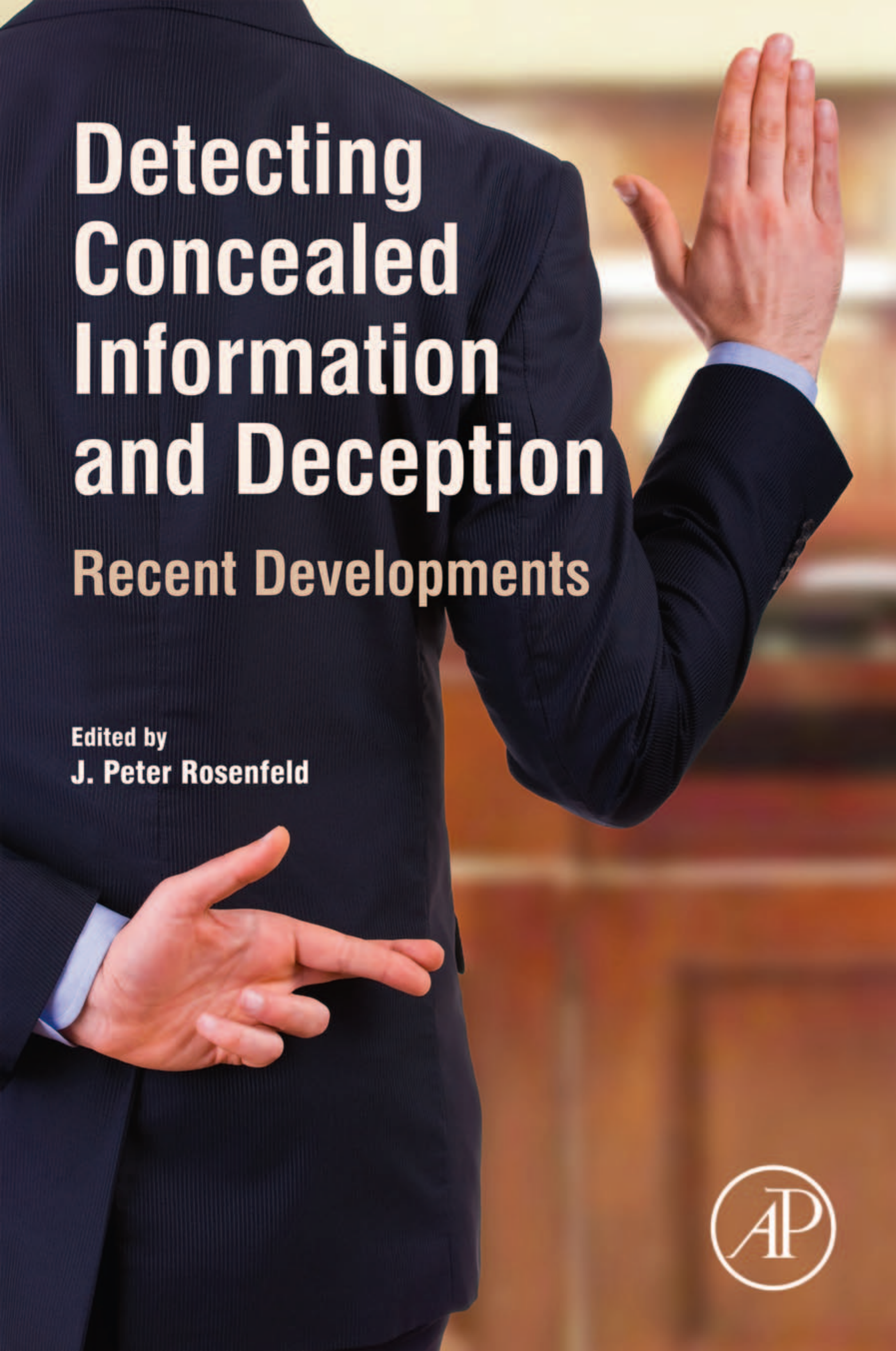

Detecting Concealed Information and Deception

Recent Developments

Edited by
J. Peter Rosenfeld

DETECTING CONCEALED INFORMATION AND DECEPTION

Recent Developments

Edited by

J. PETER ROSENFELD
Northwestern University, Evanston, IL, United States

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2018 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-812729-2

For information on all Academic Press publications visit our website at https://www.elsevier.com/books-and-journals

ELSEVIER
Book Aid International
Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org

Publisher: Nikki Levy
Acquisition Editor: Emily Ekle
Developmental Editor: Barbara Makinster
Production Project Manager: Poulouse Joseph
Cover Designer: Mark Rogers

Typeset by TNQ Books and Journals

CONTENTS

Section 1: Background, History, and Theory

CONTRIBUTORS

Wolfgang Ambach
Institute for Frontier Areas of Psychology and Mental Health (IGPP), Freiburg, Germany

Arnoud Arntz
University of Amsterdam, Amsterdam, The Netherlands

Gershon Ben-Shakhar
The Hebrew University of Jerusalem, Jerusalem, Israel

Elena Davydova
Northwestern University, Evanston, IL, United States

Eitan Elaad
Ariel University, Ariel, Israel

Matthias Gamer
University of Würzburg, Würzburg, Germany

Giorgio Ganis
University of Plymouth, Plymouth, United Kingdom

Pär Anders Granhag
University of Gothenburg, Göteborg, Sweden

John C. Kircher
University of Utah, Salt Lake City, UT, United States

Bennett Kleinberg
University of Amsterdam, Amsterdam, The Netherlands

Elena Labkovsky
Northwestern University, Evanston, IL, United States

Timothy J. Luke
University of Gothenburg, Göteborg, Sweden

Izumi Matsuda
National Research Institute of Police Science, Kashiwa, Chiba, Japan

John B. Meixner, Jr
Northwestern University, Assistant United States Attorney, Detroit, MI, United States

Merylin Monaro
University of Padova, Padova, Italy

Galit Nahari
Bar-Ilan University, Ramat-Gan, Israel

Tal Nahari
The Hebrew University of Jerusalem, Jerusalem, Israel

Hiroshi Nittono
Osaka University, Suita, Osaka, Japan

Akemi Osugi
Forensic Science Laboratory, Hyogo Prefectural Police Headquarters, Kobe, Hyogo, Japan

Yoni Pertzov
The Hebrew University of Jerusalem, Jerusalem, Israel

J. Peter Rosenfeld
Northwestern University, Evanston, IL, United States

Giuseppe Sartori
University of Padova, Padova, Italy

Nathalie klein Selle
The Hebrew University of Jerusalem, Jerusalem, Israel; University of Amsterdam, Amsterdam, The Netherlands

Evan Sitar
Northwestern University, Evanston, IL, United States

Kristina Suchotzki
University of Würzburg, Würzburg, Germany

Yaloe van der Toolen
University of Amsterdam, Amsterdam, The Netherlands

Bruno Verschuere
University of Amsterdam, Amsterdam, The Netherlands

Aldert Vrij
University of Portsmouth, Portsmouth, United Kingdom

Anne Ward
Northwestern University, Evanston, IL, United States

Joshua Wasserman
Northwestern University, Evanston, IL, United States

Andrea Zangrossi
University of Padova, Padova, Italy

PREFACE

Until about the year 2000, most field tests of deception involved the Comparison Question Test (CQT; formerly, the Control Question Test), a questioning protocol virtually always utilized with the subject connected to a polygraph machine. This machine typically recorded autonomic nervous system (ANS) responses, including skin resistance, cardiovascular activity, and breathing pattern, in conjunction with the relevant and control questions of the CQT. On the other hand, from about the 1960s forward, many deception research studies utilized a different questioning protocol called the Concealed Information Test (CIT; formerly, Guilty Knowledge Test), but also in conjunction with use of a polygraph tracking ANS responses. There were various reasons why CIT proponents rejected the CQT questioning approach, including the criticism that ANS responses to relevant questions about a suspect's personal crime involvement (e.g., Did you shoot your spouse?) could never be compared in a meaningful scientific way with ANS responses to so-called control questions (e.g., Did you ever think violent thoughts?). Such a comparison was the heart of the deception detection matter in the CQT, whose critics rightly pointed out the lack of standardization involved in interrogations designed to identify and formulate control questions for various subjects. In contrast, the CIT approach asked informational questions about crime details that would likely be known by perpetrators but not innocents. The comparison made in CIT research was between the ANS response to critical versus irrelevant items, all drawn from the same category. This comparison or difference is called the CIT effect. Thus, the guilty party, but not the innocent suspect, would recognize the presentation of the murder weapon (e.g., 356 Magnum) in a set of other possible murder weapon presentations (e.g., 45 Automatic, 38 Revolver, 22 Beretta, etc.), and this recognition would be signaled by relatively altered ANS responses only in the guilty suspect.

