the traditional polygraph test. The ODT uses a remote eye tracker to monitor eye movements and pupil size while the test subject reads and answers True/False statements presented serially by a computer. The ODT assumes that it is cognitively more demanding to deceive than tell the truth, and it assumes that deception is associated with emotional arousal. If those assumptions are correct, deception during the test should cause pupil dilation, suppress eye blinks, and produce diagnostic changes in reading patterns, response time, and error rates (Bradley, Miccoli, Escrig, & Lang, 2008; Kahneman, 1973). The results of laboratory and field research are generally consistent with those predictions. ODT accuracy rates ranged from 78% to 86% based on logistic regression analysis of features extracted from ocular-motor and behavior measures (Kircher, 2018; Kircher & Raskin, 2016).

The original ODT covered two relevant issues or topics of concern. In 2020, Potts introduced a new ocular-motor test format called the Multiple-Issue Comparison Test (MCT) that covered four relevant topics rather than two (Potts, 2020). His mock crime experiment had three conditions. One group committed a single crime, another group committed two crimes, and an innocent group committed no crimes. Potts measured pupil diameter, gaze position, blink rate, response time, and error rate. On cross-validation, logistic regression analysis correctly identified 53 of 60 truthful participants (88%) and 103 of 120 deceptive participants (86%).

Polygraph research indicates that it is better at identifying liars than that to which they lie (Barland, Honts, & Barger, 1989; Department of Defense Polygraph Institute, Research Division, 1997; Podlesny & Truslow, 1993). Potts' (2020) sought to improve the classification accuracy for role discrimination.

Potts' experiment included 60 innocent participants, 60 guilty of a single mock crime (steal cash), and 60 guilty of two crimes (steal cash and a gift card). A logistic regression function correctly identified truthful and deceptive answers on all four relevant issues in 93% of innocent participants and 78% of the guilty subjects. When the decision model was trained to classify participants as deceptive if they failed on any one or more of the four relevant issues, it correctly classified 83% of innocent subjects and 87% of guilty subjects. With that classification rule, the decision on a guilty participant was correct if they appeared deceptive to any one or more of the topics covered on the test, even those answered truthfully. On the other hand, innocent participants had to appear truthful to all four relevant issues to be classified correctly. The results were promising in that Potts was able to reliably discriminate guilty roles in a laboratory experiment. Role identification would increase the utility of credibility assessment testing in the field.

We reasoned that if both polygraph and ocular-motor techniques yield diagnostic information about a person's deceptive status, we might improve accuracy by combining features extracted from the signals recorded during polygraph and ODT phases of a combined test protocol. If the polygraph is primarily emotion-based, and the ODT is primarily cognition-based, then different psychological mechanisms would underlie the two tests; and the measures they produce might make unique contributions to the logistic regression model and increase the accuracy of outcomes.

We developed a hybrid MCT test (HMCT) that combined measures obtained from a polygraph phase and a subsequent ODT phase. One reason to position the ODT after the polygraph is that the ODT asks over 200 questions, and polygraph measures tend to habituate after only a few repetitions of the test questions. In contrast, ocular-motor measures habituate slowly and retain diagnostic within-subject differences between question types over multiple repetitions of the test questions (Kuhlman, Webb, Patnaik, Cook, Woltz, Hacker, & Kircher, J 2011). Considering the different effects of habituation on polygraph and ocular-motor measures, it appeared more likely that a prior ODT would adversely affect a subsequent polygraph test than the reverse.

Another aim of the research was to evaluate a new method for measuring blood pressure during a lie-detection test. Currently, polygraph examiners use a partially inflated blood pressure cuff on the arm called the cardiograph to record changes in cardiovascular activity. The cardiograph correlated 0.84 with diastolic blood pressure recorded continuously with a medical device for monitoring blood pressure (Podlesny & Kircher, 1999).

Despite its simple design and low cost, the cardiograph is almost as diagnostic as medical-grade equipment for measuring relative changes in blood pressure in a polygraph test. However, the longer the cuff is inflated, the more uncomfortable it becomes to the test subject, limiting the number of questions polygraph examiners can ask before deflating the cuff to restore circulation to the lower arm.

