1, 2, 3 CRIMES YOU'RE OUT: OCULAR-MOTOR METHODS FOR DETECTING DECEPTION IN A MULTIPLE-ISSUE SCREENING PROTOCOL

by

Andrew Carlos Potts

A dissertation proposal submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Educational Psychology

The University of Utah

August 2020

ABSTRACT

Previous laboratory and field studies have demonstrated that the Ocular-motor Deception Test (ODT) accurately discriminates between truthful and deceptive individuals. The ODT uses the Relevant Comparison Test (RCT), a test format that asks examinees about their involvement in two relevant issues, although examinees can be classified as deceptive to only one issue. The present study investigated whether ocular-motor measures can discriminate between truthful and deceptive individuals and identify the specific crime deceptive individuals committed on a test that asks about four relevant issues.

One hundred and eighty participants were recruited from the community and the University of Utah campus. Sixty participants stole $20 (*cash*), 60 participants stole $20 and a VISA gift card (*cash+card*), and the remaining 60 participants were innocent (*innocent*). Participants were asked about their involvement in four mock crimes: theft of $20, theft of a VISA gift card, vandalism of a parking kiosk, and filing a false police report. *Cash* participants were deceptive to cash statements, *cash+card* participants were deceptive to cash statements and card statements, and *innocent* participants were truthful to all statements. Reactions to cash, card, and vandalism statements were compared to those on false report statements to determine deception. After a participant finished the ODT, they completed a vocabulary test to assess their levels of crystallized intelligence.

As predicted, cash participants showed significant changes in pupil dilations and reading behaviors to cash statements, and cash+card participants showed significant changes in pupil dilations and reading behaviors to cash and card statements. A logistic regression function correctly classified 83.3% of innocent participants, 91.7% of cash participants, and 85.5% of cash+card participants. For innocent participants, 90% of cash items, 91.7% of card items, and 96.7% of vandalism items were accurately identified. For cash participants, 93.3% of cash items, 83.3% of card items, and 96.7% of vandalism items were accurately identified. For cash+card participants, 75% of cash items, 68.3% of card items, and 93.3% of vandalism items were accurately identified. Area under the receiver operating curve was .92 for cash classifications and .85 for card classifications.

Limitations of the present findings and implications for field applications are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ACKNOWLEDGEMENTS

Dr. John Kircher provided me with the opportunity to attend graduate school and conduct research on the ODT. For that, I will always be indebted to him. I imagine I tested his patience on more than one occasion, but he never ceased to help me and push me to become a better scientist. I will always be grateful for the impact he has had on me as a researcher and as a person.

I would like to thank my committee, Drs. Andrea Webb, Anne Cook, Dan Woltz, Doug Hacker, and Seung-Hee Claire Son. Their knowledge and insight helped me on multiple projects. Each one of them is a consummate scientist.

I would also like to thank Hannah Michael Wright, Julia Berger, Matthew Orr, Taren Rohovit, and Wei Wei for their help in running participants. This project would not have been successful without them.

I would like to thank JoLynn Yates for all of the assistance she provided not only for my dissertation, but throughout my graduate career.

Thank you to my parents who always stressed the importance of education. I would not be here were it not for their encouragement to always increase my education.

And finally, a thank you to my wife, KayCee. I could not have done this without her love and support.

INTRODUCTION

In spite of the National Research Council's (NRC) warning that the polygraph's accuracy in screening applications was "insufficient to justify reliance on its use in employee security screening in federal agencies" (National Research Council, 2003), federal agencies still administer more than 70,000 screening examinations each year (Taylor, 2013). The NRC's critiques of screening examinations stem from the protocol's inherent ambiguity. Screening examinations cover several different topics that can have varying degrees of overlap with one another. Test questions are general in scope and may cover a broad time period. For example, a question on a screening test might be, "Have you ever committed a serious crime?" The lack of a specified time interval may render it difficult for an examinee to recall specific events. But the defining characteristic of a true screening test is the absence of a well-defined relevant issue (American Polygraph Association, 2011; Nelson, 2015). These characteristics can result in a vague test where the examinees may not know if they are guilty of the activities covered by the questions.

The ambiguous nature of screening examinations also complicates attempts to assess the protocol's true validity. Laboratory experiments assign participants to commit one or several mock crimes and ask the participants about their involvement in the crime(s). This scenario results in issues that have no overlap with each other, questions that ask about a narrow time frame, and a well-defined relevant issue that elicits an indisputable deceptive or credible response from the examinee. These differences

between laboratory studies and real screening applications reduce the external validity of the laboratory studies and could result in inflated estimates of accuracy (NRC, 2003). As such, critics argue that laboratory studies cannot be used to establish accuracy in field applications (Iacono & Patrick, 1987). Nevertheless, laboratory studies that use mock-crime scenarios have been shown to provide useful information about field polygraph techniques (Honts & Thurber, 2019; Pollina et al., 2004) and can shed light on the accuracy of screening polygraphs (NRC, 2003). More generally, Anderson et al. (1999) concluded that “the psychological laboratory has generally produced truths, rather than trivialities.”

In multiple-issue screening laboratory studies, accuracy levels are significantly greater than chance for dichotomous classifications of deceptive and truthful. In Correa and Adams (1981), participants completed a preemployment data sheet, and an experimenter and the participant identified nine potential lie questions. Deceptive participants were instructed to lie to all nine questions, and truthful participants were instructed to answer all of the questions truthfully. Another experimenter then administered a preemployment polygraph examination. The polygraph examiner accurately classified 100% of the participants as truthful or deceptive. However, the test did not use the broadly-worded relevant questions that typify screening examinations.

Barland, Honts, and Barger (1989) conducted three experiments to assess the validity of screening examinations. In Experiment 1, participants were assigned to one of three groups. One group committed acts of simulated espionage, the second group went through "knowledge" scenarios where they met someone who claimed to have committed espionage and tried to recruit them to do likewise, and the third group was innocent.

Relevant questions that asked about unspecified security violations were worded in general terms to more closely represent a screening examination. Polygraph examiners accurately classified 94% of the innocent participants, but only 34% of the guilty/knowledgeable participants were correctly classified as deceptive.

Experiment 2 attempted to determine if one triple-issue test was superior to three single-issue tests. Participants were assigned to be guilty of zero, one, two, or three different acts of mock espionage or sabotage. Half of the participants were tested with one triple-issue test, and the other half were tested with three single-issue tests. For the multiple-issue approach, 55% of the innocent participants were classified as truthful, 18% were incorrect, and 27% were inconclusive. Excluding inconclusives, 75% of the innocent participants were correctly classified. Of the guilty participants, 67% were classified as deceptive to at least one crime, 5% as deceptive to none of the crimes, and 28% were inconclusive. Excluding inconclusives, 93% of the guilty participants were classified as deceptive to at least one crime. For innocent participants in the single-issue approach, 42% of the outcome decisions were correct, 8% incorrect, and 50% inconclusive. Excluding inconclusives, 83% of the innocent participants were correctly classified as truthful. For guilty participants in the single-issue approach, 82% were classified as deceptive to at least one crime, 8% deceptive to no crimes, and 10% were inconclusive. Excluding inconclusives, 91% of the guilty participants were classified as deceptive to at least one crime. The authors found there was no significant difference in the accuracy of the two approaches.

Experiment 3 assessed effects of the specificity of the relevant questions and time lag. Participants were assigned to an innocent or guilty condition. Half of the participants

were administered a polygraph test the same day they received their condition assignment, and the other half were administered the examination 6 weeks later. Two types of polygraph tests were conducted. Half of the participants were given an examination with specific relevant questions, and the other half were given a security screening type of examination with general relevant questions. The examiners accurately classified 90% of the innocent participants and 75% of the guilty participants. The authors concluded that the examiner decisions were not affected by the specificity of the questions or the time lag.

Several methodological issues call into question the validity of Barland et al.'s (1989) results. As the authors pointed out, there was a concern that the mock crime in the first experiment was not sufficiently relevant to the participants. The participants were officers loyal to the U.S. government, and comparison questions about overall honesty and integrity may have been more threatening than the programmed espionage. Likewise, the lack of explicit rewards or punishments may have reduced the motivation levels of the examinees. As most deception research suggests that motivation plays an important role (Cook et al., 2012; DePaulo et al., 2003; Elaad & Ben-Shakhar, 1989; Gustafson & Orne, 1963, 1965; Kircher et al., 1988), the lack of motivation coupled with the weak mock crime could have caused the high number of false negatives. In the second experiment, examiners were instructed to conduct the three examinations as separate examinations. However, a random sample of the single-issue examinations revealed that the time between examinations was only 68 s longer than the time between the charts within an examination. As 68 s is not sufficient time to introduce two new relevant and three new comparison questions, it is possible that some of the examiners reviewed all of the

relevant questions prior to the examinations. If that were the case, then examinees may have been thinking about all six relevant questions, even though only two relevant questions were asked on each test (U.S. Department of Defense Polygraph Institute, 1995a). In the third experiment, many participants from the first experiment were retested, so the results were confounded (NRC, 2003). As Barland et al. (1989) stated, "strong generalization of the results of these studies to the field is not possible."

An additional criticism leveled against Barland et al. (1989) concerned the wording of the relevant questions (Barland et al., 1989; U.S. Department of Defense Polygraph Institute, 1995a). Relevant questions asked examinees if they had committed an act "against the United States," but Barland et al. pointed out that because guilty examinees participated in a mock crime and did not commit any act against the United States, the wording was inappropriate. To test the effects of the phrase "against the United States" and to test a new screening format, the U.S. Department of Defense conducted a mock crime study where guilty examinees enacted one of four scenarios: 1) espionage, 2) sabotage, 3) unauthorized contact, or 4) unauthorized disclosure (U.S. Department of Defense Polygraph Institute, 1995a). Relevant questions were broadly worded, and the caveat "during this project" was added to each relevant question. For example, one question read, "Have you committed an act of espionage during this project?" Examiners accurately classified 88.9% of the innocent examinees, excluding the 3.6% inconclusive decisions, and 83.3% of the guilty examinees. Although questions were broadly worded to resemble those of a screening examination, unlike a true screening examination, examinees could be guilty of only one crime. An additional study was conducted with the same protocols that correctly classified 98% of the innocent

examinees, excluding the 1.9% inconclusive decisions, and 83.3% of the guilty examinees (U.S. Department of Defense Polygraph Institute, 1995b).

Honts and Amato (2007) developed an automated version of a commonly used polygraph screening test known as the Relevant-Irrelevant Test. A Relevant-Irrelevant test includes only relevant and irrelevant questions. Inferences about deception are based on comparisons of reactions among different relevant questions and between relevant and irrelevant questions. The Honts and Amato study is important because it shares automation and question format with the ODT (see Kircher, 2018).

Participants were assigned to an innocent condition or a guilty condition. Guilty participants were assigned to falsify two pieces of information on an employment application form such as name, birth date, address, etc. Half of the participants received their polygraph examination from a trained examiner (human condition), and the other half listened to a tape recording of the test questions (automated condition). The results showed that the automated condition was significantly more accurate than the human condition. In the human condition, 66.7% of guilty participants and 62.2% of innocent participants were classified correctly. In the automated condition, 78.9% of the guilty participants and 76.2% of the innocent participants were correctly classified. The authors concluded that generalizability of the results was limited as guilty participants lied to only one relevant item, but the results suggested that an automated procedure could be a more emotionally engaging procedure. If so, it might provide a more effective method for deception detection as compared to the traditional polygraph.