Deception research with other response systems in addition to the ANS—especially involving the central nervous system—began in the 1980s, and a burgeoning growth of all deception research work plus the introduction of yet more novel measurement methods and protocols was seen following the terrorist attack on the twin towers in New York on September 11, 2001. It is on this research that the present volume focuses. Much of the new work is by academic researchers, and is focused mainly on the CIT. Examples include chapters by myself on the now sizeable literature on event-related electroencephalography EEG potentials (especially P300) as signs of information recognition; by Ganis on the use of functional magnetic resonance imaging also to index recognition; by Gamer and Pertzov, and by Kircher on the use of oculomotor signs of familiarity and recognition; and by Sartori and by Suchotzki on behavioral indices (including the novel autobiographical Implicit Association Test and other manual dynamics measures) of recognized true versus false information. These four approaches discuss possible applications of these various novel dependent measure channels for use in field investigations. Another set of approaches to deception detection in field situations is based on novel analyses of verbal behavior. Some of this work is closely tied to considerations of the cognitive loading effects of deception. The chapters by Granhag and Luke, Vrij, and G. Nahari exemplify this approach.

Yet despite these many examples of clearly field-oriented research areas deemed critical for an up-to-date review of the field of deception detection—a goal of this book—it seemed essential for a volume like this one to include at the outset a background section devoted to a historical perspective and theoretical consideration of the psychological principles underlying the detection of concealed information and deceptive behavior. Ambach and Gamer review the physiological measurements traditionally used in conjunction with detection of concealed information. Matsuda and Nittono provide a parallel review, more oriented to central nervous system indices, and then give an original theoretical reconsideration of the roles of recognition and concealment phenomena in memory detection. Continuing this theoretical approach, Klein Selle, Verschuere & Ben Shakhar give a full traditional account of the CIT effect in terms of orienting and response inhibition theories, informed by novel findings suggesting response fractionation. Ben Shakhar and Tal Nahari consider the very important question of the external validity of CIT research by providing a thorough review of this complex literature. As a conclusion to this section, Osugi finally bridges the transition to the novel applications section by discussing how the ANS-based CIT is used in field tests in Japan, the only nation presently using this protocol as a standard technique in field investigations.

The final section of this volume considers special issues relating to modern detection of concealed information and deception. Elaad reviews psychosocial and psychophysiological correlates of self-assessed deceptive skills in individuals. Then Kleinberg reviews the topic of assessing deception on a large scale; that is, in many persons at the same time. This matter is crucial for the currently topical problem of antiterror screening at transportation portals. Finally, and importantly, attorney and biological psychologist Meixner provides a uniquely enlightened consideration about the possible admissibility of concealed information protocols in US courts.

Thus, this volume attempts to provide a comprehensive, up-to-date review of the state of the art in detection of concealed information and deception, against a background of the theoretical foundation of this area. The chapters should be of interest to forensic, clinical, and cognitive psychologists, neuroscientists, attorneys, and those interested in the new crossover field of law and neuroscience.

J. Peter Rosenfeld

CHAPTER 9

Ocular-Motor Deception Test

John C. Kircher
University of Utah, Salt Lake City, UT, United States

The present chapter summarizes the theoretical assumptions that guided development of the Ocular-Motor Deception Test (ODT), the logic that underlies its relevant-comparison test format, and empirical evidence of its criterion-related validity. The chapter then outlines areas in need of research including mechanisms responsible for observed effects on ocular-motor measures and generalizability. Elsewhere, Hacker, Kuhlman, Kircher, Cook, and Woltz (2014) summarize the physiological basis of ocular-motor measures and psychological factors other than deception that can affect those measures.

OVERVIEW OF THE OCULAR-MOTOR DECEPTION TEST

The Ocular-Motor Deception Test (ODT) is an automated psychophysiological test for deception designed for use in a screening environment. A computer presents voice-synthesized instructions followed by written true/false test statements concerning the examinee's possible involvement in illicit activities. The computer informs examinees that if they do not answer quickly and accurately, they will fail the test. The computer then presents a single true/false statement in the center of the screen. The examinee reads the statement and presses a key to answer true or false. Half a second later, the computer presents the next statement. While the examinee reads and responds to test items, a remote eye tracker records eye movements and changes in pupil size 60 times per second (60 Hz). The computer measures response times and error rates, extracts features from recordings of gaze position and pupil size, combines its measurements in a logistic regression equation to compute the probability of deception, and classifies the individual accordingly.

The ODT uses a test format known as the Relevant Comparison Test (RCT). The RCT includes statements about the two relevant issues (R1 and R2). The RCT uses the difference between reactions to the two sets of relevant statements to determine if the examinee was truthful or deceptive

Detecting Concealed Information and Deception
ISBN 978-0-12-812729-2
https://doi.org/10.1016/B978-0-12-812729-2.00009-4

Table 9.1 A Subset of Test Statements for an Ocular-Motor Deception Test

Type	Statement	Expected answer
Neutral	The sky is blue on sunny days.	True
R1	I was uninvolved in the theft of the $20.	True
R2	I copied the credit card information from the computer.	False
R1	I admit to stealing the cash that was in the secretary's purse.	False
Neutral	I am reading this on a day that is not Sunday.	True
R2	The stolen credit card information is not in my possession.	True
Neutral	Trees that grow in the forest are never harvested for lumber.	False
R2	I made a copy of the professor's credit card.	False
R1	I did not leave the office until I had taken the $20 that was in the purse.	False

to either of the relevant issues. Each relevant issue serves as a control for the other. If the examinee reacts more strongly to statements concerning one of the two issues, the ODT classifies that person as deceptive about that relevant issue. Examinees who show little or no difference in reactions to the two sets of relevant statements are classified as truthful to both issues.

True/false statements about neutral topics are intermixed with the R1 and R2 statements. We designed the neutral statements to require relatively little cognitive effort and an opportunity for recovery from reactions to the prior statement. Table 9.1 contains a portion of a sequence of statements in an ODT.

RATIONALE UNDERLYING THE OCULAR-MOTOR DECEPTION TEST

The ODT is based on two assumptions: it assumes that deception is cognitively more demanding than telling the truth, and it assumes that deception is associated with emotional arousal. The cognitive workload hypothesis appears throughout the literature on deception detection techniques (e.g., Johnson, Barnhardt, & Zhu, 2005; Kircher, 1981; Raskin, 1979; Steller, 1987; Vrij, Fisher, Mann, & Leal, 2006). All examinees must comprehend the test statement, evaluate its relationship with autobiographic memory, and make a motor response. In addition, a deceptive individual must distinguish between two classes of test items: statements answered truthfully and statements answered deceptively. When they recognize a statement as inculpatory, they must inhibit the correct, truthful answer and issue an incorrect, deceptive one, and they must do so consistently, quickly, and accurately over the course of the test. Whereas truthful individuals should attend similarly to the two sets of relevant statements, we expect deceptive individuals to invest more mental effort when they process potentially incriminating statements. While they perform the task, deceptive individuals also may self-monitor their performance for signs that they are revealing their deception, for example, by answering too slowly or by making mistakes.