For some time, psychophysiologists have explored the value of pulse transit time (PTT) as a continuous, indirect measure of arterial blood pressure (Geddes, Voelz, Babbs, Bourland, & Tacker, 1981; Obrist, Light, McCubbin, Hutcheson, & Hoffer, 1978; Obrist, Light, McCubbin, Hutcheson, & Hoffer, 1979). Contractions of the ventricles of the heart produce pulse waves that travel throughout the arterial system. PTT is the time it takes a blood pressure pulse wave to travel from the heart to a peripheral site such as a finger. As the pressure in the arterial system increases, the time it takes the pulse to travel from the heart to the finger decreases. Webb and Kircher (2005) measured PTT from the R-wave of the electrocardiogram (ECG) to the occurrence of the pulse in a finger photoplethysmogram (PPG). The R-wave of the ECG is associated with contraction of the left ventricle and initiation of the pulse wave. They found that PTT was as effective as the cardiograph for discriminating between truthful and deceptive people on polygraph tests.

3. EXPERIMENT 1

3.1. Objectives

Experiment 1 had two objectives: It assessed the efficacy of the HMCT that combined an automated polygraph test with an ODT, and it assessed the efficacy of PTT for detecting deception at the categorical level (innocent or guilty) and at the level of involvement (guilty to zero, one, or two mock crimes).

3.2. Methods

3.2.1. Participants

We recruited 180 participants with advertisements in the temporary help wanted section of an online job site (59% female). Ages ranged from 19 to 56 years (M = 28.4). Participants were offered $40 for their time and promised a $30 bonus if they passed the HMCT. Participants were randomly assigned to three groups of equal size (n=60). One group was guilty of a single mock crime. Another group was guilty of two mock crimes, and the third group was innocent.

Participants who committed a mock crime located a secretary's office in a business setting, had a brief interaction with the secretary, left the office, waited for the secretary to leave the office, and then reentered the office. Participants who committed a single mock crime stole $20 from the secretary's backpack. Participants who committed two crimes stole the $20 and a ring from the desk. After completing their tasks, participants reported to the test proctor. The proctor calibrated the eye tracker and started the computer-administered test. Guilty participants were instructed to lie to questions about one or two of the four relevant issues on the test, whereas innocent participants were instructed to answer all questions truthfully.

3.2.2. Hybrid multiple-issue comparison test (HMCT)

The HMCT covered four relevant issues. The first relevant issue was about the theft of $20 (R1); the second was about the theft of the ring (R2); the third was about the theft of a cell phone (R3); and the fourth was about the theft of a set of AirPods (R4). No participants stole a cell phone or AirPods. A computer introduced the four relevant topics, provided instructions, and presented the test questions. Participants used the left and right mouse buttons to answer test questions Yes or No questions during the polygraph phase. During the ODT phase, participants used the left and right mouse buttons to answer True or False statements.

A GP-12 Physiology Monitor (J&J Engineering, Poulsbo, WA, USA) recorded skin conductance, respiration, electrocardiogram (ECG), and finger pulses from a photoplethysmograph (PPG). A Tobii 4C remote eye tracker (Tobii, Stockholm, Sweden) attached to the bottom of the computer monitor monitored left and right pupil size, horizontal and vertical gaze position, and fixations. The GP-12 and Tobii eye tracker recorded all signals continuously while the participant completed both test phases.

3.2.3. Polygraph phase

The polygraph phase began with a preamble that introduced the relevant issues and a six-question practice test, after which the computer presented 12 Yes/No questions about each of the relevant issues (e.g., Did you steal the $20?) plus 18 neutral questions (e.g., Is looking both ways before crossing the street a wise thing to do?). The computer presented test questions aurally and visually every 22 seconds to allow physiological reactions time to recover between questions.