In addition to dichotomous classifications of deceptive and truthful, several of these studies also tested if the polygraph could accurately identify the items to which a

guilty individual was deceptive. In the Correa and Adams (1981) study, the examiner correctly identified 82.7% of the individual lie responses. In Experiment 1 of Barland et al. (1989), examiners correctly classified 24% of the relevant questions answered deceptively. In Experiment 2, 33% of the outcomes on specific individual crimes were correct. Although neither of the U.S. DOD studies evaluated specific-issue accuracy, the Department of Defense Polygraph Institute Research Division Staff (Department of Defense Polygraph Institute Research Division Staff, 1997) subsequently conducted a mock crime study similar to the U.S. DOD studies. They analyzed guilty participants' reactions to the relevant issues on a polygraph examination and discovered that 59% of the guilty examinees responded most strongly to the question specific to the crime they committed. Honts and Amato (2007) found that guilty participants exhibited the strongest reactions to deceptively answered questions 79% of the time.

One field study compared various methods for scoring polygraph recordings from multiple-issue screening examinations. Raskin and Kircher (2014) used computer algorithms to analyze two law enforcement screening tests for 58 prospective employees of a federal law enforcement agency (a total of 116 examinations). Excluding inconclusives, 90% of the deceptive examinations and 82.4% of the truthful examinations were correctly classified. The authors did not report accuracy at the level of individual relevant issues.

The literature supports the notion that the polygraph can classify individuals as deceptive or truthful on multiple-issue screening examinations with approximately 85% accuracy, but accuracy drops to approximately 65% when attempts are made to identify the specific issue(s) to which the person was deceptive (American Polygraph Association,

2011; Correa & Adams, 1981; Department of Defense Polygraph Institute Research Division Staff, 1997; Honts & Amato, 2007). Probability theory may account for this large decline in accuracy.

Probability theory says that if the issues are independent and the probability of deception on any given issue is .50, a correct classification of all relevant issues has a chance accuracy equal to $1/2^n$, where *n* is the number of issues to be tested. When there is a single issue, chance accuracy would be is $1/2^1 = .50$. A multiple-issue test with three issues has a chance accuracy of $1/2^3 = .125$. To achieve an accuracy of .85, a credibility test that classifies the person as deceptive or truthful on one issue requires an improvement of 35 percentage points, but a three-issue test requires an improvement of 72.5 percentage points.

In addition to low chance rates of accuracy, multiple-issue screening tests also may suffer from a low prior probability of guilt (NRC, 2003). Prior probability of guilt refers to the relative frequency of guilt in the population being tested. To illustrate, Krapohl (2002) proposed the following thought experiment. He supposed that the validity of a screening examination was 80% in the discrimination between truthful and deceptive examinees. He also assumed that 10 of 1,000 people applying for positions in the government were spies. If 1,000 applicants were tested, with an accuracy of 80%, 792 applicants would clear the process. Of the 10 spies, eight would be caught, but these spies would be among the 198 non-spies who also failed the examination. If all the applicants had been classified as non-spies, the error rate would have been 1%. With the screening examination, the error rate would increase to 20%, and 99% of those errors would be false positives.

Critics argue that this high error rate should disqualify credibility assessment tools from use within the federal government (NRC, 2003). In contrast, Martin and Terris (1991) argued that the false-positive argument in personnel selection is problematic because it assumes that all applicants would have been hired were it not for the screening test. However, most organizations have more applicants than vacant positions. When the number of positions available is less than the number of applicants, then any tool that discriminates between truthfulness and deception better than chance reduces the number of deceptive individuals in the applicant pool and increases the odds that a truthful applicant will be hired (Krapohl, 2002). A valid predictor, by definition, produces fewer false-positive and false-negative decision errors than a predictor that has little or no validity. Therefore, despite the base rate problem, valid screening tools provide the best chance to make a correct hiring decision when there are more applicants than available positions.

The ODT

In 2012, Cook et al. introduced a new psychophysiological screening tool called the ocular-motor deception test (ODT). A computer presents written instructions and true/false test statements concerning the examinee's possible involvement in two illicit activities, also called relevant issues. In a laboratory experiment, the relevant issues might be stealing $20 from a purse or stealing a credit card from an unattended backpack. In a field setting, the two relevant issues might be recent illicit drug use or espionage. The examinee uses the keyboard or mouse keys to answer True or False while a remote eye tracker and computer records eye movements, blink rate, response time, answer accuracy,

and changes in pupil size. The computer compares ocular-motor responses to questions concerning the two relevant issues, combines its measurements in a logistic regression, and classifies the individual as truthful or deceptive on the test. Accuracy rates in lab and field studies range from 78% to 86% (Kircher, 2018).

Although the ODT and the polygraph are both psychophysiological tests for deception, there are several differences between them. A polygraph test begins with a pretest interview. In this interview, the examiner obtains basic biographical information from the examinee, discusses the purpose of the test, explains why deception causes physiological reactions, and reviews the test questions with the examinees. The ODT has no pretest interview.

There are several types of polygraph examinations. The most common type of test for criminal investigation is known as the probable lie test. On a probable lie test, there are three main types of test questions: relevant questions, probable-lie questions, and neutral questions. Relevant questions ask the examinee about their involvement in illicit activities. Probable-lie questions are designed to encourage the examinee to answer with a denial (Raskin & Honts, 2002), and they are presented to the examinee in a manner that would make the examinee believe that "admissions would negatively influence the examiner's opinion and that strong reactions to these questions would produce a deceptive result" (Raskin, 1986). An example of a probable-lie question may be, "Before the age of 18, did you ever take something that did not belong to you?" Neutral questions provide a rest between questions. Although an ODT has relevant and neutral questions, it does not contain probable-lie questions.

A sequence of questions in a polygraph test will contain approximately 10

questions, two to four of which are relevant questions. The sequence of questions is repeated three times, though it can be repeated up to five times if three repetitions provides insufficient data to draw a conclusion. Test questions are presented at a rate of one question every 25 to 30 s. During the test, examinees' electrodermal, cardiovascular, and respiration responses are recorded. Participants are expected to react most strongly to the question that poses the greatest threat to their appearing truthful on the test (Podlesny & Raskin, 1977). Deception or uncertainty about their deception to the probable-lie questions causes innocent participants to view probable-lie questions as the greatest threat, whereas guilty participants are expected to view the relevant questions as the greatest threat (Honts et al., 2002). A polygraph takes several hours to conduct, and additional time is required for the examiner to score the results.

A sequence of questions on the ODT contains 48 questions: 32 relevant questions, 16 for each of the two relevant issues, and 16 neutral questions. Questions are presented at a rate of 2 – 4 s, and the examinee's pupillary responses and reading behaviors are recorded. Deceptive examinees are expected to react to the relevant issue that poses the greatest threat, and truthful examinees should be equally concerned over both sets of relevant statements (Kircher, 2018). Therefore, only deceptive examinees are expected to display differential reactions. A test takes approximately 30 min, and results are available within five min.

In spite of these differences, the ODT and the polygraph share similar theoretical frameworks. Originally, the ODT was based on the cognitive-workload hypothesis (Cook et al., 2012). More recently, Bovard et al. (2019) stated that the four-factor theory of deception introduced by Zuckerman, DePaulo, and Rosenthal (1981) may better explain

effects on ocular-motor and other psychophysiological measures. The four-factor theory posits that differential reactions observed in deceptive participants on the polygraph or the ODT are a result of changes in arousal, emotion, control, or cognitive load.

According to Arnold, (as cited in Kircher, 1981), arousal is manifested by changes in visceral activity brought about by the autonomic nervous system. Raskin (1979) suggested that for polygraph tests, questions vary in arousal value or salience. A salient stimulus is significant to the individual because it is novel, surprising, familiar, complicated, pertinent, or otherwise important. Consistent with Raskin (1979), Handler and Nelson (2007) and others (Senter et al., 2010) have proposed that different types of questions possess different degrees of salience for truthful and deceptive people, and differential salience accounts for diagnostic effects on physiological measures.

Autonomic responses also reflect changes in the intensity of emotional reactions to stimuli (Zuckerman et al., 1981). Generally, physiological arousal increases in response to positively and negatively valenced stimuli (Bradley, Miccoli, Escrig, & Lang, 2008). Relevant questions answered deceptively should pose the greatest threat to a deceptive examinee. When the consequences of failing the test are serious, relevant questions may cause the person to experience fear and evoke strong physiological reactions (Kircher, 2018).

To avoid detection, a deceiver attempts to control their behavior, and these attempts may serve as cues to deception (Zuckerman et al., 1981). Cook et al. (2012) reported within-subject differences between relevant statements answered truthfully and deceptively; deception was associated with fewer fixations and shorter reading and rereading time. The authors found that deceptive participants attempted to read the

incriminating statements quickly and suppress rereading behaviors. This finding is consistent with the view that participants can exert some conscious control over their reading behaviors to implement specific reading strategies (Hyönä & Nurminen, 2006).

In addition to increased levels of arousal and emotion, deception results in increased cognitive load (Johnson et al., 2005; Vrij et al., 2006). In an ODT examination, examinees are informed that if they do not respond quickly and accurately, they will fail the test. As a deceptive individual takes the test, they are under time pressure to differentiate between questions that require either truthful or deceptive responses. Bovard et al. (2019) suggested that truthful answers are prepotent, and deceptive examinees must inhibit truthful answers to answer deceptively. A deceptive examinee must also respond in a timely and consistent manner to avoid detection. The additional mental effort required to (1) differentiate between questions answered truthfully and deceptively, (2) inhibit truthful answers, and (3) monitor the consistency of responses to different types of test questions could account for some effects of deception on pupil dilation (Bradley & Janisse, 1979; Janisse & Bradley, 1980; Lubow & Fein, 1996; Webb et al., 2009), response time (Seymour et al., 2000; Spence et al., 2001; Vendemia et al., 2005; Walczyk et al., 2003, 2009), and reading behaviors (Cook et al., 2012).

The magnitude of change in cognitive load a deceptive person experiences, and consequently the magnitude of observed responses, may depend on individual differences such as intelligence and working memory (Carroll, 1993; Jonassen et al., 1993; Plass, Kalyuga, & Leutner, 2010). The Cattell-Horn-Carroll Theory of Intelligence posits that intelligence has two components: fluid intelligence (Gf) and crystallized intelligence (Gc). Fluid intelligence is the ability to adapt to new situations and solve novel problems

(Cattell, 1963). Crystallized intelligence (Gc) is considered to be language and general world knowledge. Reading comprehension, reading speed, and vocabulary knowledge reflect levels of crystallized intelligence (Baghaei & Tabatabaee, 2015; Carroll, 1993; Chin et al., 2015). Because the ODT incorporates reading behaviors to determine deception (Hacker et al., 2014), the present study assessed the effects of Gc on deceptive participants' behavioral and ocular-motor measures.

Crystallized intelligence has been shown to correlate with pupil dilations. Ahern and Beatty (1979) used SAT (Scholastic Aptitude Test) scores, a crystallized measure, to determine the intelligence levels of the participants. They showed that high intelligence individuals displayed smaller pupil dilations than low intelligence individuals when performing mental calculations. More recently, Lee and colleagues (2015) investigated whether pupil dilations can distinguish between high and low intelligence. The authors used the Korean intelligence test to measure both crystallized and fluid intelligence and administered a series of linguistic, mathematical, and visuo-spatial tasks of varying difficulties while recording pupil levels. The results showed that difficult linguistic and visuo-spatial tasks evoked larger pupil dilations from intelligent individuals than less intelligent individuals. However, consistent with Ahern and Beatty, the authors also found that highly intelligent individuals showed smaller task–evoked pupil dilations than less intelligent individuals on mathematical tasks, regardless of difficulty.