The recruitment of mental resources to accomplish these additional cognitive and meta-cognitive activities could explain effects on pupil dilation, eye movements, response time, and error rates. For instance, pupil size has been found to covary with level of difficulty on cognitive tasks such as mental arithmetic (Ahern & Beatty, 1979; Bradshaw, 1968), rehearsal of digit strings (Kahneman & Beatty, 1966; Klingner, Tversky, & Hanrahan, 2011), sentence processing (Just & Carpenter, 1993), letter processing (Beatty & Wagoner, 1978), and lexical tasks (Hyönä, Tommola, & Alaja, 1995). Consistent with the cognitive workload hypothesis, deception has been associated with pupil enlargement (Berrien & Huntington, 1943; Dionisio, Granholm, Hillix, & Perrine, 2001; Heilveil, 1976; Lubow & Fein, 1996), and evoked pupil reactions have been found to discriminate between truthful and deceptive individuals in common polygraph test formats (Bradley & Janisse, 1981; Webb, Honts, Kircher, Bernhardt, & Cook, 2009). Research on eye movements have shown that the number and duration of fixations increase and intersaccade differences decrease when people experience difficulty reading text (Rayner, 1998; Rayner & Pollatsek, 1989). If deceptive individuals find it more difficult to read and respond to inculpatory statements, eye movement reading patterns could be diagnostic. Finally, Seymour et al. have published several studies showing effects of concealing information on response times (Seymour & Fraynt, 2009; Seymour & Kerlin, 2008; Seymour, Seifert, Shafto, & Mosmann, 2000). Consistent with the increased workload hypothesis, deception was associated with longer response times.

In addition to association with increased cognitive workload, the ODT assumes that deception is associated with emotional arousal. Whether examinees are truthful or deceptive, they are likely to believe there is a chance they will fail the test, and if they fail, they will experience negative consequences. Whereas deceptive examinees are expected to be most concerned about the subset of relevant test items answered deceptively, truthful examinees should be equally concerned about both sets of relevant statements. Differential concern over the consequences of detection for one or the other relevant issue could contribute to interaction effects on pupil and other physiological measures that distinguish deceptive from truthful individuals. The research by Bradley and Janisse (1981) and Webb et al. (2009) is consistent with the idea that emotional stimuli are associated with sympathetically mediated pupil enlargement (Bradley, Micolli, Escrig, & Lang, 2008), and there is substantial literature on effects of deception on other sympathetically mediated measures in concealed information (Ben-Shakhar & Furedy, 1990; Elaad & Ben-Shakhar, 2006) and probable-lie deception tests (Kircher & Raskin, 2001).

RELEVANT COMPARISON TEST

We originally proposed the RCT as a new polygraph test format for use at ports of entry to screen travelers for trafficking of drugs or transporting explosives (Kircher, Kristjansson, Gardner, & Webb, 2012). The Computerized Screening System (CSS) was not conceptualized as a primary screening system. Rather, we thought it might be used as a secondary or tertiary assessment if there was reason to believe that a passenger posed a threat to other travelers or infrastructure. We tested the CSS in a mock-crime experiment. Some guilty participants transported what appeared to be illegal drugs ($n = 119$), other guilty participants transported a device that appeared to be a bomb ($n = 111$), and a third group was innocent of both crimes ($n = 124$). All participants were instructed to deny involvement in either crime and were promised and paid a monetary bonus if they could pass the test. A laboratory assistant attached the physiological sensors and ran a computer program that presented prerecorded auditory instructions and relevant questions about the drugs (e.g., Did you take illegal drugs from a locked cabinet?), relevant questions about the bomb (e.g., Did you put a bomb in a flight bag?), and neutral questions (e.g., Is this the year 1996?).

Deceptive answers to questions about drugs (R1) or explosives (R2) were associated with increases in skin conductance, systolic blood pressure, diastolic blood pressure, total peripheral resistance, and pupil diameter (PD); and decreases in finger pulse amplitude and respiration, but there were no effects on stroke volume or cardiac output. On cross-validation, mean accuracy of classification into drugs, bomb, and innocent groups was 67.5%. Although an accuracy rate of 67% represents a 34% improvement in accuracy over the chance probability of a correct decision for three groups (33%), decision accuracy was insufficient to recommend use of the CSS as a supplemental screening system at ports of entry.

THE RELEVANT COMPARISON TEST AND RELEVANT—IRRELEVANT TEST

Except in rare circumstances, an RCT would be problematic for specific-incident testing because it would be difficult to identify a credible, unrelated comparison issue for the particular matter under investigation. Reid (1947) once suggested that so-called "guilt-complex" questions about a fictitious crime could serve as a control for the relevant issue. Unfortunately, in an actual criminal investigation, people usually are well aware that they are suspected of involvement in a particular crime long before they are asked to take a polygraph test. By that time, it would be difficult to convince them that they are suspects in another crime. Even if it were possible to convince examinees that the authorities suspect them of a fictitious crime, the value of the guilt-complex question would be short-lived given the ready availability of information about various polygraph techniques on the Internet. Knowing that one of the relevant issues on the test is fictitious would likely cause innocent examinees to focus more on questions that address the real crime, leading to high false-positive rates. For these reasons, though conceptually sound, the guilt-complex question is impractical and rarely used (Ben-Shakhar & Furedy, 1990; Krapohl & Shaw, 2015).

The RCT is not well suited to specific-incident criminal investigation, but it might be used for screening applications. Currently, the US federal government relies on the polygraph for preemployment screening of applicants for positions in law enforcement and for periodic tests of employees with security clearances (DoDPI, 2002). In 2011, over 90% of polygraph examinations conducted by the US Department of Defense were for screening rather than criminal investigation (Office of the Under Secretary of Defense for Intelligence, 2011). Although most agencies use probable-lie or directed-lie polygraph formats for these applications, some still use a test format known as the Relevant—Irrelevant (RI) test (Krapohl & Rosales, 2014). The RI screening test includes questions about several relevant topics such as illegal drug use, past criminal activity, and falsification of the job application. The test also includes questions about irrelevant (neutral) topics such as "Are the lights on in this room?" Applicants who are deceptive to any one or more of the relevant issues are likely to perceive those questions as threats and react more strongly to them than to questions about neutral topics. However, because the relevant questions are easily identified as more important to the outcome of the test than irrelevant questions, truthful subjects also are likely to be more attentive to the relevant questions and react more strongly to them, resulting in high false-positive error rates.

Consistent with these predictions, Horowitz, Kircher, Honts, and Raskin (1997) conducted a mock crime experiment and compared reactions to relevant questions to those produced by neutral questions. They correctly classified 100% of deceptive but only 22% of truthful participants. Subsequently, Krapohl and Rosales (2014) obtained similar results in a field study of the RI test. They reported 81.5% correct decisions on deceptive cases but only 47% correct decisions on truthful cases.

Although there is good reason to expect that the RI test will have low accuracy on truthful cases when reactions to relevant and irrelevant questions are compared, it is not clear that all field polygraph examiners who use the RI test format make decisions based on such comparisons. Indeed, there are no formal rules for evaluating the polygraph protocols from RI tests (Bancroft, 2015). Some examiners might compare reactions of relevant questions to those of irrelevant questions, whereas others might compare reactions to different relevant questions. It may be that accuracy on truthful cases was higher in the Krapohl and Rosales study than in the Horowitz et al. experiment because some field examiners based their decisions on comparisons of reactions to different relevant questions. Lack of standardization and variability in the procedures examiners use to decide if a person was deceptive on the test limits the reliability and validity of the RI polygraph test. Nevertheless, if polygraph examiners were to base their decisions on comparisons of reactions to relevant questions, then the RI format would share some essential features with the RCT.