We arranged the test questions to form all possible pairwise comparisons of relevant issues across the three sessions. The first half of session 1 asked about R1 and R2, and the second half of session 1 asked about R3 and R4. Session 2 paired R1 and R3 and then paired R2 and R4. The third session paired R1 and R4 and then R2 and R3, completing the set of possible pairwise comparisons. The computer informed participants about the forthcoming topics during the test, e.g., "Now you will be asked about the ring and the $20." This arrangement allowed participants to focus on only two relevant issues at a time. Test questions were presented in random order, subject to the constraint that no two questions of the same type appeared in immediate succession. Between sessions, the computer asked three simple arithmetic questions to clear working memory of the test topics.

3.2.4. Ocular-motor test phase

The second phase was an ODT that contained 64 True/False statements about the four relevant issues divided into two sessions. The computer instructed participants to answer True or False statements as quickly and accurately as possible, or they might fail the test. The ODT put the test subject under time pressure to increase cognitive load. The computer then gave participants a six-item practice test with feedback about the number of statements they answered correctly and their mean response time.

During the ODT, the computer presented statements about the four relevant issues in random order, except that no two statements of the same type appeared in succession. Unlike the polygraph phase, the ODT contained no neutral statements, only statements about the four relevant topics, e.g., "I am guilty of taking the $20 from the secretary's purse." For each relevant issue, the exculpatory answer was True for half the statements and False for the remaining statements. We balanced statements concerning the four relevant issues for length, negation, and passive voice. Between the two sessions, participants answered 10 True/False arithmetic statements.

3.2.5. Polygraph measures

Skin conductance was recorded at 350 Hz from disposable Ag-AgCl snap electrodes attached to the palmar surface of the middle phalanges of two fingers on the left hand. Respiration was recorded at 350 Hz with a strain gauge in an elastic belt attached with Velcro around the chest. ECG was recorded at 1000 Hz from disposable Ag-AgCl snap electrodes attached to the dorsal surface of the left and right wrists. Finger pulses were recorded at 1000 Hz with a photoplethysmograph (PPG) attached to the middle finger of the left hand.

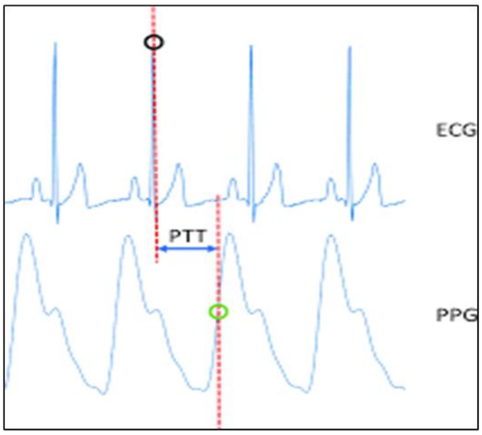
3.2.6. Pulse transit time (PTT)

Figure 1 illustrates the measurement of PTT. The algorithm identified the R-waves in the ECG (spikes) and the steepest slope in ascending limb of finger pulses recorded by the PPG. PTT was the time interval between the R-wave and steepest slope in the subsequent finger pulse. Before measuring PTT, the algorithm removed baseline drift from the ECG with a 2-pole high-pass Butterworth filter, $f_c = 2$ Hz. It then used a slope detector to identify R-waves in the ECG. It computed the range for an interval that began at the first sample and ended at the 25th (25 ms) and stored that range. It incremented the scoring window by 1 ms, measured the range of filtered ECG values for samples 2 through 26, and stored that range. The computer incremented the scoring window by 1 ms and repeated that process for the entire ECG signal. The algorithm then transformed the array of ranges to standard scores. Outliers in the array of standard scores ($z > 4$) started a forward search for a maximum z score within 70 ms. We took the occurrence of the maximum z score as an R-wave. The algorithm skipped forward 300 ms from the detected R-wave and started a new search for the next outlier ($z > 4$). This process continued until the algorithm reached the end of the array. The algorithm occasionally missed an R-wave in the ECG signal and inserted heartbeats into the array to interpolate across interbeat intervals that exceeded 1300 ms. The algorithm missed 25 of 18,372 R waves in ECGs of 13 randomly selected subjects (<.02%).