In addition to intelligence, cognitive load is associated with an individual's working memory (Paas, Renkl, & Sweller, 2003; Sweller, 1988). Working memory (WM) is the cognitive capacity to hold information in the memory while processing incoming information (Baddeley, 2000). Multiple studies have shown that WM is

correlated with crystallized intelligence (Cantor et al., 1991; García-Madruga et al., 2013; Swanson, 2011; Tillman et al., 2009) and especially with reading comprehension (García-Madruga et al., 2013; Just & Carpenter, 1992; Seigneuric et al., 2000). Thus, measures of Gc should, to some degree, capture variance associated with WM.

In general, task-evoked cognitive load is smaller for individuals with high Gc and high WM compared to individuals with low Gc and low WM (Kelley & McLaughlin, 2012; Paas, Tuovinen, Tabbers, & Gerven, 2003). Because pupil dilations and reading behaviors reflect cognitive load, deceptive individuals with high Gc and high WM may display smaller pupil dilations and faster response times than deceptive individuals with low Gc and low WM. To evaluate this prediction, participants completed a vocabulary test after the ODT. Vocabulary is commonly used to measure crystallized intelligence and correlates between .3 and .4 with WM (Daneman & Merikle, 1996; Woltz, 1990). Although previous research has shown that WM does not affect the diagnostic validity of the ODT (Patnaik, 2015), the present study explored the possibility that vocabulary affects ODT outcomes.

Present Study

The most common protocol for the ODT is the Relevant Comparison Test (RCT). The RCT tests whether the examinee is deceptive to either of two relevant issues, or the target issues of concern (Krapohl et al., 2012), by comparing reactions to the two issues. Examinees who show little or no difference in reactions to the two sets of relevant statements are classified as truthful to both issues. If the examinee reacts more strongly to statements concerning one of the two relevant issues, the ODT classifies the person as

deceptive about that relevant issue.

The present study developed a new, multiple-issue screening protocol and tested whether it can accurately classify an examinee as truthful or deceptive and accurately identify the issues that elicited a deceptive response. The new protocol expanded the RCT by asking the examinee about their possible involvement in four crimes. Three of those crimes were the relevant issues, while the fourth crime was the comparison issue. Comparison issues provide a baseline of nondeceptive behavior. Reactions to the relevant issues were compared to the comparison issue to assess the relative magnitude of the person's reactions to the relevant issues. If reactions to the relevant issue were significantly greater than reactions to the comparison issue, the subject was classified as deceptive on that particular issue. The three relevant issues included the theft of $20, the theft of a VISA gift card, and the vandalism of a parking kiosk. The comparison issue was filing a false police report. The study also tested if Gc, as measured by a vocabulary test, affected deceptive participants' behavioral and ocular-motor measures.

Research questions and aims

1. With a multiple-issue screening format, do ocular-motor and behavioral measures discriminate between truthful and deceptive participants?
2. With a multiple-issue screening format, do deceptive participants react specifically to the issue(s) that required a deceptive response?
3. Are behavioral and ocular-motor measure differences between items answered deceptively and items answered truthfully greater for guilty participants with lower levels of Gc than for guilty participants with high levels of Gc?

METHODS

Design and Analysis

The design was a 3 x (4 x 5) mixed design with one between-group factor and two within-subject factors. The between-group factor was guilt with three levels (innocent, guilty of one crime, guilty of two crimes). The two within-subject factors were statement type (cash, card, vandalism, false report) and repetition. A set of 48 items was presented 5 times. The test included 12 cash, 12 card, 12 vandalism, and 12 false report statements. The correct (nonincriminating) answer was True for six of the 12 statements that address an issue and False for the remaining statements that address that issue. The test statements are presented in Appendix A.

Participants

Two hundred and two participants were recruited from the general community and the University of Utah. Of the 202 participants, 11 opted out of the study once they discovered they were assigned to a mock crime condition. An additional 11 were dropped from the data analyses due to insufficient data or a failure to follow instructions. The final analyses included 23 participants from the university and 157 participants from the general community for a total sample size of 180. The mean age for participants from the university was 24.73, and the mean age of participants from the community was 28.98. Sixty participants were assigned to the innocent condition (*innocent*), 60 participants

were assigned to steal $20 (*cash*), and 60 participants were assigned to steal $20 and the VISA gift card (*cash+card*). Table 1 provides the demographic information for each of the conditions.

Advertisements were posted on KSL.com and Craigslist, and flyers were posted on the University of Utah campus. The ads promised individuals $40 in pay and the chance to earn a $40 bonus if they participated in a psychology experiment at the University of Utah. Respondents who spoke English as their first language, were over the age of 18, did not wear glasses, had not participated in a previous lie detection experiment, and were not currently taking psychotropic medications were allowed to participate.

Apparatus

A Tobii 4C remote eye tracker affixed to a 15.6-inch Asus flat screen monitor recorded eye movements and pupil diameter at 60 Hz. Viewing was binocular, but analyses used data only from the right eye. Although the eye tracker allowed for free head movement, a chin rest was used to keep the participant's head still. EyeDetect by Converus© version 3.27 presented stimuli to the participant and collected and edited the ocular-motor data. The 60 Hz PD data were imported into CPSLAB 11 (Scientific Assessment Technologies, Inc, Salt Lake City, UT), a general-purpose computer program for psychophysiological research. Stimuli were presented to the participant on the computer monitor positioned 60 cm from the participant's eyes.

The 18-item vocabulary test presented in Appendix B was used to assess the participant's Gc and WM.

Ocular-Motor Deception Test

The ODT consisted of 48 test statements, and these same 48 statements were presented five times in random order, subject to the constraint that a statement of particular type was never followed by a statement of the same type. All participants had the same randomized order. The statements appeared one at a time, halfway up the computer screen on the left side. To the right of the statements, the words "True" and "False" appeared in separate boxes. Instructions and test items appeared in black characters on a gray background to minimize fluctuations in illumination that might have affected the pupil. Participants answered by pressing the left arrow key or the right arrow key on the keyboard. The selected answer's background briefly changed to a lighter gray color. There was a 600 ms pause after an examinee's response until the next item appeared. Between the five sessions of test statements, examinees performed 10 simple arithmetic tasks. The purpose of the intervening task was to clear working memory of ODT test statements (Hacker et al., 2014).

Procedures

Participants reported alone to a room in a building on campus. Instructions in an envelope addressed to the participant and taped to the door instructed the participant to enter the room and read and sign the consent form. The participant then listened over headphones to a recording that provided the instructions for the study. A hard copy of the recorded instructions was included as well. A phone number was provided for participants to call if they did not wish to participate.

All participants were told that some participants were guilty of committing one or

more of four mock crimes: vandalize a parking kiosk on campus, take $20 from a secretary's backpack, make a false report to the campus police, or steal a VISA gift card from an office; whereas other participants were innocent of all four crimes. Guilty participants were assigned to one of two guilty conditions. Guilty participants in the first condition were told were told to take $20 from a secretary's backpack (*cash*). They were instructed to go to a secretary's office on the same floor of the same building and ask the secretary where Dr. Mitchell's office was located. The secretary informed the participant that there was no Dr. Mitchell in the building, and the participant left. The participant waited inconspicuously for the secretary to leave the office unattended, then entered the office, found the backpack with a wallet in it, removed $20 from the wallet, and concealed the money on their person. Participants were told to have an alibi in case they were caught and not to leave fingerprints. They were informed that they had no more than 20 min to commit the crime and report to the experimenter, who was in another building on campus. In the second guilty condition, the procedures remained the same, except in addition to the $20, participants were told to steal a VISA gift card that was in the same wallet as the $20 (*cash+card*).

Innocent participants did not commit any of the crimes. They were told that some participants had to commit one or more of the crimes, but they were assigned to the *innocent* group and should not commit any of the crimes. *Innocent* participants were instructed to wait approximately 20 min before reporting to the experimenter.

Although all participants were informed that some participants vandalized a parking kiosk or made a false police report, no one committed those crimes.

Participants reported to the experimenter after committing their crime(s) or after

an appropriate waiting period. After the experimenter calibrated the Tobii eye tracker, the participant was tested about the four crimes.

After the completion of the ODT, participants were administered a vocabulary test. Once the participant finished the vocabulary test, she or he was informed of the decision, paid, and debriefed.

Discriminating Variables

To maximize the reliability of each outcome measure, measurements of responses were averaged across items of a given type (cash, card, vandalism, false report) and across repetitions. There was a mean for the cash items, a mean for the card items, a mean for the vandalism items, and a mean for the police report items. For each outcome measure, I defined the false report issue as the comparison issue and then computed the difference between each remaining relevant item and the comparison issue. The difference from the comparison issue served as the discriminating variable. *Innocent* participants were expected to exhibit small differences for all relevant issues. Guilty participants were expected to display large differences for relevant items answered deceptively.

Because participants did not commit any acts of vandalism or file a false police report, vandalism could have been the comparison question, and filing a false police report could have been a relevant issue. Analyses indicated that vandalism as a comparison issue would have yielded comparable results to those obtained with using filing a false report as the comparison issue. Thus, filing a false report was arbitrarily chosen as the comparison issue.

Behavioral Outcome Measures

For each person, the mean of 12 items was obtained for each of the four relevant issues for each of the five sessions.

Response Time (RT)

RT was the time in s from the appearance of the item on the screen to a key press response from the subject.

Proportion Wrong

Proportion wrong was the number of incorrect responses divided by the number of items.

Ocular-Motor Outcome Measures

An area of interest (AOI) was defined for each T/F test item. The AOI began with the first character of the item and ended at the period at the end of the statement. Ocular-motor reading measures were computed for the fixations in each AOI. Fixations were determined from the data files produced by the Tobii eye tracker by identifying a sequence of samples where the eye showed little movement for at least 100 ms.

Number of Fixations

Number of fixations was the number of fixations detected in the AOI.

First Pass Duration

First pass duration was the sum of all fixation durations in the AOI before the eye fixated outside the AOI.

Reread Duration

Reread duration was the sum of all fixation durations that followed leftward eye movements within the AOI. This measure assessed rereading done by the participant whether or not the eye fixated outside the AOI.

PD Waveform

PD waveform was the pupil response curve in mm from statement onset for a period of 4 s. To establish a common baseline of 0 for each response curve, the initial level at statement onset was subtracted from each subsequent level for the 4-s interval.

PD Peak Amplitude

PD peak amplitude was obtained from a pupil response curve. The response curve began the moment the test statement appeared on the computer screen and ended four s later. The computer identified high and low points in the response curve and computed the difference between each low point and every subsequent high point. PD peak amplitude was the greatest observed difference.

PD Area Under the Pupil Response Curve

PD AUC was the area under the pupil response curve from response onset to the point at which the response returned to the initial level or to the end of the 4-s sampling interval, whichever occurred first. Response onset was defined at the low point in the response curve from which peak amplitude was measured.

Pupil Diameter (PD) Level at Response

The computer calculated the mean and standard deviation of PD samples from the onset of the first test statement to the participant's answer to the 48th test statement. The mean and standard deviation were used to convert the raw PD samples to standard scores. PD level was the mean of standard scores beginning 500 ms before the answer and ending 1500 ms after the answer.