APPLICATIONS OF THE OCULAR-MOTOR DECEPTION TEST

Similar to the RI test, the ODT is designed to screen applicants for employment or to conduct periodic assessments of individuals subject to some restrictions, such as government employees with security clearances or people on parole or court-ordered restrictions. In contrast to the RI test and all other polygraph tests, the ODT does not require a trained polygraph examiner. The ODT takes less time than a polygraph test, and it is less invasive because it does not require attachment of surface electrodes or other sensors to the examinee. For a given application, the pretest information, instructions, test items, analysis, and interpretation of the data are standardized.

Because the ODT is faster and less costly than a polygraph test, an agency might use it at the front end of a screening program to reduce the number of applicants that move on to the next more costly stage of screening. There might be an advantage in using the ODT in tandem with the polygraph to minimize the risk of a particular type of error. For example, if the goal were to minimize the risk of false positive errors, and each of two independent tests had a false positive rate of 20%, then the risk that a truthful person would fail both tests would be $0.2 \times 0.2 = 0.04$, or only 4%. Of course, we do not know the extent to which ODT and polygraph outcomes are independent, and a reduction in the risk of one type of error (false positive) would increase the risk of the other error (false negative). Thus, if the two independent tests each had false negative rates of 20%, then the probability that a deceptive person would fail the first test and fail the second test would be 0.8×0.8, or 64%. Stated differently, there would be a 36% chance that a deceptive person would pass at least one of the two tests and continue on as a candidate for employment. The false positive error rate on truthful individuals would be only 4%, but 36% of deceptive individuals would pass through the screening system. Still, if the ODT and polygraph were at least partially independent, then use of the ODT and polygraph in combination could reduce the risk of a particularly undesirable decision error.

MOCK CRIME LABORATORY RESEARCH ON THE OCULAR-MOTOR DECEPTION TEST

We have conducted a series of mock crime laboratory experiments to determine if ocular-motor measures discriminate between truthful and deceptive people, and we borrowed those procedures from our laboratory research on polygraph techniques (Podlesny & Raskin, 1978). Realistic mock crime experiments produce diagnostic effects on electrodermal, cardiovascular, and respiration reactions that are similar to those obtained from actual suspects in specific-incident criminal investigations (Kircher, Horowitz, & Raskin, 1988; Kircher, Raskin, Honts, & Horowitz, 1994).

In our ODT experiments, we recruit participants from the university campus or the general community for pay and randomly assign them to guilty and innocent treatment conditions. Guilty participants commit a mock crime and then lie about it on the test. In one experiment, we instructed one group of guilty participants to take $20 from a secretary's purse and another group to download credit card information from a professor's computer. In other experiments, to simplify the procedures, we told all participants that guilty subjects committed one of two crimes, but in actuality, guilty participants committed only one crime. Because truthful and deceptive examinees in field settings usually are highly motivated to pass the test, we promised all participants a monetary bonus that would double their pay if they were able to pass the test.

OCULAR-MOTOR DECEPTION TEST ADMINISTRATION

Examinees were seated at a computer with a keyboard in a small room without windows and indirect lighting. Over the years, we have used several different eye trackers. In our last several experiments, we used a remote 60-Hz eye tracker that was affixed to the bottom of the computer monitor (SMI REDm, Sensomotoric Instruments, Berlin). The examinees placed their chin in a chin rest positioned approximately 70 cm from the monitor. To calibrate the eye tracker, the examinee gazed at an illuminated disk that appeared in several locations of the screen. Calibration was necessary to determine where fixations were in relation to the text.

The computer informed examinees with written and audio-based instructions that they would be tested about two relevant issues. The computer instructed the examinee to read and answer each true/false statement by pressing one of two keys on the keyboard. The computer also informed them that the test was based on the idea that it is more difficult to lie than to tell the truth, that deceptive people respond more slowly and less accurately than truthful people, and it was in their best interest to answer all the statements as quickly and accurately as possible. We provided this information because we believe that the effects of deception on cognitive load would be reduced if examinees chose to take a long time to consider each statement before they answered.

The standard ODT consisted of a set of 48 test statements: 16 statements concerning one relevant issue (R1), 16 statements concerning the other relevant issue (R2), and 16 neutral statements. The expected, exculpatory answer was True to half of each type of statement (e.g., I did not take the $20 from the secretary's purse.) and was False to the remaining statement (e.g., I am guilty of taking the $20 from the secretary's purse.). The test began with two neutral statements to give the examinee an opportunity to orient to the task. Thereafter, we ordered statements such that no two statements of the same type appeared in succession. The computer presented a written statement in black font on a gray background on a single line in the middle of the screen beginning on the left side. We used black font on a gray background to minimize effects of changes in illumination on the pupil. The examinee read the statement and pressed a key to answer True or False. The examinee's answer appeared on the right side of the monitor adjacent to the text for 500 ms, at which time the computer replaced the statement with the next item. When the examinee completed the block of 48 statements, the computer presented a brief unrelated task to clear working memory of the test statements. For example, examinees might have been asked to indicate if each of 10 simple arithmetic statements was true or false (e.g., $4 + 5 = 8$). The computer then presented the 48 ODT statements again in a different order. This process was repeated a total of five times. Altogether, the eye tracker provided recordings of gaze position and left and right pupil size at 60 Hz for 80 R1 statements (16 statements × 5 repetitions), 80 R2 statements, and 80 neutral statements. The speed at which examinees answered the statements typically varied between 2 and 4 s.

Cook et al. (2012) described an experiment in which all guilty participants were deceptive to statements about the theft of cash from a purse. The control issue was the theft of an exam from a professor's office. Mean change in pupil size is shown in Fig. 9.1 for 4 s following the onset of the neutral, cash, and exam statements. As predicted, guilty participants (left) reacted more strongly to cash than exam statements, whereas innocent

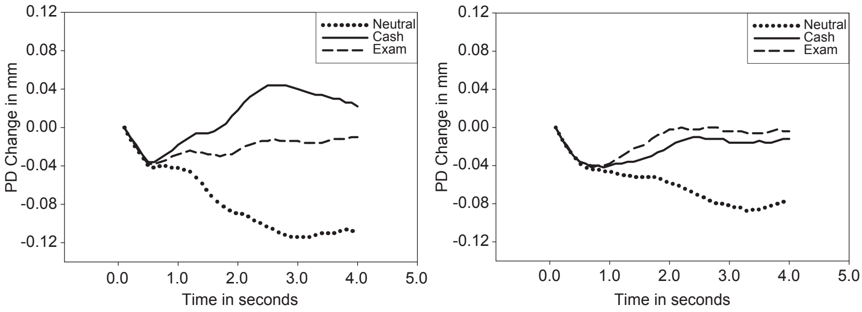

Figure 9.1 Mean change in pupil diameter (PD) from statement onset for guilty (left) and innocent participants (right).

participants (right) showed little difference between cash and exam statements. The mean change in PD associated with deception was less than 0.1 mm but is evident with signal averaging.

FEATURE EXTRACTION

For each test item, the computer extracted a set of physiological, reading, and behavioral measures. Depending on the particular eye tracker we used at the time, we recorded PD from only the right eye or from both eyes. The computer extracted two features from each signal independently. Prior to feature extraction, we replaced data losses due to eye blinks with interpolated values and smoothed the signal with a 0.5 s Savitsky-Golay filter that used linear and quadratic components to predict the midpoint of a sliding 0.5 s interval. The computer then transformed the smoothed time series of PD samples to standard scores. From the standardized signal, the computer extracted the area under the evoked pupil response. Integration of the area under the curve began at a low point that followed statement onset and lasted until the pupil response curve returned to the initial low point or to the end of the 4 s interval, whichever occurred first (Kircher & Raskin, 2001). The second feature was the level (mean) of the standardized response curve from 0.5 s before the examinee's answer to 0.5 s after the answer.