The PPG signal also was conditioned prior to the measurement of PPT. The computer smoothed the PPG with a 2^{nd}-order Savitsky-Golay filter (length = 401 ms) and then applied a 2-pole Butterworth high-pass filter (fc = 5 H.Z.) to the smoothed signal. The steepest slope in the original finger pulse wave was the maximum value in the filtered photoplethysmogram between two R-waves in the ECG (Webb & Kircher, 2005).

The computer measured the interval from each R-wave to the steepest slope in the filtered PPG signal between 90 ms and 350 ms after the R-wave. It stored PTT as a square wave at 60 Hz that showed PTT change in ms at each heartbeat.

Figure 1.
Measurement of pulse transit time (PTT) from the R-wave in the electrocardiogram (ECG) to the steepest slope in the photoplethysmogram (PPG).



3.2.7. Cohens' d feature scores

We extracted 14 features from the signals generated by the physiology monitor and eye tracker, such as the amplitude of the pupil reaction and mean pulse transit time. Most of those features are described elsewhere (Kircher & Raskin, 2002; 2016). The computer obtained a score on each feature for the 66 polygraph questions and 64 ODT statements. For some features, such as skin conductance, a high score indicated that the person showed a strong reaction to the test question. For other features, such as respiratory activity, a low score indicated that the person reacted to the question. We reversed the sign of features when relatively small values indicated strong reactions. Thus, for all features, higher scores were indicative of stronger reactions to test items.

The scores for each feature were used to compute Cohen's d, a within-subject standardized distance between relevant issues. For example, we measured the amplitude of pupil reactions to each of 48 polygraph questions, 12 for each of the four relevant topics, and calculated the mean and variance of each set of 12 within-issue measurements. The square root of the mean of the four variances provided a pooled measure of the within-issue standard deviation. Cohen's d was the difference between the mean for a relevant issue and the smallest of the four observed means divided by the pooled within-issue standard deviation. The relevant issue with the smallest observed mean served as the person's baseline, and it varied over people and features. The Cohen's d score for each relevant issue was its distance from the person's minimum relevant reaction.

3.3. Results of experiment 1

The data matrix consisted of a Cohen's d score for each participant's four relevant issues, phase (ODT or polygraph), and feature. An exploratory logistic regression identified a combination of nine features that distinguished between relevant issues answered truthfully or deceptively. The analysis selected six features from the polygraph phase and three features from the ODT phase. The most diagnostic feature was the change in pupil size during the polygraph phase of the test. PTT correlated significantly with deceptive status ($r_{pb} = -.418$, $p<.001$) and was among the variables selected for the logistic regression model.

A procedure known as k-fold validation provided estimates of how well the model would perform if tested on a new sample of cases. The "k" refers to the number of subgroups formed from the entire sample. We divided the sample of 720 relevant issues into six subsamples (k=6) of 120 issues and conducted a 6-fold validation. Of the 120 questions in each subsample (fold), participants answered 90 truthfully and 30 deceptively.

The first subset comprised a "hold-out subsample." We removed it from the dataset and combined the remaining subsets to create a training set. We used the training set to develop a logistic regression equation that was then used to classify the relevant issues in the hold-out subsample. We recorded the accuracy for the hold-out subsample. The accuracy of classifications in the hold-out subsample was less biased than the accuracy in the training set because the hold-out relevant issues were not used to optimize feature coefficients in the regression equation.

We returned the first subset to the training set and removed the second subset. The second subset served as a new hold-out subsample. We created a new logistic regression equation with all but the second subset of relevant issues. That new model was used to classify the relevant issues in the hold-out subsample, and we recorded its accuracy. We repeated this process for each of the remaining subsets. The best estimate of accuracy for the model was the mean accuracy across the six hold-out samples.

Table 1 reports the percent correct decisions for questions answered truthfully or deceptively for each hold-out subsample (fold). Accuracy estimates ranged from 80.0% to 96.7% correct. Mean accuracy was 91.1%.