Item Blink Rate and Next Item Blink Rate

Blink rate was the number of blinks per minute. As the average blink duration is 258 ms (Davson, 1990), missing data gaps that were greater than 170 ms and less than 350 ms were classified as blinks (Draper Laboratory, 2008). Blink rate was computed for each item (item blink rate) with an extraction interval that began 2550 ms before the response and ended when the participant answered. Blinks were also extracted for the item that followed (next item blink rate) with an extraction interval that began when the participant answered and ended 3150 ms later. Extraction intervals were based on participants' mean response time (2550 ms).

Table 1. Participant demographics by condition.

		Innocent	Cash	Cash+Card
Age				
	n	60	60	60
	Mean	29.75	28.45	27.12
	Min–Max	18–69	18–76	18–50
	Median	26	24.5	25
Gender				
	Female	26	29	26
	Male	34	31	34
Vocabulary				
	n	43	52	52
	Mean	11.4	11.38	11.38
	Min - Max	5–17	4–17	3–17
	Median	11	11	11

RESULTS

Repeated measures analysis of variance (RMANOVA) was used to analyze each behavioral and ocular-motor measure. Only the Guilt X Statement type interaction was of interest and will be discussed in the Results section. Appendix C contains the complete set of statistically significant main effects and interactions for each outcome measure. Significance for tests that involved a repeating factor used Huynh-Feldt corrections for degrees of freedom.

Table 2 shows means and standard deviations for outcome measures broken down by Guilt and Statement Type.

RMANOVA provided an omnibus test of the Guilt X Statement Type interaction to determine if the three groups reacted differently to the four types of statements. To test if the groups reacted in the manner predicted by the rationale that underlies the multiple-issue ODT, three planned comparisons were conducted when the omnibus test was significant. The first comparison tested for the expected difference between guilty and *innocent* groups. The remaining two comparisons tested for predicted differences between the two guilty groups.

To facilitate these comparisons, three within-subject contrasts were computed. The $\psi_{Cash-False\ report}$ contrast was the difference between cash and false report scores. *Innocent* participants were expected to show no significant difference between cash and false report items, whereas both groups of guilty participants were expected to react more

strongly to cash than false report statements. The $\psi_{Cash-Card}$ contrast was the difference between cash and card scores, and the $\psi_{Card-False\ report}$ contrast was the difference between card and false report scores. *Cash* participants were expected to show a significant $\psi_{Cash-Card}$ difference but not a $\psi_{Card-False\ report}$ difference. Conversely, *cash+card* participants were expected to show a significant $\psi_{Card-False\ report}$ difference but no $\psi_{Cash-Card}$ difference. Table 3 shows the results of the contrasts.

Due to missing data, one participant from the *cash+card* condition was dropped from the following analyses: number of fixations, first pass duration, reread duration.

Preliminary RMANOVAs were conducted to test for moderating effects of sessions. Only pupil level showed a significant Guilt X Statement type X Session interaction, $F(23.213, 2054.381) = 1.855$, $p < .01$, $\eta_p^2 = .021$. Because the effect size was inconsequential and the interaction was session was limited to one outcome measure, all subsequent analyses were conducted with means based on all sessions.

The Guilt X Statement type interaction was not significant for the following variables: proportion wrong, item blink rate, next item blink rate.

Behavioral and Reading Measures

Figure 1 presents the means for response time. Error bars represent 95% confidence intervals calculated for within-subject comparisons of statement type. The Guilt X Statement type interaction was significant, $F(4.336, 377.273) = 19.445$, $p < .001$, $\eta_p^2 = .183$. Results of the planned comparisons revealed that the difference between cash and false report statements was 0.34 s greater for guilty participants than for *innocent* participants. The difference between cash statements and card statements was 0.244 s

greater for *cash* participants than for *cash+card* participants. Finally, the difference between card and false report statements was 0.193 s greater for *cash+card* participants than for *cash* participants.

Figure 2 shows the results for number of fixations. The Guilt X Statement type interaction was significant, $F(4.631, 401.506) = 22.057$, $p < .001$, $\eta_p^2 = .202$. All three planned comparisons were significant. The difference between cash and false report statements was 0.974 fixations greater for guilty than for *innocent* participants. On average, *cash* participants showed a 0.730 greater difference between cash statements and card statements than did *cash+card* participants. *Cash+card* participants showed a 0.664 greater difference between card statements and false report statements than did *cash* participants.

Figure 3 presents the results for first pass duration. The Guilt X Statement type interaction was significant, $F(4.794, 414.668) = 3.041$, $\eta_p^2 = .034$. Planned contrasts revealed that guilty participants showed a 79 ms greater difference between cash and false report statements than did *innocent* participants. The remaining two contrasts were not significant.

Figure 4 shows the mean reread duration times. The Guilt X Statement type interaction was significant, $F(5.014, 433.729) = 10.951$, $p < .001$, $\eta_p^2 = .112$. Planned contrasts revealed that difference between cash items and false report items was greater for guilty than for *innocent* participants; the difference between cash statements and card statements was greater for *cash* than *cash+card* participants; and the difference between card statements and false report statements was greater for *cash+card* than *cash* participants.

Pupil Measures

Figures 5a, 5b, and 5c show the mean change in pupil diameter (PD) from statement onset for *innocent*, *cash*, and *cash+card* participants over a period of four s. The first observation occurred at statement onset and served as a baseline. It was subtracted from each subsequent observation in the response curve. A positive value indicated that pupil diameter increased relative to statement onset, and a negative value indicated that it decreased.

As predicted, *innocent* participants showed similar changes in pupil size to all four types of statements. In contrast, *cash* participants reacted more strongly to cash than to other statements, and *cash+card* participants reacted more strongly to cash and card statements than to vandalism and false report statements.

Figure 6 presents the results for the peak amplitude of the pupil response. The Guilt X Statement type interaction was significant, $F(4.938, 429.58) = 16.219$, $p < .001$, $\eta_p^2 = .157$. The first contrast indicated that the difference between cash statements and false report statements for guilty participants was 0.024 mm greater than the difference for *innocent* participants. The second contrast showed that the difference between cash statements and card statements for *cash* participants exceeded the difference for *cash+card* participants. The third contrast revealed that the difference between card statements and false report statements was greater for *cash+card* than for *cash* participants.

Figure 7 shows the results for area under the pupil response curves. The Guilt X Statement type interaction was significant, $F(4.318, 375.627) = 23.183$, $p < .001$, $\eta_p^2 = .209$. Consistent with predictions, the three planned contrasts were significant. The

difference between cash and false report statements was greater for guilty participants than for *innocent* participants. In addition, the difference between cash and card statements was greater for *cash* than for *cash+card* participants, and the difference between card and false report statements was greater for *cash+card* than *cash* participants.

Figure 8 presents within-session standardized pupil levels when the participant answered the test statement. The Guilt X Statement type interaction was significant, $F(5.087, 442.573) = 43.286$, $p < .001$, $\eta_p^2 = .332$. The difference between cash statements and false report statements for guilty participants exceeded the difference for *innocent* participants by 0.272 mm. The difference between cash statements and card statements was greater for *cash* participants than for *cash+card* participants; and the difference between card statements and false report statements for *cash+card* participants was greater than for *cash* participants.

Age and Vocabulary

Effects of Age and Vocabulary were of secondary interest. To test for effects of Age and Vocabulary, for each outcome measure, the difference between cash and false report statements for guilty participants was regressed onto Age and Vocabulary. Because a test proctor did not realize that the vocabulary test had two sides, data were missing for 16 of the 120 guilty participants. Those without a completed test were dropped from the regression analyses.

In the regression analyses, first pass duration was significantly related to Vocabulary, $p < .01$, $\beta = .283$. For a one-unit increase in vocabulary score, the difference

between cash and false report statements decreased by 0.021 s.

Valence

Innocent participants unexpectedly showed significant differences among statement types in response times. Because valence has been shown to affect reading behaviors (Citron et al., 2014; Megalakaki et al., 2019), I analyzed the valence values of the statements. I collected the valence values of each statement by using the NRC Valence, Arousal, and Dominance Lexicon (Mohammad, 2018). The lexicon rates words on a continuous scale from 0 to 1, where higher scores reflect more positively-valenced stimuli. For each of the 48 statements, I used the lexicon to determine the valence values of the content words. I summed the valence values and divided by the number of content words in the item to calculate the mean valence value per word. Words not found in the lexicon were excluded from the analyses. Results of the valence calculations are shown in Appendix D. I then ran ANOVA contrasts to determine if the valence differences among relevant issues were significant. The results are shown in Table 4. Of the six comparisons, the difference between cash statements and card statements was nonsignificant, and the difference between vandalism statements and false report statements approached significance. All other contrasts were significant and consistent with the idea that statements with larger valence values also had faster response times. To assess the impact of the valence values on number of fixations and response times, for each *innocent* participant, I regressed the valence values for the 48 test statements on mean number of fixations and mean response times for each of the 48 statements. The mean coefficients among the 60 participants were significantly different from zero,

$b_{FIX} = -1.179, t(59) - 3.257, p < .01;\ b_{RT} = -0.714, t(59) = -5.769,\ p < .001.$ For a one-unit increase in valence, the number of fixations decreased by 1.179, and response times decreased by 0.714 s.

Discriminating Variables

For each outcome measure, repeated measurements were averaged across items of a given type and across repetitions, yielding a mean for cash items, a mean for card items, a mean for vandalism items, and a mean for false report items for each participant. The person means for the four statement types were used to compute two discriminating variables. One variable was the difference between the mean for cash items and the mean for false report items. The second variable was the difference between the means for card and false report items.

Point-biserial correlations were calculated to assess the diagnostic validity of these derived outcome measures. The difference between cash and false report items was correlated with a dichotomous variable that distinguished between *innocent* (coded 0) and all guilty participants (coded 1). The difference between card and false report items was correlated with a dichotomous variable where *innocent* and *cash* participants were coded as 0 (truthful) and *cash+card* participants were coded as 1 (deceptive).

To assess the reliability of these measures, responses were averaged within statement types and within repetitions. This resulted in one mean for each of the relevant issues for each of the five repetitions. The difference between the cash and the false report items was computed for each participant. Coefficient alpha was then computed to assess the agreement among difference scores across repetitions. The same procedure was

used to assess the reliability of differences between card and false report statements. Table 5 reports the point-biserial correlations and Cronbach's alpha for each measure.

The negative correlations for first pass duration, reread duration, response time, and number of fixations indicate that guilty participants were faster to respond, made fewer fixations, and spent less time reading and rereading when they were deceptive than when they were truthful. The positive correlations for PD area, PD level, and PD peak amplitude indicate that guilty participants showed greater pupil dilations to statements answered deceptively than statements answered truthfully.

Of the eight significant outcome measures, response time, number of fixations, PD peak amplitude, PD area, and PD level were most diagnostic of deception and were therefore selected for further analyses. For each of these variables, a Cohen's *d* effect size was calculated between cash and false report statements, card and false report statements, and vandalism and false report statements for each participant. Cohen's *d* used a pooled standard deviation based on the 60 presentations of each statement type for the individual.