Reading was characterized by measures derived from eye fixations on the test statement. To compute fixations, we used an algorithm developed by the Applied Science Laboratory (Bedford, MA). Briefly, the computer scanned the 60 Hz series of horizontal and vertical gaze positions for periods of little movement in either direction, where movement was measured in degrees of visual angle. Periods of quiescence less than 100 ms or greater than 1000 ms were considered outside the acceptable range and were not considered fixations (Rayner, 1998). The algorithm used the mean of horizontal and vertical samples that met measurement criteria for a fixation to determine the X and Y coordinates for the fixation on the computer screen. The duration of each fixation in ms was based on the number of samples; that is, (number of samples/60) × 1000 (Cook et al., 2012).

The computer derived measures of reading behavior from fixations that fell within the area of interest. *Number of fixations* was a count of the fixations in the region of interest. *First pass duration* was the sum of fixation durations for all fixations that occur in the forward direction (left-to-right) in the region of interest before a fixation fell outside the region of interest. First pass duration was a presumed measure of lexical processing during which the reader determined the meaning of words. *Re-read duration* was the sum of durations of fixations in the region of interest that followed leftward saccades and may reflect higher-order cognitive activities, including readers' efforts to resolve comprehension failures (Hacker et al., 2014).

Behavioral measures included response time and errors. *Response time* was the time in ms from the appearance of the test item on the computer monitor to the moment the examinee pressed a key to answer True or False. *Errors* were proportions of test items of a given type answered incorrectly.

Periods during which a person is deceptive have been associated with reductions in eye blinks, whereas periods following deception have been associated with increased blink rates (Leal & Vrij, 2008, 2010; Marchak, 2013). During an eye blink, the eye tracker loses its image of the eye and there is a brief period of data loss. When the eye opens, the eye tracker reacquires the signal and resumes storage of gaze position and pupil size.

In our experiments, we measured the number of times we lost data over a 3 s interval prior to the examinee's answer (*item blink rate*), and again for 3 s after the examinee answered (*next item blink rate*). Because two statements of the same type never were presented in immediate succession, when the examinee was deceptive on the ODT, a statement that was answered truthfully always followed a statement that was answered deceptively. We expected that deceptive individuals would show a reduction in blink rates on incriminating items followed by an increase on the next item; and we expected that truthful individuals would show little difference among statement types.

Statistical adjustments for individual differences are common in psychophysiological research. As noted earlier, we transformed pupil size in mm to standard scores within item blocks. Although we have not observed an advantage to standardizing reading measures, we did divide each reading measure by the number of characters in the statement to adjust for differences in the length of test items. For response time, we transformed raw response times for the 48 items within each block to standard scores. Finally, we transformed the proportion of incorrect answers to R1 and R2 statements for the entire test to a z-test statistic for the difference between proportions.

DISCRIMINATING FEATURES

The computer calculated the mean of its 80 measurements of a given feature for each statement type (neutral, R1, and R2). With three levels of statement type, there were two degrees of freedom, and we could compute two orthogonal contrasts. Since Patnaik (2015), we have focused exclusively on the (R1-R2) contrast to reduce the number of measures and minimize opportunities to capitalize on chance when we construct multivariate decision models.

RELIABILITY AND VALIDITY OF OCULAR-MOTOR MEASURES

Table 9.2 reports internal consistency statistics (Cronbach's alpha) for the various ocular-motor measures from two dissertation experiments (Patnaik, 2015; Webb, 2008). For each participant, we computed a value for each (R1-R2) feature contrast for each block of 48 test items, and used alpha to assess the extent to which measurements from the five repetitions were consistent. If one presentation of test items suggested that the subject was deceptive, did the remaining four presentations of test items also suggest that the person was deceptive?

Table 9.3 reports validity coefficients for the various features in four mock-crime experiments, three of which were dissertation projects (Osher,

Table 9.2 Reliability Coefficients in Laboratory Experiments

	Webb (2008)[a]	Patnaik (2015)[b]	Mean
Pupil diameter			
Area under the curve	0.609	0.615	0.612
Level at answer	0.465	0.510	0.488
Reading			
Number of fixations	0.528	0.627	0.578
First pass duration	0.508	0.540	0.524
Reread duration	0.494	0.397	0.446
Behavioral			
Response time	0.397	0.329	0.363
Error rate	0.184	0.209	0.197
Blink rate			
Item blink rate	0.391	0.182	0.287
Next item blink rate	0.251	0.351	0.301

[a] Webb's (2008) dissertation experiment was reported as Experiment 2 in Cook et al. (2012).
[b] Patnaik's (2015) dissertation experiment has not been published.

Table 9.3 Validity Coefficients in Laboratory and Field Studies of the Ocular-Motor Deception Test

	Osher (2006)[a]	Webb (2008)[b]	Patnaik (2015)	Patnaik et al. (2016)	Kircher and Raskin (2016)[c]	Mean[d]
Sample size	40	112	80	145	154	
Pupil size						
Area under the curve	**0.550**	**0.464**	**0.586**	**0.546**	**0.484**	**0.517**
Level at answer	NA	**0.523**	**0.585**	**0.587**	**0.536**	**0.556**
Reading						
Number of fixations	**−0.555**	**−0.529**	**−0.406**	−0.139	**−0.202**	**−0.310**
First-pass duration	−0.075	**−0.530**	**−0.253**	**−0.452**	−0.074	**−0.301**
Reread duration	**−0.562**	**−0.489**	**−0.342**	**−0.192**	**−0.287**	**−0.332**
Behavioral						
Response time	**−0.489**	**−0.480**	**−0.497**	**−0.544**	**−0.474**	**−0.499**
Error rate	NA	0.057	0.093	0.056	**−0.370**	−0.071
Eye blink rate						
Item blink rate	NA	−0.071	**−0.388**	**−0.260**	−0.059	**−0.175**
Next item blink rate	NA	0.079	−0.088	0.049	0.023	0.025

Bolded validity coefficients were statistically significant at $P < 0.05$.

[a] One condition in Osher's (2006) dissertation experiment was reported as Experiment 1 in Cook et al. (2012).

[b] Webb's(2008) dissertation experiment was reported as Experiment 2 in Cook et al. (2012).

[c] Field study of applicants for government positions with $n = 83$ truthful and $n = 71$ deceptive applicants.

[d] Significance of mean correlation was based on total available sample size ($N = 531$ or 491).

2006; Patnaik, 2015; Webb, 2008), and one a field validity study (Kircher & Raskin, 2016). The validity coefficients were point-biserial correlations between the (R1-R2) contrast and deceptive status, where deceptive status was coded 0 if the examinee was truthful and coded 1 if the examinee was deceptive to the R1 issue. These correlations indicate the extent to which the feature discriminated between truthful and deceptive individuals. The squared point-biserial correlation is equivalent to the estimated η^2 measure of effect size. The results in Table 9.3 represent only standard testing conditions, as described earlier, and are neither exhaustive nor representative of our research on alternative test protocols that yielded inferior results.

Although the reliability coefficients presented in Table 9.1 for the various features were lower than those commonly reported for established psychological tests, they were similar to those obtained for automated polygraph systems (Kircher et al., 2012). As compared to reliability coefficients, the validity coefficients in Table 9.2 provide more information about the usefulness of ocular-motor features for detecting deception. A validity coefficient indicates the extent to which the variable discriminates between groups of truthful and deceptive individuals. The correlation of the variable with the dichotomous criterion is the figure of merit with regard to its criterion-related validity (Nunnally, 1978). Nevertheless, the low reliability values indicate that we might improve the diagnostic validity of all the available ocular-motor measure with better test construction, longer test length, improved instrumentation, or better algorithms. For example, item blink rate was not highly correlated with deceptive status ($r = -0.175$), but it also was not reliably measured ($\text{alpha} = 0.287$). If we can develop an algorithm that distinguishes bona fide eye blinks from other failures of the tracker to monitor the eyes, we should be able to improve the diagnostic validity of this measure. Although response time is highly correlated with deceptive status, we might increase its correlation with deceptive status by measuring response time from the first fixation in the area of interest, rather than from when the computer presents the statement. In general, the reliability data suggest that there is significant room for improvement in test construction, administration, instrumentation, or analysis.