Table 1.
Percent Correct Decisions for Questions Answered Truthfully or Deceptively in 6-fold Validation.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Mean
n	120	120	120	120	120	120	720
Truthful	92.2	87.8	94.4	93.3	93.3	92.2	92.2
Deceptive	93.3	96.7	93.3	80	86.7	91.7	90.0
					Mean Accuracy		91.1

The computer used the posterior probabilities of deception in the hold-out samples to classify participants as either truthful to all four of the relevant questions or deceptive to any one or more of the relevant questions. Participants were classified as innocent if the posterior probability of deception was less than or equal to .50 on all four relevant questions. Otherwise, the computer classified the participant as guilty.

3.4. Conclusions of experiment 1

We conducted z tests to compare the accuracy of the statistical classifier to chance (50%). All of the accuracy rates reported below were significantly greater than 50% at $p < .001$.

The regression equations with polygraph and ocular-motor features correctly classified 90% of the 60 innocent subjects and 91.7% of 120 guilty participants. Accuracy on cross-validation of a logistic regression model that included only ocular-motor measures was 88.3% for innocent participants and 85.8% for guilty participants. On average, accuracy was 4% higher with polygraph measures than without.

Guilty participants were deceptive to some questions and truthful to others. If we considered the classification of a guilty participant correct only when the decisions on all four relevant issues were correct, the accuracy on that group dropped from 91.7% to 75.8%.

4. EXPERIMENT 2

4.1. Objectives

We designed Experiment 2 to explore the use of polygraph and ocular-motor features for a hybrid directed-lie test. All polygraph techniques compare a person's physiological reactions to two types of questions. The person is classified as deceptive if reactions to target or relevant questions are stronger than their reactions to comparison questions. Conversely, the person is classified as truthful if reactions to comparison questions are similar or stronger than their reactions to target or relevant questions.

The directed-lie test compares reactions to relevant questions about the crime to directed lie questions about transgressions made sometime in the person's lifetime. An example directed lie question is, "Have you ever broken a rule or regulation?" Before the test, the test subject is instructed to lie to directed lie questions. Since everyone has broken a rule or regulation at some point in their life, to deny it would be a lie. Test subjects are told it is essential to know what it looks like when the person lies, and if they do not lie and react to the directed lie questions, they will fail the test. The directed-lie test predicts that innocent subjects will be more concerned about the directed lie questions and react more strongly to them than relevant questions. It also predicts that guilty subjects will be more concerned, and react more strongly, to relevant questions than directed-lie questions (Bell, Kircher & Bernhardt, 2008; Honts & Reavy, 2015).

The procedures in Experiment 2 were the same as those in Experiment 1, except where noted below.

4.2. Methods

4.2.1. Participants

We recruited a new sample of 124 participants (44% female), paid them \$40-\$70 for one hour of participation, and randomly assigned them to two groups of equal size (n=62). One group stole \$20 as described above, and the other group was innocent. Ages ranged from 19 to 74 years (M = 28.6).

4.2.2. Hybrid directed lie comparison test (HDLC)

The HDLC began with a preamble that introduced the relevant issue and described the directed lie questions. The computer then administered a practice test to ensure participants understood the requirement to lie to directed lie questions. After the practice test, the computer asked a set of 10 Yes/No questions three times in different orders at a rate of one question every 22 seconds. The set included an initial question to evoke an orienting response, three relevant questions about the theft of the $20 (R1), three directed lie questions, and three simple arithmetic questions.

In the subsequent ODT phase, the computer presented 36 T/F statements twice in different orders. Twelve of the 36 statements asked about the theft of the $20, 12 were directed-lie statements, and the remaining 12 were simple arithmetic statements.

4.3. Results of experiment 2

Eleven of 15 features extracted from the polygraph and ocular-motor phases correlated significantly with deceptive status. PTT correlated -.464 with deceptive status, $p < .001$. Five of the 11 features contributed significantly to a logistic regression equation and included PTT. The analysis indicated that features obtained only during the polygraph phase of the test contributed significantly to the regression model. Thus, the ODT phase of testing did not contribute significantly to the model.