A 4-folds cross-validation was performed to assess the accuracy of a logistic regression equation. Participants were randomly assigned such that each fold contained 45 participants: 15 *innocent*, 15 *cash*, and 15 *cash+card*. A logistic regression model was trained on three folds and validated on the fourth. Each fold was used three times in a training model and once as the validation fold. For each of the four partitions, all possible combinations of the five aforementioned variables were analyzed in R (R Core Team, 2017) with the package leaps (Lumley, 2020) to determine the best model in terms of accuracy and parsimony. The analyses indicated that number of fixations and PD level

best predicted guilt for all partitions.

For each validation, accuracy is reported for the dichotomous classification (truthful or deceptive) and for the three relevant issues. Participants were classified as innocent if deception was not indicated to any of the three relevant issues. Participants were classified as guilty if they showed a deceptive response to at least one of the three relevant issues. The 4-fold validation results are shown in Table 6. A summary table is provided in Table 7.

The mean dichotomous classification accuracy was 85.5%: 83.3% for *innocent* participants and 86.6% for guilty participants, with 91.7% of *cash* participants and 81.4% of *cash+card* participants correctly identified. The mean specific-issue accuracy for cash, card, and vandalism statements was 87.6%: 92.8% for *innocent* participants, 91.1% for *cash* participants, and 78.9% for *cash+card* participants. ROC curves were generated for cash classifications and card classifications. The area under the ROC curve was .92 for cash classifications and .851 for card classifications.

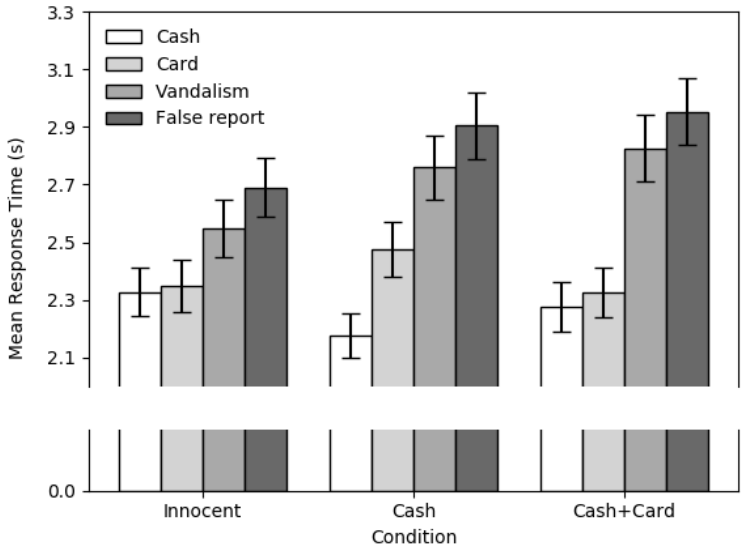


Figure 1. Response times by statement type and guilt condition.

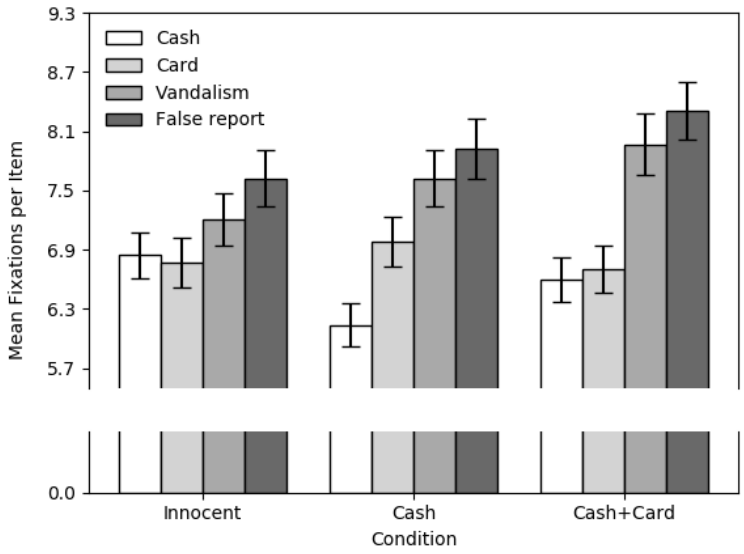


Figure 2. Number of fixations by statement type and guilt condition.

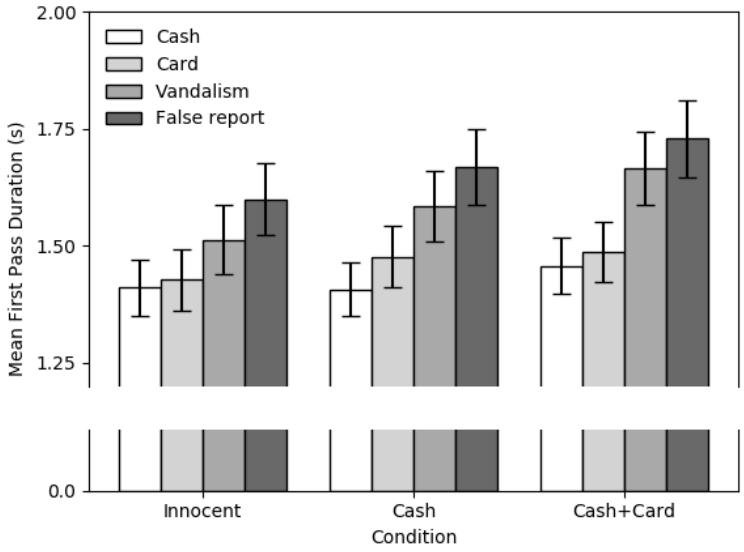


Figure 3. First pass duration times by statement type and guilt condition.

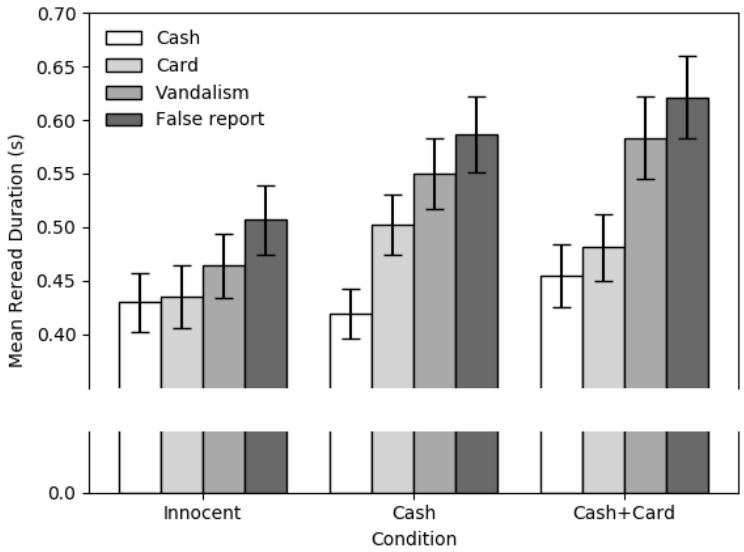
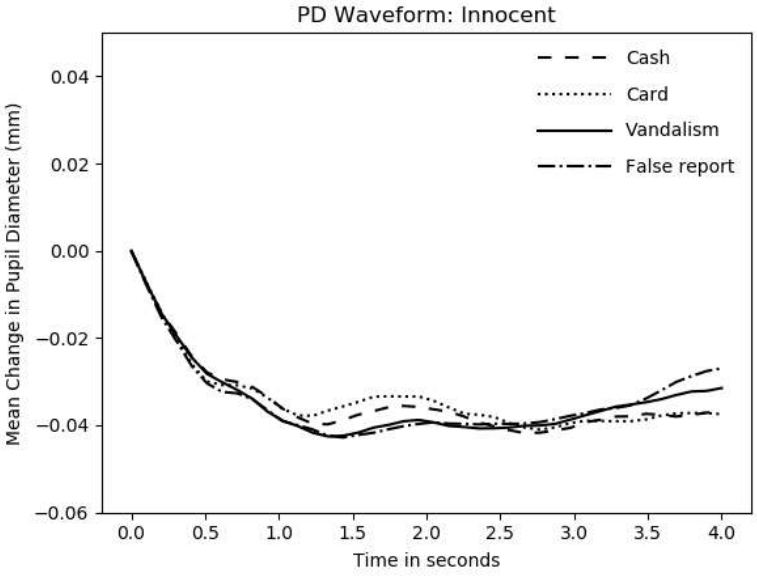


Figure 4. Reread duration times by statement type and guilt condition.

a)



b)

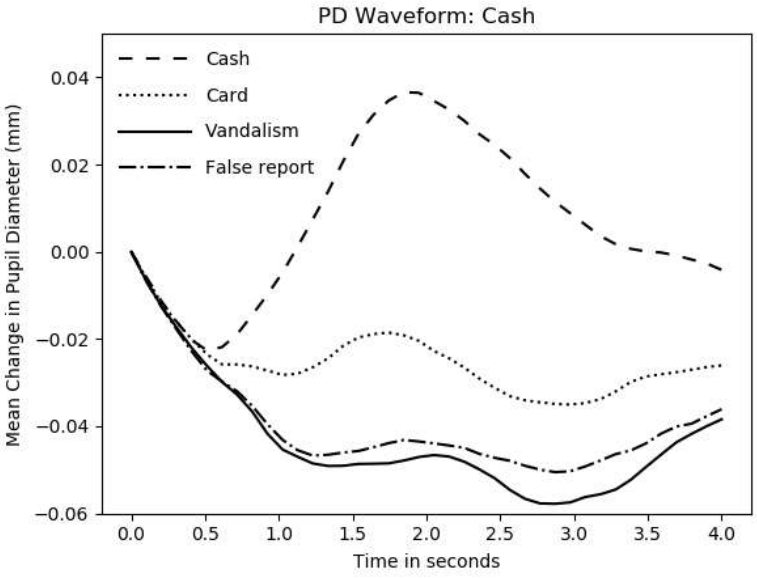


Figure 5. Pupil response by statement type, a) for *innocent* participants, b) for *cash* participants, c) for *cash+card* participants.

c)

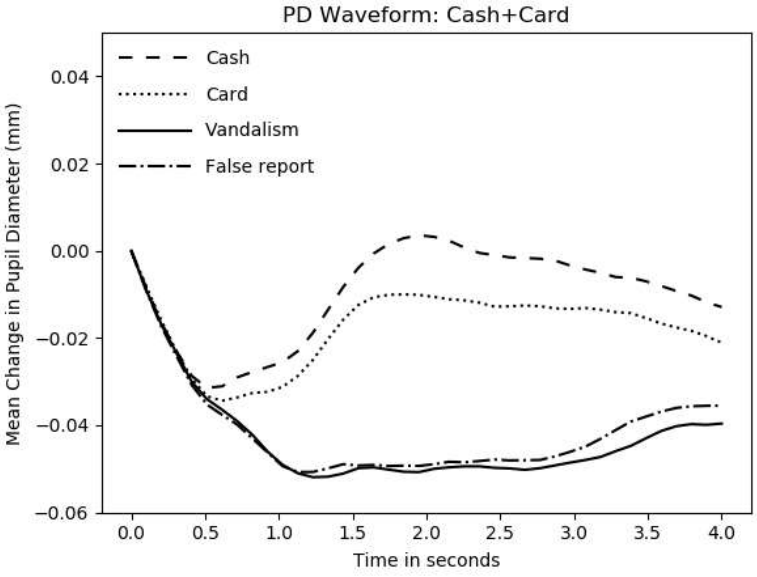


Figure 5. Continued

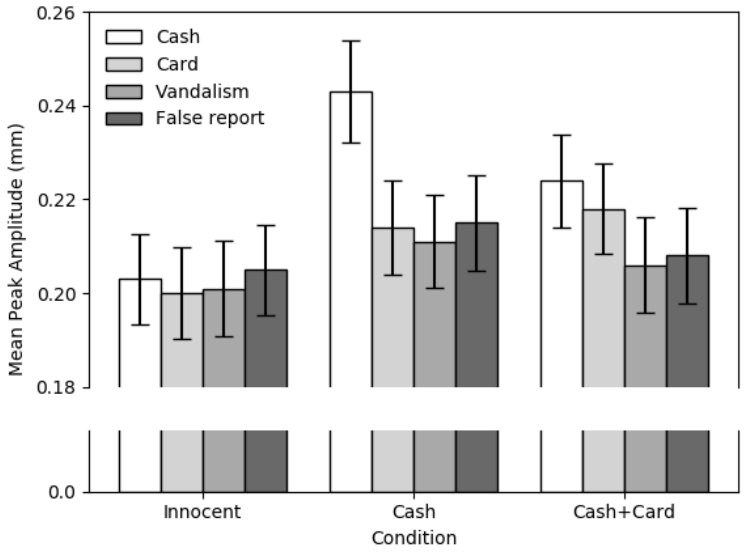


Figure 6. PD peak amplitude by statement type and guilt condition.