Examination of the mean validity coefficients indicate that the pupil measures were more diagnostic than reading, behavioral, and blink rate measures. The (R1-R2) contrast for response time was almost as diagnostic as were the pupil measures. On average, error rates were not diagnostic, but in the field study, error rates were moderately correlated with deceptive status. Blink rate measures were the least predictive of deceptive status. The

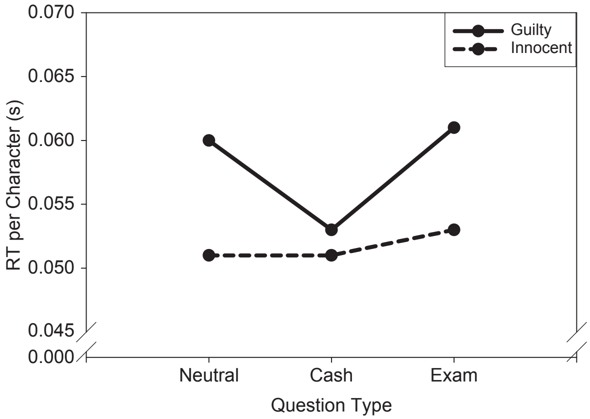

Figure 9.2 Response time for guilty and innocent groups per question type.

results also indicate that the effects on pupil size and response time were consistent across experiments and settings. Effects of deception on reading, error rate, and blink rate measures were more variable.

The two pupil measures correlated positively with deceptive status. Whereas truthful people reacted similarly to relevant and control statements across all measures, deceptive individuals reacted more strongly to relevant statements, as illustrated in Fig. 9.1. For all other measures, the correlations were negative. As compared to truthful subjects, deceptive individuals made fewer fixations, spent less time reading, and spent less time rereading relevant than control statements. Results from Webb's (2008) dissertation experiment illustrate the general nature of this effect. Fig. 9.2 shows that her guilty participants, on average, took longer than innocent participants to answer. However, when guilty participants were deceptive to cash items, their response times were shorter than when they answered truthfully to neutral and exam (control) items. This pattern of results suggests that deceptive examinees invested more mental effort in processing the relevant than control statements, as indicated by increases in PD and a reduction in blink rate to cash items. I believe they did so because they wanted to make a rapid response when they were deceptive to avoid detection. The later effect was evident in measures of response time, number of fixations, first pass duration, and reread duration.

DECISION MODELS

To classify individuals as truthful or deceptive, we used a logistic regression equation or discriminant function to compute the probability of deception from a subset of optimally weighted ocular-motor measures. If the probability of deception exceeded 0.5, we classified the person as deceptive; if the probability was less than 0.5, we classified the person as truthful. The weights for measures in the decision model were optimal in the sense that they attempted to maximize the percentage of individuals classified correctly.

Kircher and Raskin (2016) summarized the accuracy of classifications using our standard mock-crime protocol and standard ODT. Those results are reproduced in Table 9.4. The decision models yielded approximately 86% correct classifications in the original, standardization sample, and 83% correct when tested on independent samples (cross-validation). Generally, accuracy was higher for innocent (84.1%) than for guilty participants (82.1%). We attributed the relatively poor performance on cross-validation in the Osher (2006) study to the small number of participants and small subject-to-variable ratio.

Table 9.5 summarizes results from nonstandard conditions (Kircher & Raskin, 2016). Osher (2006) found that serial presentations of individual test statements (Table 9.4) yielded better ocular-motor data than did the simultaneous display of multiple test statements (Table 9.5). Webb (2008) found that sex did not moderate the effects of deception on ocular-motor measures, whereas higher motivation to pass the test and semantic simplicity in phrasing of test statements improved the diagnostic validity of some ocular-motor measures.

Together, the USTAR and Patnaik (2013) studies indicated that test statements that referred directly to the matter at hand (I did not take the $20.) produced stronger reactions in deceptive individuals than did statements that indirectly asked if the person falsified their answers on a pretest questionnaire about their involvement in the crime (I did not falsify my answers on the questionnaire about the $20.). In the NSA studies, we recruited employees and tested them about minor security violations. The studies used a nonstandard protocol because we relied on self-report for ground truth, and we were not permitted to provide meaningful incentives to government employees to pass the ODT. The agency did allow us to offer participants 1 h of release time to participate and a second hour of release time if they passed the test. In addition, most of the participants were federal polygraph examiners who may have participated because they were curious about a new technology for credibility assessment, not because they were trying to earn an hour or two of release time.

Table 9.4 Percent Correct Decisions Under Standard Conditions in Mock-Crime Experiments

Experiment	Independent variables	N	n_G	n_I	Guilty	Innocent	Mean	Validation$_G$	Validation$_I$	Mean
Osher (2006)[a]	Issues; serial format	40	20	20	85.0	85.0	85.0	85.0	70.0	77.5
Webb (2008)[b]	Sex; motivation; difficulty	112	56	56	82.1	89.2	85.7	89.3	80.4	84.9
Patnaik (2013)[a]	Direct interrogation	48	24	24	83.3	95.8	89.6	83.3	83.3	83.3
Patnaik (2015)[a]	Distributed item types; pretest feedback; postresponse interval	80	40	40	82.5	90.0	86.3	80.0	90.0	85.0
Patnaik et al. (2016)[c]	Language; culture	145	82	63	84.1	87.3	85.7	81.9	87.5	84.7
Middle East[d]	Language; culture	112	51	61	80.4	88.5	84.5			
Middle East[e]	Language; culture	101	52	49				75.0	85.7	80.4
Standard Protocol		638	325	313	82.8	89.0	85.9	82.1	84.1	83.1

[a] Validation results were obtained with the leave-one-out procedure.

[b] We used the decision model based on Patnaik et al. (2016) to classify participants in Webb's (2008) dissertation.

[c] The decision model based on Webb's (2008) dissertation was used to classify participants in Patnaik et al. (2016).

[d] The decision model was developed on this Middle Eastern sample.

[e] The decision model was tested on this Middle Eastern sample.

Table 9.5 Percent Correct Decisions Under Nonstandard Conditions in Mock-Crime Experiments

Experiment	Independent variables	N	n_G	n_I	Guilty	Innocent	Mean
Osher (2006)	Issues; parallel format	40	20	20	70.0	95.0	82.5
USTAR[a,b]	Pretest questionnaire; issues	71	47	27	59.6	77.8	68.7
NSA[a,c]	Standardization	94	51	43	72.5	88.4	80.5
NSA[a,c]	Validation	60	34	26	50.0	80.8	65.4
Patnaik (2013)[a]	Indirect interrogation	48	24	24	58.3	79.2	68.8
Patnaik (2015)[a]	Blocked	80	40	40	77.5	85.0	81.3
Nonstandard Protocols		393	216	180	65.3	84.5	74.9

[a] Unpublished.
[b] Utah Science, Technology, and Research Initiative.
[c] National Security Agency.

Patnaik (2015) found that the standard sequencing of neutral, R1, and R2 statements yields more accurate outcomes than does the presentation of several items of the same type in sequence. Patnaik also found that feedback about speed and accuracy on a pre-ODT practice test and lengthening the interval between the answer and the presentation of the next item had no discernible effect on outcomes. Patnaik et al. (2016) found that the effects on ocular-motor measures were similar for tests administered to university students in their native language in the United States and Mexico. The experiments conducted in the Middle East required modification of the software to present Arabic text from right to left. Accuracy rates on cross-validation in the Middle East were lower than those obtained in the United States and Mexico, particularly for guilty participants. Although the differences in decision accuracy between Middle Eastern and Western participants were not statistically significant, we found it necessary to reduce the number of repetitions of test statements for measures of pupil response to achieve near-comparable levels of accuracy for Arabic-speaking participants as for English- and Spanish-speaking participants. It is possible that differences between Middle Eastern and Western cultures or their languages moderated the effects of deceptive status on ocular-motor measures.