For experiment 2, a k-fold (k=4) validation was conducted. Table 2 reports the four-fold validation of the 5-feature regression model for each of the four subsets of participants.

Table 2.
Percent Correct Decisions for Questions Answered Truthfully or Deceptively in 4-fold Validation.

	Fold 1	Fold 2	Fold 3	Fold 4	Mean
n	30	30	30	34	124
Truthful	94.1	86.7	80.0	86.7	87.1
Deceptive	76.5	73.3	93.3	100.0	85.5
			Mean Accuracy		86.3

4.4. Conclusions from experiment 2

The mean accuracy for the polygraph phase was 86.3% when the posterior probability of deception was assumed to be 0.5. However, in a criminal investigation, it may be difficult to defend a decision when the posterior probability of deception is near 0.50 (chance). We re-computed accuracy rates using posterior probabilities of deception from the four hold-out samples and classified tests as inconclusive if they produced a probability of deception between .45 and .55. Of the 124 HDLC tests, 10 (8.1%) were inconclusive. Excluding inconclusive outcomes, the mean percent correct decisions was 89.3% for innocent participants and 87.9% for guilty participants. The mean accuracy of the HDLC on cross-validation was 88.6%.

We also examined the effect of eliminating polygraph features that required the attachment of sensors to the test subject and used only measures obtained with the remote eye tracker. The accuracy based on probabilities from the 4-fold validation of a model with only ocular-motor features was 82.3% for innocent participants and 87.1% for guilty participants. The mean accuracy was 84.7%. With an inconclusive region for unbiased posterior probabilities of deception that ranged from .45 to .55, 9 of 124 tests were inconclusive (7.3%). The mean percent correct decisions was 89.5% for innocent participants and 84.5% for guilty participants. Excluding inconclusive outcomes, mean accuracy was 87.0%. Adding polygraph measures to ocular-motor measures improved decision accuracy from 87% to almost 89%.

5. FUTURE RESEARCH DIRECTIONS AND LIMITATIONS

Field research on credibility assessment is challenging because absolute knowledge of the person's deceptive status (ground truth) is rarely available, especially for innocent people. Reviews of laboratory and field studies on ODT accuracy show similar accuracies in English-speaking, Spanish-speaking, and Middle Eastern cultures (Kircher, 2018; Kircher & Raskin, 2016). Although the present findings are promising, future research should assess the accuracy of HMCT and HDLC tests in real-world settings and other cultures and evaluate the theoretical basis of ocular-motor correlates of deception.

6. CONCLUSION/DISCUSSION

The findings of both experiments indicate that polygraph and ocular-motor measures achieve high levels of discrimination between truthful and deceptive people in a laboratory setting. The findings also suggest that polygraph measures make small but significant contributions to classification accuracy compared with only ocular-motor measures. Decision accuracy with polygraph measures ranged from 86% to 91%, and accuracy without polygraph measures ranged from 85% to 87%. Whether a slight gain in accuracy justifies the inconvenience of attaching multiple sensors to the body may depend on the circumstances. For example, even a slight gain in accuracy would be justified in a capital case to ensure the most valid decision. It would not be as essential in a pre-employment screening setting when many people are competing for a few positions, and the goal is to narrow the pool of candidates for subsequent evaluations, such as background checks, work history, or psychological testing.

The results obtained with the PTT measure derived from the ECG and PPG were consistent with those reported previously (Webb & Kircher, 2005). PTT contributed significantly to the logistic regression models for both the HMCT and the HDLC. PTT is less invasive than the cardiograph and could replace the cardiograph in polygraph tests since both signals provide indirect measures of arterial blood pressure (Numaguchi, Kircher & Raskin, 1994; Podlesny & Kircher, 1999). Although PTT is less invasive than the cardiograph, it still requires that ECG and PPG sensors be attached to the test subject. In addition, PTT requires high-quality ECG and PPG recordings. Poor signal quality or artifact in either the ECG or the PPG signal corrupts PPT and causes data loss.