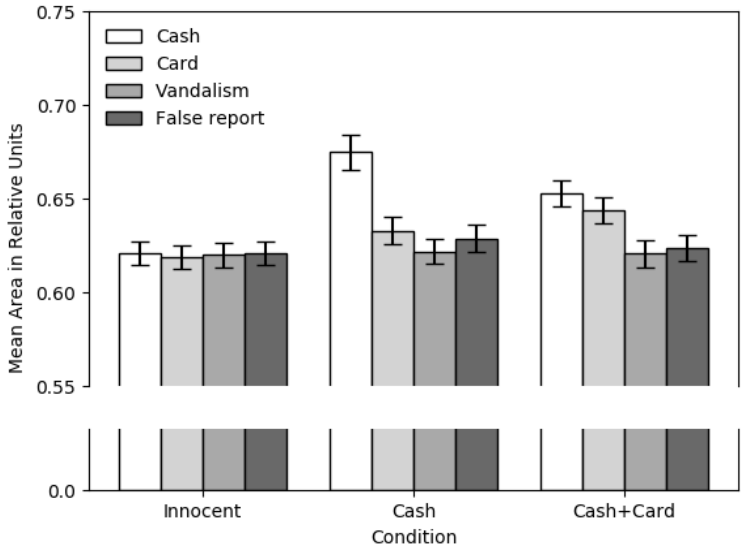


Figure 7. PD area under the curve by statement type and guilt condition.

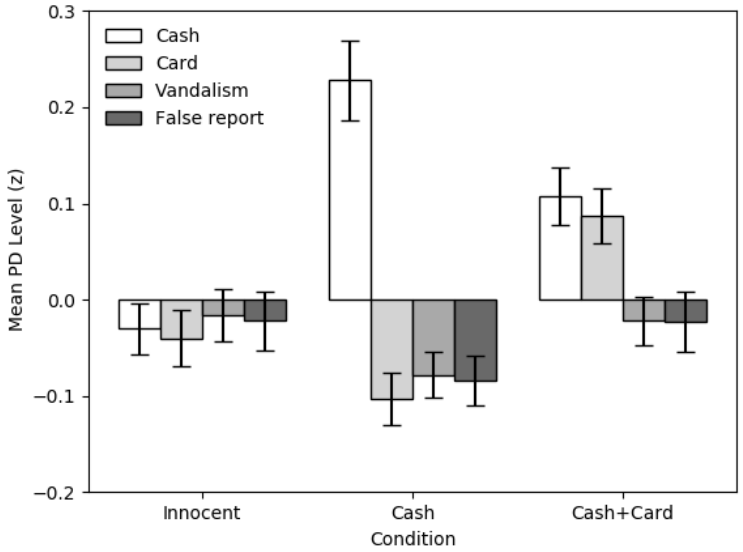


Figure 8. PD level by statement type and guilt condition.

Table 2. Means and standard deviations for dependent variables by statement type for all participants.

		Cash		Card		Vandalism		Police Report	
Dependent Variable	Condition	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Response Time	Innocent	2.327	.567	2.348	.623	2.548	.686	2.689	.700
	Cash	2.177	.525	2.473	.651	2.759	.752	2.905	.795
	Cash+Card	2.275	.588	2.326	.599	2.826	.793	2.952	.789
Proportion Wrong	Innocent	.022	.029	.028	.037	.033	.034	.037	.037
	Cash	.030	.041	.036	.044	.046	.060	.038	.039
	Cash+Card	.030	.061	.034	.066	.049	.071	.051	.074
Number of Fixations	Innocent	6.847	1.582	6.772	1.732	7.204	1.835	7.621	1.960
	Cash	6.138	1.484	6.979	1.707	7.622	1.937	7.919	2.074
	Cash+Card	6.594	1.541	6.705	1.599	7.968	2.139	8.309	2.019
First Pass Duration	Innocent	1.411	0.413	1.427	0.448	1.513	0.505	1.599	0.526
	Cash	1.406	0.392	1.477	0.454	1.584	0.518	1.669	0.557
	Cash+Card	1.457	0.414	1.487	0.432	1.665	0.538	1.729	0.561
Reread Duration	Innocent	.430	.187	.435	.198	.464	.207	.507	.221
	Cash	.419	.158	.502	.194	.550	.226	.587	.242
	Cash+Card	.455	.200	.481	.213	.583	.263	.621	.264
PD Waveform	Innocent	-.032	.046	-.031	.042	-.031	.046	-.033	.040
	Cash	.001	.039	-.027	.036	-.046	.038	-.043	.034
	Cash+Card	-.006	.039	-.014	.036	-.040	.038	-.037	.034
PD Peak Amplitude	Innocent	.203	.066	.200	.067	.201	.069	.205	.066
	Cash	.243	.075	.214	.068	.211	.068	.215	.070
	Cash+Card	.224	.068	.218	.066	.206	.070	.208	.070
PD Area	Innocent	.621	.043	.619	.044	.620	.047	.621	.043
	Cash	.675	.064	.633	.049	.622	.047	.629	.049
	Cash+Card	.653	.048	.644	.047	.621	.050	.624	.048
PD Level	Innocent	-.030	.180	-.040	.197	-.016	.188	-.022	.210
	Cash	.228	.286	-.103	.184	-.078	.164	-.084	.177
	Cash+Card	.108	.206	.087	.198	-.022	.170	-.023	.216
Item Blink Rate	Innocent	3.385	2.379	3.062	2.175	3.393	2.209	3.017	2.417
	Cash	3.032	1.619	3.133	1.718	3.230	1.837	2.927	1.782
	Cash+Card	2.963	1.671	2.795	1.576	3.242	1.789	2.858	1.829
Next Item Blink Rate	Innocent	3.708	2.554	3.032	2.304	3.093	2.016	3.370	2.361
	Cash	3.565	1.991	3.077	1.703	2.975	1.497	3.003	1.419
	Cash+Card	3.453	2.086	2.967	1.767	2.888	1.471	2.852	1.731

Table 3. Planned contrasts for outcome measures.

Measure	Contrast	Value	t	p	η_p^2
Response Time	$\psi_{Cash-False\ report}$	-0.340	-6.574	< .001	.196
	$\psi_{Cash-Card}$	0.244	6.790	< .001	.207
	$\psi_{Card-False\ report}$	-0.193	-3.911	< .001	.080
Number of Fixations	$\psi_{Cash-False\ report}$	-0.974	-6.918	< .001	.214
	$\psi_{Cash-Card}$	0.730	6.541	< .001	.196
	$\psi_{Card-False\ report}$	-0.664	-5.377	< .001	.141
First Pass Duration	$\psi_{Cash-False\ report}$	-0.079	-2.235	< .05	.028
	$\psi_{Cash-Card}$	0.041	1.541	ns	-
	$\psi_{Card-False\ report}$	-0.050	-1.557	ns	-
Reread Duration	$\psi_{Cash-False\ report}$	-0.090	-4.744	< .001	.113
	$\psi_{Cash-Card}$	0.058	3.766	< .001	.075
	$\psi_{Card-False\ report}$	-0.056	-3.220	< .01	.056
PD Peak Amplitude	$\psi_{Cash-False\ report}$	0.024	5.975	< .001	.168
	$\psi_{Cash-Card}$	-0.023	-5.657	< .001	.153
	$\psi_{Card-False\ report}$	0.011	3.194	< .01	.054
PD AUC	$\psi_{Cash-False\ report}$	0.038	6.837	< .001	.209
	$\psi_{Cash-Card}$	-0.034	-6.132	< .001	.175
	$\psi_{Card-False\ report}$	0.015	3.316	< .01	.065
PD Level	$\psi_{Cash-False\ report}$	0.272	8.492	< .001	.289
	$\psi_{Cash-Card}$	-0.350	-10.485	< .001	.383
	$\psi_{Card-False\ report}$	0.155	5.436	< .001	.143

Table 4. ANOVA contrasts for the valence values of the test items.

Contrast	p	Value
Cash – False report	< .001	0.165
Cash – Vandalism	< .01	0.109
Cash – Card	ns	-
Card – False report	< .001	0.170
Card – Vandalism	< .01	0.114
Vandalism – False report	.079	0.056

Table 5. Validity and reliability coefficients

	Cash minus False Report		Card minus False Report	
Outcome Measure	Correlation	Reliability	Correlation	Reliability
Response Time	-.572***	.655	-.436***	.522
Proportion Wrong	.004	.306	-.156*	.000
Number of fixations	-.587***	.636	-.491***	.473
First Pass Duration	-.178*	.570	-.096	.233
Reread Duration	-.360***	.547	-.298***	.357
PD Peak Amplitude	.430***	.667	.352***	.495
PD AUC	.469***	.728	.401***	.477
PD Level	.510***	.808	.446***	.623
Item Blink Rate	-.045	.000	.009	.000
Next Item Blink Rate	.041	.193	.016	.059

*$p < .05$
**$p < .01$
***$p < .001$

Table 6. Results from the 4-folds validation.

Training/Validation Folds	Group Membership	Predicted Group Membership			Specific Issue (% Correct)			
		Innocent	Guilty	% Correct	Cash	Card	Vandalism	Mean % Correct
1, 2, 3/4	Innocent	13	2	86.7	13 (86.7)	14 (93.3)	15 (100)	93.3
	Cash	3	12	80	12 (80)	13 (86.7)	14 (93.3)	86.7
	Cash+Card	3	12	80	11 (73.3)	10 (66.7)	15 (100)	80
	Mean % Correct			82.2	80	82.2	97.8	86.7
1, 2, 4/3*	Innocent	14	1	93.3	15 (100)	14 (93.3)	15 (100)	97.8
	Cash	2	13	86.7	13 (86.7)	13 (86.7)	15 (100)	91.1
	Cash+Card	2	12	85.7	11 (78.6)	11 (78.6)	13 (92.9)	77.8
	Mean % Correct			88.6	86.7	86.4	97.7	88.9
1, 3, 4/2	Innocent	12	3	80	13 (86.7)	14 (93.3)	15 (100)	93.3
	Cash	0	15	100	15 (100)	13 (86.7)	15 (100)	95.6
	Cash+Card	6	9	60	8 (53.3)	8 (53.3)	15 (100)	68.9
	Mean % Correct			80	80	77.8	100	85.9
2, 3, 4/1	Innocent	11	4	73.3	13 (86.7)	13 (86.7)	13 (86.7)	86.7
	Cash	0	15	100	15 (100)	11 (73.3)	14 (93.3)	88.9
	Cash+Card	0	15	100	15 (100)	12 (80)	13 (86.7)	88.9
	Mean % Correct			91.1	95.6	80	88.9	88.1

*Validation Fold 3 included the *cash+card* participant without fixation data. Therefore, the participant was dropped from validation analyses, and *cash+card* accuracy was based on 14 participants.