FIELD STUDY OF THE OCULAR-MOTOR DECEPTION TEST

We are interested in developing a screening test, but in our laboratory experiments, we ask participants to commit a specific mock crime. The effect sizes on ocular-measures in the laboratory are encouraging, but questions can be raised about the generalizability of these effects to field settings for screening applications.

To address these concerns, we conducted a field validity study of the ODT that evaluated applicants for positions in the Mexico attorney general's office, immigration, and federal police (Kircher & Raskin, 2016). We compared reactions to statements about recent use of illegal drugs (R1) to statements about either corruption or affiliation with a religious terrorist organization (R2). We had ground truth on the issue of corruption because it involved communication with ODT test developers, and we assumed that no applicants were affiliated with a religious terrorist organization because the base rate of that activity is very low. Confirmation of deception on the ODT was based on admissions of illegal drug use by applicants during a subsequent polygraph test, or the applicant failed a hair or urine test for prohibited substances ($n = 71$). We planned to use negative hair and urine test results to establish that applicants for positions at immigration had been truthful on the ODT. However, of the 35 applicants at that organization who confessed, 32 passed the urine test (91% false negatives) and 24 passed the hair test (69% false negatives). Therefore, we had no confidence that a person who passed the drug tests was, in fact, truthful on the ODT; urine and hair tests miss far too many deceptive individuals.

Since passing a drug test was inadequate to establish conclusively that an applicant was truthful on the ODT, we created a second ODT and administered it to applicants for positions in immigration to determine if they had committed espionage (R1) or sabotage (R2). We assumed that all the tested individuals were truthful in their answers to both relevant issues because the base rates of deception on those issues are very low, especially for people who have had no prior government employment and no apparent access to state secrets or equipment ($n = 83$).

To develop and validate a decision model with the field data, we extracted ocular-motor measures from the eye tracker data and used linear regression to select a subset of four measures to distinguish between the confirmed truthful and deceptive groups. We then used the selected variables in a five-fold validation of a logistic regression model to classify cases as truthful or deceptive. To conduct the five-fold validation, we partitioned

Table 9.6 Accuracy Rates for Five Independent Subsamples

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Mean
	N = 30	N = 30	N = 31	N = 31	N = 32		N = 154
Truthful	75.0	87.5	88.2	88.2	100.0	87.8	86.1
Deceptive	100.0	71.4	85.7	78.6	86.7	84.5	

the sample of 154 field cases into five random subsamples such that each subsample consisted of approximately 20% of the deceptive cases (n = 14 or 15) and 20% of the truthful cases (n = 16 or 17). The first subsample of 14 truthful cases and 16 deceptive cases (N = 30) was removed, and a decision model was created with the remaining truthful and deceptive cases in subsamples 2, 3, 4, and 5 (N = 124). We used that decision model to classify the holdout sample of 30 cases and recorded the percent correct for truthful and for deceptive cases in the holdout sample. The second subsample then was set aside (N = 30), a new decision model was developed with the remaining cases in subsamples 1, 3, 4, and 5 (N = 124), and the accuracy of classifications was calculated for the second holdout sample. We repeated this process for the remaining three subsamples. The results are reproduced in Table 9.6.

Consistent with the observed similarity in effect sizes for ocular-motor measures in laboratory and field settings shown in Table 9.1, decision accuracy in an actual screening context with applicants for positions in the Mexican government was similar to that obtained in mock-crime experiments. On average, the standard ODT produces between 80% and 86% accuracy in laboratory and field settings.

LIMITATIONS AND AREAS OF FUTURE RESEARCH

We know little about the relative importance of cognition and emotion in the ODT. We assumed that being deceptive is cognitively more demanding than being truthful, and we attempted to design a test that would reveal the effects of cognitive workload on physiological, reading, and behavioral measures. The data generally are consistent with the cognitive workload hypothesis. However, for most people, taking a deception test is unusual, and that request often occurs when adverse consequences to the individual are associated with failing the test. Under these conditions, we can expect the general levels of arousal to increase to a greater or lesser degree depending on the individuals' deceptive status, the perceived consequences of failing the test, and their disposition. Examinees should be invested in the outcome, and we have evidence from the NSA studies and Webb's (2008) dissertation that low levels of motivation reduce accuracy. Unless the individual is motivated to pass, the relevant items will not be perceived as threats to that end. An enhanced sensitivity to the particular subset of test items that an individual perceives as threats could explain effects on ocular-motor measures just as well as differential cognitive workload. Research that explores the roles of cognition and emotion in the ODT would contribute to our understanding of mechanisms responsible for the observed effects on outcome measures.

Alone, the ODT will not mitigate practical concerns about screening large numbers of people for threats to national security that occur only rarely in the target population (National Research Council, 2003). For example, screening tests for espionage and sabotage are unlikely to be useful because the base rate of deception is so low. Even if a test is 90% accurate, about 10% of the tested population would fail it, and the vast majority of those individuals who fail the test would be innocent of the crimes. Certainly, no single test would provide a solution to the problem of identifying the rare spy in a population of people with security clearances, although a series of screens with criteria set to avoid missing the deceptive individual could be a way to reduce the pool of possible threats to national security (Krapohl & Stern, 2003). Although screening for such low probability events is problematic, other undesirable behaviors are far more common and would be candidates for a moderately effective screening technology such as the ODT.

Our field validation study revealed that the same ocular-motor measures that are most effective in mock crime experiments also are most effective when testing job applicants in a screening environment. It was encouraging to learn that the accuracy rates achieved in a field setting were at least as high as those obtained in our laboratory experiments. Moreover, the similarity between effect sizes obtained in laboratory and field settings suggest that the mock crime paradigm is an ecologically valid means of conducting research on the ODT. It remains to be seen if discrepancies between the two settings in reading and error rate measures are systematic or due to chance. More data would help.

Although the field study was important, the five-fold validation was flawed in the sense that the entire sample of confirmed cases was used to select variables for the decision model. In the five-fold validation, only the weights for the variables changed from one phase of the validation process to the next, not the variables themselves. The decision model from the current field study should be reevaluated and refined with independent and representative samples from this and other target populations.

Unpublished efforts to assess credibility with the ODT in Colombia were unsuccessful. Although the data were limited, the ODT appeared to work well when we tested well-educated people who had applied to work for an airline, but the ODT was ineffective when we tested less well-educated applicants for security companies. We hypothesized that the reading ability of applicants for security companies may have been inadequate. If a person struggles to read and comprehend the test items, those difficulties might overshadow effects of deception on our measures. Since those early efforts to conduct research in Colombia, we began to use response times and error rates to determine whether or not a person has sufficient reading ability to take the test. In addition, we are exploring alternative, audio-based ODTs that may or may not include electrodermal, cardiovascular, or respiration measures. With an audio-based format, we would lose the eye movement–based reading measures, but we might gain diagnostic information from another physiological channel. Preliminary results suggest that an audio version will work, but we do not yet know if the audio version will be as effective as the standard reading version.

Theoretically, the RCT should misclassify examinees who are deceptive to both sets of relevant statements. If examinees are equally concerned about the two relevant issues, there should be no difference in their cognitive or emotional responses to those to those issues, and the algorithm should misclassify those individuals as truthful. We conducted one laboratory study in which one of four groups was deceptive to both sets of relevant items (USTAR, unpublished). Consistent with these predictions, accuracy on deceptive individuals was near chance. However, deception to both relevant issues was confounded with several other factors that distinguished the USTAR study from our other experiments. Patnaik (2013) explored one possibility that the adverse effects on accuracy in the USTAR study were a consequence of testing participants on whether they had falsified information on a pretest questionnaire about the crime, rather than asking if they committed the crime. Asking if the participant committed the crime was more effective than asking if they lied on a pretest questionnaire about their involvement in the crime. However, we have not yet tested the possibility that the RCT does not work for examinees who are deceptive to both relevant issues, which also might explain the high false-negative error rate in the USTAR study.