In Experiment 1 and Experiment 2, pupil enlargement was more diagnostic than traditional polygraph measures, and these findings agree with those reported by Webb, Honts, Kircher, Bernhardt, and Cook (2009). Among the traditional polygraph measures, changes in skin conductance are usually more diagnostic than cardiovascular or respiratory signals (Kircher & Raskin, 2002; Meijer, Selle, Elber, & Ben-Shakhar, 2014). In Experiment 1 and Experiment 2, pupil size correlated more highly with deceptive status than skin conductance and contributed more to the decision models. The present findings suggest that traditional polygraph techniques would likely benefit from measures of pupil size.

Kircher and Raskin (2016) review evidence that supports the idea that the present findings will generalize to field settings. However, we estimated accuracy with people seeking temporary work, all of whom we asked about specific incidents. Questions that address a specific event, such as the theft of $20, were justified in Experiment 2 because it tested a protocol designed for use in specific-incident criminal investigations. However, the HMCT was designed for screening applications where questions are likely to be broadly worded, such as "Did you ever commit espionage?" or "Have you withheld any work-related discipline?" The accuracy estimates in Experiment 1, where we asked participants about specific criminal acts, might not generalize to the more broadly worded questions that characterize screening tests in field settings.

REFERENCES

American Polygraph Association. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, *40*(4), 196–305.

Bell, B. G., Kircher, J. C., & Bernhardt, P. C. (2008). New measures improve the accuracy of the directed-lie test when detecting deception using a mock crime. Psychology & Behavior, *94*(3), 331–340.

Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the accuracy of security screening polygraph examinations*. DTIC AD Number A304654. Department of Defense Polygraph Institute.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S., Woltz, D. J., & Kircher, J. C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. *Journal of Experimental Psychology: Applied*, *18*(3), 301-313.

Department of Defense Polygraph Institute Research Division Staff. (1997). A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats. *Polygraph*, *26*, 79-106.

Geddes L. A., Voelz, M. H., Babbs C.F., Bourland J. D., & Tacker, W. A. (1981). Pulse transit time as an indicator of arterial blood pressure. *Psychophysiology*, *18*(1), 71–74.

Granhag, P. A., & Strömwall, L. A. (2004). *The detection of deception in forensic contexts*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511490071

Hacker, D. J., Kuhlman, B. B., Kircher, J. C., Cook, A. E., & Woltz, D. J. (2014). Detecting deception using ocular metrics during reading. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 159–216). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-394433-7.00005-1

Hartwig, M., & Bond, C.F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661-676.

Honts, C. R., & Reavy, R. (2015). The comparison question polygraph test: A contrast of methods and scoring. Physiology and Behavior, *143*, 15-26. http://dx.doi.org/10.1016/j.physbeh.2015.02.028

Honts, C.R., Thurber, S. & Handler, M. (2021). A comprehensive meta-analysis of the comparison question polygraph technique. *Applied Cognitive Psychology*, 35(2), 411-427. https://doi.org/10.1002/acp.3779

Honts, C. R., & Thurber, S. (2019). Analyzing Iacono's thought experiment about polygraph field studies: Reason or fantasy? *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice*, *48*, 76-86.

Iacono, W. G., & Ben-Shakhar, G. (2019). Current Status of forensic lie detection with the comparison question test: An Update of the 2003 National Academy of Sciences report on polygraph testing. *Law and Human Behavior*, *43*(1), 86-98.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Englewood Cliffs, N.J.

Kircher, J. C. (2018). Ocular-motor Deception Test. In P. Rosenfeld (Ed.), *Detecting Concealed Information and Deception* (pp. 187–212), Elsevier Academic Press.

Kircher, J. C., Kristjansson, S., Gardner, M. K., & Webb, A., K. (2012). Human and computer decision making in the psychophysiological detection of deception. *Polygraph*, *41*(2), 77-126.

Kircher, J. C., & Raskin, D. C. (2002). Computer methods for the psychophysiological detection of deception. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 287–326). London: Academic Press.