Table 7. Mean percent correct from 4-folds validation.

		Specific Issue			
	Guilt Classification	Cash	Card	Vandalism	Mean Accuracy
Innocent	83.3	90.0	91.7	96.7	92.8
Cash	91.7	90.1	83.3	96.7	90.0
Cash+Card	81.4	75.0	68.3	93.3	78.9
Mean Accuracy	85.5	86.1	81.1	95.6	87.6

DISCUSSION

The primary goal of the present study was to determine if a multiple-issue screening format could discriminate between truthful and deceptive participants and accurately identify which of the three crimes a deceptive participant committed. In addition, this study assessed the effects of age and crystallized intelligence on the accuracy of the multiple-issue ODT.

Multiple-Issue Responses

The four-factor theory argues that deception causes changes in emotion, arousal, control, and cognitive load (Zuckerman et al., 1981). These changes result in physiological differences that enable the detection of deception. Based on this theory, it was hypothesized that a multiple-issue screening protocol would distinguish between innocent and deceptive participants and permit identification of the particular relevant statements to which the person was deceptive.

To test these hypotheses, data for 10 behavioral and ocular-motor measures were analyzed with RMANOVA. The Guilt X Statement interaction indicated whether ocular-motor and behavioral reactions to various statement types differed as a function of guilt. The Guilt X Statement type interaction was significant for 8 of the 10 measures. For those measures with a significant omnibus interaction, three contrasts were calculated: $\psi_{Cash-False\ report}$, $\psi_{Cash-Card}$, and $\psi_{Card-False\ report}$. Tests of the $\psi_{Cash-False\ report}$

contrast revealed that 8 of 10 outcome measures discriminated between guilty and *innocent* groups with a mean partial η^2 effect size of 0.16. These findings are consistent with those obtained in laboratory and field studies of the RCT (Kircher, 2018). Generally, deception was associated with decreases in response time, number of fixations, and reading times, and increases in pupil size.

Tests of the $\psi_{Cash-Card}$ contrast revealed that 7 of 10 outcome measures discriminated between guilty participants who committed one or two of the mock crimes. Generally, there was a greater difference between statements about the cash and card for participants who lied about taking the cash than for those who lied about both the cash and the card. The mean effect size for the difference between guilty groups on the $\psi_{Cash-Card}$ contrast was 0.16, which suggests that the differences between the two guilty groups were as great as the differences between guilty and *innocent* groups.

The final contrast ($\psi_{Card-False\ report}$) tested for differences between guilty groups that did or did not take the gift card. Only participants who took the cash and the card were expected to react differentially to card and false report statements. Seven of eight contrasts were significant, and all seven showed greater differences between card and false report statements for participants who lied about the gift card than for participants who were truthful. The mean effect size of .077 suggests that differences between relevant and comparison issues decreased for *cash+card* participants compared to *cash* participants.

The multiple-issue screening protocol introduced in the present study accurately discriminated between innocent and guilty examinees, and it identified which issue(s) elicited a deceptive response with greater than 85% accuracy. These results stand in

contrast to prior PDD research. Although some research suggests that the polygraph may be used to identify the specific issue that elicits a deceptive response (Correa & Adams, 1981; Honts & Amato, 2007), the general consensus is that the rate of error is too great to employ this practice (Nelson, 2015). On multiple-issue polygraph tests, deceptive examinees may react to questions about crimes they did not commit in addition to, or instead of, the crime they committed (Handler et al., 2009). The high rate of error may be due to the low number of questions about each specific issue on the polygraph test (Correa & Adams, 1981). For example, the Utah Comparison Question Test may contain four relevant issues per chart and three charts per test, though up to five charts may be used (Raskin & Kircher, 2014). With five charts, reactions to 20 relevant questions are used to decide if the person was deceptive on the test. However, if an attempt is made to identify the specific issue(s) that caused the deceptive outcome, decisions may be based on reactions to a maximum of five repetitions of each relevant question. In contrast, the ODT contains 12 relevant questions per issue within a single session, and the test consists of five sessions. As compared to the polygraph, the ODT has 10 to 20 times the number of questions per relevant issue. This increase in relevant questions increases the reliability of the results (Raskin & Kircher, 2014) and allows the ODT to classify examinees as truthful or deceptive to various issues with greater accuracy than the polygraph.

When deriving the algorithm to calculate accuracy, two methods were used to measure within-subject differences between reactions to relevant issues. Both approaches used the mean of reactions across the 12 statements for each relevant issue across the five repetitions (n = 60). The first method used the simple difference between the mean for cash, card, or vandalism statements and the control issue, false police report. The second

method divided the simple difference by the pooled within-statement standard deviation (Cohen's *d*). Cohen's *d* was uniformly more diagnostic than the simple difference score across all outcome measures. Use of Cohen's *d* provided a common scale for the measures included in the logistic regression equation and increased its accuracy, relative to using the raw differences.

As compared to participants who committed one crime, participants who committed two crimes displayed attenuated pupillary responses when they were deceptive. Neuroscientists have shown that the amygdala acts as a gateway for emotion processing, identification, and evaluation of the affective value of a stimulus (Sladky et al., 2012). Likewise, emotional tasks with cognitive demand involve the anterior cingulate cortex (ACC, Phan et al., 2002). The amygdala and ACC have been shown to activate during deception (Abe, 2011; Abe et al., 2006), arguably because deception elicits emotional reactions and increases cognitive load (Zuckerman et al., 1981). These two neural structures have also been shown to habituate to repeated presentations of emotional stimuli (Phan et al., 2003; Wright et al., 2001). Participants who committed two crimes lied twice as often on the test as those who committed one crime. Consequently, the relatively weak pupillary responses obtained from participants who lied to two relevant issues may be a result of the amygdala and the ACC habituating to the stimuli.

The amygdala and the ACC modulate pupillary responses through the locus coeruleus (LC). Stimuli that cause changes in cognitive load, attention, arousal, or emotion activate the LC (Aston-Jones et al., 1999; Berridge, 2008; Bouret & Sara, 2005; Rajkowski et al., 2004; Samuels & Szabadi, 2008; Tully & Bolshakov, 2010), and pupil

dilations have been shown to be an indirect measure of LC activation (Aston-Jones & Cohen, 2005; Eldar et al., 2013; Laeng et al., 2012; Samuels & Szabadi, 2008). The amygdala and ACC both project to the LC (Aston-Jones & Cohen, 2005; Bouret et al., 2003). Therefore, as the amygdala and the ACC habituate to the repeated questions, LC activation is reduced, and the pupillary responses decrease.

In addition, the attenuated pupillary responses may be a result of decreased cognitive load. Cash statements and card statements were similar to each other in that they both asked about theft. *Cash+card* participants knew that whenever the question pertained to theft, regardless of what was stolen, they had to lie. *Cash* participants lied only to statements that asked about the theft of $20. The pupil dilations observed in *cash* participants suggest that they had to focus more carefully than did cash+card participants on the content of the statements to determine which statement required a deceptive response. The pupil dilations to cash statements might indicate that *cash* subjects had to pay more attention to content of the statement. On the contrary, because *cash+card* participants did not need to differentiate between cash statements and card statements, their cognitive load was less than that of *cash* participants, and the reduced cognitive load resulted in reduced pupillary responses.

Age and Intelligence

There was no evidence that age affected the predictive validity of any ocular-motor or behavioral outcome measure. Likewise, there was no evidence that intelligence moderated the predictive validity of the remaining reading or pupil outcome measures. As such, within the ranges of age and intelligence represented in the present study, it does

not appear that these individual differences affect the diagnostic accuracy of the multiple-issue ODT.

Interestingly, guilty participants with greater Gc showed less differences between cash and false report statements on first pass duration than did participants with lower Gc. This finding suggests that the initial processing of the text was easier for participants with high levels of intelligence than those with moderate levels of intelligence (Liversedge et al., 1998). However, there were no apparent effects of intelligence on reread behavior or pupil responses. These results suggest that, once guilty participants realized that the statement required a deceptive response, the subsequent processing load was similar across all levels of intelligence.

Innocent Participants' Response Times

The response time data indicated that *innocent* participants responded faster to cash statements and card statements compared to vandalism statements and false report statements. *Innocent* participants were expected to show similar reading behaviors across the four issues since they did not commit a crime, and previous studies on the ODT have not observed a significant differences in reading behaviors among innocent participants (Cook et al., 2012; Patnaik et al., 2016). The lack of differential pupil levels between the relevant issues suggested that the items did not significantly differ as a function of cognitive load or arousal for *innocent* participants. Instead, the response time differences might be due to the wording of the stimuli. Studies have shown that positively-valenced words elicit faster responses than negatively-valenced words (Scott et al., 2012, 2014). To evaluate if this hypothesis could account for differences among statement types

observed in the present experiment, valence values of the statements were averaged together by statement type and then compared with ANOVA contrasts. The results showed that statements with higher valence values resulted in faster response times than items with low valence values. These results suggest that future studies should control for the valence values of the stimuli.

Limitations

The present study was a laboratory study. Though attempts were made to replicate real-life situations and instill a sense of jeopardy, the consequences of failing the ODT were trivial compared to failure in a real-life scenario. Although no substantive differences between results obtained in laboratory and field settings have been reported for the RCT (Kircher, 2018), efforts should be made to replicate the present findings with representative samples of confirmed field cases.

As described previously, screening examinations typically include general questions that cover a broad period of time and lack well-defined issues. The questions in the present study referred to behaviors that occurred the same day and were well-defined. These differences may limit the generalizability of the present findings to real-life screening situations.

To increase generalizability of the results, participants were recruited from the general community and university campus. Only 12% of participants were university students, and most were Caucasian, had at least a high school education, and were under the age of 30. Although generalizability to the general population may be limited, previous studies have demonstrated the ODT's ability to generalize to different

populations and cultures (Kircher, 2018).

Summary

The multiple-issue format developed in this study has greater practical utility than the RCT format. Whereas the RCT format allows for a classification of truth or deception on only one topic, the multiple-issue format allows for decisions on multiple topics. The multiple issue format includes the same number of test items as the RCT and is as diagnostic of deception as the RCT, but it covers more relevant issues. Supplemental analyses yielded no evidence that either age or intelligence affected the diagnostic validity of any behavioral or ocular-motor measure.

APPENDIX A

TEST ITEMS

R1 – Theft of $20

1. I am guilty of taking money from the secretary's wallet.
2. I am the one who took the secretary's money from the backpack.
3. I took a $20 bill from a backpack in a secretary's office.
4. The $20 in the secretary's backpack was taken by me.
5. The reason the $20 is gone is because I took it.
6. The secretary's money was stolen by me.
7. I was uninvolved in the theft of the $20 from the secretary.
8. I am innocent of taking the $20 from the secretary's wallet.
9. I had nothing to do with the theft of the secretary's $20.
10. The money in the secretary's backpack was untouched by me.
11. The theft of the money from a backpack was not my fault.
12. The cash in the secretary's wallet was not stolen by me.