One potential solution to this problem is to construct ODTs that pair a high base-rate relevant issue, such as drug use, with a low base-rate relevant issue, such as espionage. Among federal employees, both relevant issues have face validity because employees know that those issues are of concern to their employer. Although being deceptive to both issues would be no more common than being a spy, if a person is deceptive to both issues on the ODT, we would expect the person to fail the test because the consequences of failing on the espionage issue are far more severe than failing on the drug issue. We have not tested this prediction.

To date, we have conducted no research to investigate the effects of countermeasures on ODT outcomes. We are about to start a mixed-methods investigation of countermeasures against the ODT. We will provide half of the guilty and half of the innocent participants with detailed information about how the ODT works and how we use the various ocular-motor measures to make a decision. The remaining guilty and innocent participants will serve as controls and not be so informed. Following the ODT, the experimenter will conduct interviews with the participants and ask them to complete a posttest questionnaire. From those participants' reports, we will attempt to identify strategies people develop to pass the test. In subsequent research, we would train participants to use those strategies that appear to help deceptive individuals defeat the test and attempt to develop counter-countermeasures.

DISCLOSURE

The author has a financial interest in Converus, Inc. (www.converus.com), a company that has commercialized the technology described in this report. I have disclosed those interests to the University of Utah and have in place an approved plan for managing any potential conflicts that arise from involvement in Converus.

REFERENCES

Ahern, S. K., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*, 1289–1292.

Bancroft, H. H. (2015). Polygraph screening. In D. Krapohl, & P. Shaw (Eds.), *Fundamentals of polygraph practice*. Cambridge, MA: Academic Press.

Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science, 199*, 1216–1218.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Spring-Verlag.

Berrien, F. K., & Huntington, G. H. (1943). An exploratory study of papillary responses during deception. *Journal of Experimental Psychology, 32*, 443–449.

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18*, 307–315.

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

Bradshaw, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology, 20*, 116–122.

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.

Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology, 38*, 205–211.

DoDPI. (2002). *Law enforcement pre-employment test*. Department of Defense Polygraph Institute. Available online at https://antipolygraph.org/documents/dodpi-lepet.pdf.

Elaad, E., & Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *International Journal of Psychophysiology, 61*, 226–234.

Hacker, D. J., Kuhlman, 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.

Heilveil, I. (1976). Deception and pupil size. *Journal of Clinical Psychology, 32*, 675–676.

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.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48A*, 598–612.

Johnson, R., Jr., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research, 24*, 386–404.

Just, M. A., & Carpenter, P. A. (1993). The intensity of dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310–339.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583–1585.

Kircher, J. C. (1981). *Psychophysiological processes in the detection of deception*. Salt Lake City, UT: Department of Psychology, University of Utah (Unpublished manuscript).

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., 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. (2001). Computer methods for the psychophysiological detection of deception. In M. Kleiner (Ed.), *Handbook of polygraph testing*. 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.

Kircher, J. C., Raskin, D. C., Honts, C. R., & Horowitz, S. W. (1994). Genereralizability of statistical classifiers for the detection of deception. *Psychophysiology, 31*, S73 (Abstract).

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology, 48*(3), 323–332.

Krapohl, D., & Rosales, T. (2014). Decision accuracy for the relevant-irrelevant screening test. *Polygraph, 43*(1), 20–29.

Krapohl, D. J., & Shaw, P. K. (2015). *Fundamentals of polygraph practice*. New York: Elsevier.

Krapohl, D. J., & Stern, B. A. (2003). Principles of multiple-issue polygraph screening: A model for applicant, post-conviction offender, and counterintelligence testing. *Polygraph, 32*, 201–210.

Leal, S., & Vrij, A. (2008). Blinking during and after lying. *Journal of Nonverbal Behavior, 32*, 187–194.

Leal, S., & Vrij, A. (2010). The occurrence of eye blinks during a guilty knowledge test. *Psychology, Crime and Law, 126*, 349–357. https://doi.org/10.1080/10608316090276843.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied, 2*, 164–177.

Marchak, F. (2013). Detecting false intent using eye blink measures. *Frontiers in Psychology, 4*. https://doi.org/10.3389/psyq.20133.00736.

National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Office of the Under Secretary of Defense for Intelligence. (2011). *Department of defense polygraph program process and compliance study report.*

Osher, D. (2006). *Multimethod assessment of deception: Oculomotor movement, pupil size, and response time measures* (Doctoral dissertation). University of Utah, Department of Educational Psychology.

Patnaik, P. (2013). *Ocular-motor methods for detecting deception: Direct versus indirect interrogation* (Master's thesis). University of Utah, Department of Educational Psychology.

Patnaik, P. (2015). *Oculomotor methods for detecting deception: Effects of practice feedback and blocking* (Doctoral dissertation). University of Utah, Department of Educational Psychology.

Patnaik, P., Woltz, D. J., Hacker, D. J., Cook, A. E., Ramm, M. L., Webb, A. K., & Kircher, J. C. (2016). Generalizability of an ocular-motor test for deception to a Mexican population. *International Journal of Applied Psychology, 6*(1), 1–9.

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology, 15*, 344–358.

Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans*. New York: Wiley & Sons.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Reid, J. (1947). A revised questioning technique in lie-detection tests. *Journal of Criminal Law and Criminology, 37*, 542–557.

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback, 34*(3), 177–187.

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology, 22*, 475–490.

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". *Journal of Applied Psychology, 85*, 30–37.

Steller, M. (1987). *Psychophysiologische Aussagebeurteilung* [Psychological assessment]. Gottingen: Hogrefe.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10*, 141–142.

Webb, A. K. (2008). *Effects of motivation, and item difficulty on oculomotor and behavioral measures of deception* (Doctoral dissertation). University of Utah, Department of Educational Psychology, ISBN 9780549980032.

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.

FURTHER READING

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349–350.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science, 143*, 1190–1192.

Honts, C. R. (2012). Countermeasures and credibility assessment. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 131–156). Elsevier.

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall.

Loewenfeld, I. E. (1999). *The pupil: Anatomy, physiology, and clinical applications* (Vol. 1). Boston: Butterworth-Heinemann.

Office of Technology Assessment. (1983). *Scientific validity of polygraph testing: A research review and evaluation*. OTA-TM-H-15. Washington, DC: U.S. Government Printing Office.

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.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*, 241–255.

Steinhauer, S. R., Boller, F., Zubin, J., & Pearlman, S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology, 20*, 472.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology, 21*, 22–33.

USTAR. (2010). *Oculomotor deception detection. Unpublished research funded by the Utah Science Technology and Research Initiative*. University of Utah, Department of Educational Psychology.

INDEX

‘*Note*: Page numbers followed by “f” indicate figures, “t” indicate tables and “b” indicate boxes.’

Detecting Concealed Information and Deception

Recent Developments

Edited by

J. Peter Rosenfeld

Detecting Concealed Information and Deception brings together the world's leading experts on all aspects of concealed information detection. The book examines an array of different methods—behavioral, verbal interview, and physiological—and chapters address how to make use of detected information for present and future legal purposes. With a theoretical and empirical foundation, *Detecting Concealed Information and Deception* also covers burgeoning new human interviewing techniques, including the highly influential Implicit Association Test, among others.

- Presents research from Concealed Information Test (CIT) studies
- Explores the legal implications and admissibility of the CIT
- Covers electroencephalography, event-related brain potentials, and autonomic detection measures
- Reviews multiple verbal lie detection tools
- Discusses ocular movements during deception and evasion
- Identifies how to perceive malicious intentions
- Explores personality dimensions associated with deception including religion, age, and gender

ISBN 978-0-12-812729-2

9 780128 127292

ELSEVIER

ACADEMIC PRESS

An imprint of Elsevier

elsevier.com/books-and-journals