Kircher, J. C. & Raskin, D. C. (2016). Laboratory and field research on the ocular-motor deception test. *European Polygraph*, *10*(4), 159-172.

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the concealed information test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*(9), 879-904 National Research Council (2003). *The Polygraph and Lie Detection*. Washington, D.C.: National Academy of Sciences.

Kuhlman, B., Webb, A., Patnaik, P., Cook, A., Woltz, D., Hacker, D., & Kircher, J. (2011). Evoked Pupil Responses Habituate During an Oculomotor Test for Deception. Poster presented at Society for Psychophysiological Research, Boston, MA.

Numaguchi, G., Kircher, J.C., & Raskin, D. C. (1994). The Erlanger method for measuring cardiovascular activity: Covariation with blood pressure, blood volume, and heart rate. Final report to the U.S. Government. Scientific Assessment Technologies, 1865 Herbert Avenue, Salt Lake City, UT 84108

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1978). Pulse transit time: Relationship to blood pressure. *Behavior Research Methods & Instrumentation*, *10*, 623-626.

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1979). Pulse transit time: Relationship to blood pressure and myocardial performance. Psychophysiology, *16*(3), 292-301.

Podlesny, J. A. & Kircher, J. C. (1999) The Finapres (volume clamp) recording method in psychophysiological detection of deception examinations: Experimental comparison with the cardiograph method. *Forensic Science Communication*, *1*(3), 1-17.

Podlesny, J. A., & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, *78*(5), 788-797.

Potts, A.C. (2020). 1, 2, 3 Crimes you're out: Ocular-motor methods for detecting deception in a multiple-issue screening protocol [Unpublished doctoral dissertation], University of Utah.

Raskin, D. C., & Kircher, J. C. (2014). Validity of polygraph techniques and decision methods. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 63–129). Elsevier Academic Press.

Taylor, M. (2013, January 24). After criticism, Obama officials quietly craft new polygraph policy. *McClatchy Washington Bureau*. Retrieved from https://www.mcclatchydc.com/news/special-reports/article24743677.html

Vrij, A. (2008). Detecting Lies and Deceit: Pitfalls and Opportunities (2nd ed.). Chichester, UK: Wiley.

Webb, A. K, Honts, C. R., Kircher, J. C., Bernhardt, P. C., & Cook, A.E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminal Psychology*, *14*(2), 279-292.

Webb, A. K. & Kircher, J. C. (2005). *Use of Pulse Transit Time for the Psychophysiological Detection of Deception* (Final report to the U.S. Department of Defense). Salt Lake City: University of Utah, Department of Educational Psychology.

AUTHORS' INFORMATION

Full name: Mark David Handler
Institutional affiliation: Converus, Inc.
Institutional address: 610 S. 850 E., Suite 4, Lehi, UT 84043-2925
Email address: polygraphmark@gmail.com
Short biographical sketch: Mark D. Handler is the Director of Professional Services for Converus, Inc. He is a veteran of the United States Naval Nuclear Power Submarine Program, a former law enforcement officer and polygraph examiner, a credibility assessment instructor/lecturer, and the author of over 75 peer-reviewed articles relating to credibility assessment.

Full name: Monika Nacházelová
Institutional affiliation: Masaryk University, Faculty of Arts, Department of Psychology
Institutional address: Arne Nováka 1, 602 00 Brno, Czech Republic
Email address: mo.kupcova@gmail.com
Short biographical sketch: Monika Nacházelová has obtained her BSc degree in psychology at Juniata College, USA, and Master's degree in psychology at Masaryk University, Czech Republic. Currently she is a PhD candidate at Masaryk University, Faculty of Arts, Department of Psychology. Her research focuses on psychophysiology of deception detection, especially the Ocular-Motor Deception Test (ODT). Her main research interest centers on the effect of countermeasures on the ODT accuracy. Monika works as a lecturer at the Department of Education, Masaryk University and also works as a child psychologist at two separate facilities: Children's Hospital Brno, Department of Children's Neurology and Children's Counseling Center (PPP Brno).

[i] Address correspondence concerning this article to Contact: polygraphmark@gmail.com