R2 – Theft of Credit Card

1. I took the VISA gift card from the office.
2. I am responsible for the theft of the gift card from the office.
3. I am guilty of stealing the VISA gift card from the office.
4. The VISA gift card is missing from the office because of me.
5. The VISA gift card is missing because I took it.
6. The VISA gift card was taken from the office by me.
7. I was uninvolved in the recent theft of the VISA gift card.
8. I am completely innocent of taking the VISA gift card from the office.
9. I did not steal a VISA gift card from an office today.
10. The stolen VISA gift card is not in my possession.
11. It is not my fault that the VISA gift card is missing.
12. The VISA gift card was not taken by me.

R3 – Vandalism

1. I am to blame for vandalizing the parking kiosk.
2. I intentionally damaged the parking kiosk on campus.
3. I am not innocent of vandalizing the campus parking kiosk.
4. The damage to the parking kiosk on campus was my fault.
5. The recent damage to the parking kiosk was my fault.
6. The campus parking kiosk was vandalized by me.
7. I was uninvolved in vandalizing the parking kiosk on campus.
8. I am innocent of damaging the campus parking kiosk.
9. I did not engage in vandalizing the campus parking kiosk.
10. The university's parking kiosk was undamaged by me.
11. Vandalizing the parking kiosk on campus was done by someone else.
12. The parking kiosk at the university was not damaged by me.

R4 – Police Report

1. I am responsible for the false police report.
2. I called campus police to falsely report a crime.
3. I am not innocent of making a false police report.
4. The person who reported the fake crime on campus was me.
5. A false report to the police was made by me.
6. The campus police received a false crime report from me.
7. I was not responsible for the false report to the campus police.
8. I am innocent of reporting a fake crime to the campus police.
9. I had no part in falsely reporting a crime on campus.
10. The false report received by campus police was not from me.
11. The incorrect report of a campus crime was not my doing.
12. The false report of a campus crime was by someone else.

APPENDIX B

VOCABULARY TEST

EVOKE

A. wake up
B. surrender
C. reconnoiter
D. transcend
E. call forth

Answer: _____

PLACATE

A. rehabilitate
B. plagiarize
C. depredate
D. apprise
E. conciliate

Answer: _____

APATHETIC

A. wandering
B. impassive
C. hateful
D. prophetic
E. overflowing

Answer: _____

EFFULGENCE

A. prominence
B. outline
C. change
D. radiance
E. energy

Answer: _____

RANCOROUS

A. malignant
B. jubilant
C. abashed
D. inglorious
E. careless

Answer: _____

VACILLATION

A. purification
B. wavering
C. expulsion
D. tempting
E. foolishness

Answer: _____

COLLUSION

A. nerve
B. rest
C. prayer
D. conspiracy
E. disguise

Answer: _____

FEIGN

A. pretend
B. prefer
C. wear
D. be cautious
E. surrender

Answer: _____

SUCCINCT

A. sudden
B. concise
C. prosperous
D. literary
E. cunning

Answer: _____

ALIGNMENT

A. formation
B. accusation
C. emblem
D. brightness
E. buoyant

Answer: _____

MASTICATE

A. chew
B. massage
C. manufacture
D. create
E. pollute

Answer: _____

GRITTY

A. frigid
B. windy
C. adhesive
D. granular
E. unwieldy

Answer: _____

ARIDITY

A. bitterness
B. surface
C. sonority
D. dryness
E. torridity

Answer: _____

DUPLICITY

A. extent
B. double-dealing
C. agreement
D. cleverness
E. overlapping

Answer: _____

DURESS

A. period of time
B. distaste
C. courage
D. hardness
E. compulsion

Answer: _____

CAPRICIOUSNESS

A. stubbornness
B. courage
C. whimsicality
D. amazement
E. greediness

Answer: _____

CORROBORATORY

A. plausible
B. anticipatory
C. confirmatory
D. explanatory
E. esoteric

Answer: _____

FIGURINE

A. metaphor
B. wine
C. poem
D. organ
E. statuette

Answer: _____

APPENDIX C

SIGNIFICANT EFFECTS FOR DEPENDENT VARIABLES

Table 8. Effect sizes for dependent variables.

Measure	Source	p	η_p^2
Response Time	Guilt		
	Statement type	< .001	0.673
	Sex	< .01	0.057
	Guilt X Statement type	< .001	0.182
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex		
Proportion Wrong	Guilt		
	Statement type	< .001	0.08
	Sex	< .05	0.024
	Guilt X Statement type		
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex	< .05	0.027
Number of Fixations	Guilt		
	Statement type	< .001	0.604
	Sex		
	Guilt X Statement type	< .001	0.202
	Guilt X Sex		
	Statement type X Sex	< .05	0.018
	Guilt X Statement type X Sex		
First Pass Duration	Guilt		
	Statement type	< .001	0.436
	Sex	< .001	0.101
	Guilt X Statement type	< .05	0.032
	Guilt X Sex		
	Statement type X Sex	< .001	0.053
	Guilt X Statement type X Sex		
Reread Duration	Guilt		
	Statement type	< .001	0.441
	Sex	< .05	0.034
	Guilt X Statement type	< .01	0.112
	Guilt X Sex		
	Statement type X Sex	< .05	0.018
	Guilt X Statement type X Sex		
PD	Guilt		
	Statement type	< .001	0.37
	Sex		
	Time	< .001	0.101
	Guilt x Statement type	< .001	0.262
	Guilt x Sex		
	Guilt X Time	< .05	0.025

Table 8 (continued)

Measure	Source	p	η_p^2
	Statement type X Sex		
	Statement type X Time	< .001	0.223
	Time X Sex	< .01	0.028
	Guilt X Statement type X Sex		
	Guilt X Statement type X Time	< .001	0.15
	Guilt X Time X Sex		
	Statement type X Time X Sex		
PD Peak Amplitude	Guilt		
	Statement type	< .001	0.204
	Sex	< .001	0.097
	Guilt X Statement type	< .001	0.157
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex		
PD Area Under the Curve	Guilt		
	Statement Type	< .001	0.273
	Sex	< .001	0.101
	Guilt X Statement Type	< .001	0.209
	Guilt X Sex		
	Guilt X Statement Type X Sex		
PD Level at Response	Guilt		
	Statement type	< .001	0.299
	Sex		
	Guilt X Statement type	< .001	0.332
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex	< .05	0.027
Item Blink Rate	Guilt		
	Statement type		
	Sex		
	Guilt X Statement type		
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex		
Next Item Blink Rate	Guilt		
	Statement type	< .001	0.192
	Sex		
	Guilt X Statement type		
	Guilt X Sex		
	Statement type X Sex		
	Guilt X Statement type X Sex		

APPENDIX D

VALENCE CALCULATIONS

Table 9. Valence values for content words by issue.

Issue	Statement	Content Words	Valence
Cash	1	guilty	0.135
		taking	0.828
		money	0.844
		secretary	0.612
		wallet	0.680
	2	one	0.542
		took	0.828
		secretary	0.612
		money	0.844
		backpack	0.728
	3	took	0.828
		twenty	0.520
		dollar	0.755
		bill	0.396
		backpack	0.728
		secretary	0.612
		office	0.427
	4	twenty	0.520
		dollar	0.755
		secretary	0.612
		backpack	0.728
		taken	0.828
	5	reason	0.781
		twenty	0.520
		dollar	0.755
		gone	0.292
		took	0.828
	6	secretary	0.612
		money	0.844
		stolen	0.156
	7	uninvolved	N/A
		theft	0.135
		twenty	0.520
		dollar	0.755
		secretary	0.612
	8	innocent	0.729
		taking	0.828
		twenty	0.520

Table 9 (continued)

Issue	Statement	Content Words	Valence
		dollar	0.755
		taking	0.828
		secretary	0.612
		wallet	0.680
	9	nothing	N/A
		theft	0.135
		secretary	0.612
		twenty	0.520
		dollar	0.755
	10	money	0.844
		secretary	0.612
		backpack	0.728
		untouched	0.531
	11	theft	0.135
		money	0.844
		backpack	0.728
		fault	0.188
	12	cash	0.867
		secretary	0.612
		wallet	0.680
		stolen	0.156
Card	1	took	0.828
		Visa	0.735
		gift	0.880
		card	0.531
		office	0.427
	2	responsible	0.837
		theft	0.135
		gift	0.880
		card	0.531
		office	0.427
	3	guilty	0.135
		stealing	0.021
		VISA	0.735
		gift	0.880
		card	0.531
		office	0.427
	4	VISA	0.735
		gift	0.880

Table 9 (continued)

Issue	Statement	Content Words	Valence
		card	0.531
		missing	0.306
		office	0.427
	5	VISA	0.735
		gift	0.880
		card	0.531
		missing	0.306
		took	0.828
	6	VISA	0.735
		gift	0.880
		card	0.531
		taken	0.828
		office	0.427
	7	uninvolved	N/A
		recent	0.592
		theft	0.135
		VISA	0.735
		gift	0.880
		card	0.531
	8	completely	0.730
		innocent	0.729
		taking	0.828
		VISA	0.735
		gift	0.880
		card	0.531
		office	0.427
	9	steal	0.021
		VISA	0.735
		gift	0.880
		card	0.531
		office	0.427
		today	0.806
	10	stolen	0.156
		VISA	0.735
		gift	0.880
		card	0.531
		possession	0.677
	11	fault	0.188
		VISA	0.735

Table 9 (continued)

Issue	Statement	Content Words	Valence
		gift	0.880
		card	0.531
		missing	0.306
	12	VISA	0.735
		gift	0.880
		card	0.531
		taken	0.828
Vandalism	1	blame	0.115
		vandalizing	0.083
		parking	0.777
		kiosk	0.500
	2	intentionally	0.450
		damaged	0.059
		parking	0.777
		kiosk	0.500
		campus	0.604
	3	innocent	0.729
		vandalizing	0.083
		campus	0.604
		parking	0.777
		kiosk	0.500
	4	damage	0.059
		parking	0.777
		kiosk	0.500
		campus	0.604
		fault	0.188
	5	recent	0.592
		damage	0.059
		parking	0.777
		kiosk	0.500
		fault	0.188
	6	campus	0.604
		parking	0.777
		kiosk	0.500
		vandalized	0.083
	7	uninvolved	N/A
		vandalizing	0.083
		parking	0.777
		kiosk	0.500

Table 9 (continued)

Issue	Statement	Content Words	Valence
		campus	0.604
	8	innocent	0.729
		damaging	0.059
		campus	0.604
		parking	0.777
		kiosk	0.500
	9	engage	0.771
		vandalizing	0.083
		campus	0.604
		parking	0.777
		kiosk	0.500
	10	university	0.735
		parking	0.777
		kiosk	0.500
		undamaged	N/A
	11	vandalizing	0.083
		parking	0.777
		kiosk	0.500
		someone	N/A
		else	N/A
	12	parking	0.777
		kiosk	0.500
		university	0.735
		damaged	0.059
False report	1	responsible	0.837
		false	0.120
		police	0.458
		report	0.583
	2	called	0.344
		campus	0.604
		police	0.458
		falsely	0.140
		report	0.583
		crime	0.071
	3	innocent	0.729
		making	0.684
		false	0.120
		police	0.458
		report	0.583

Table 9 (continued)

Issue	Statement	Content Words	Valence
	4	person	0.646
		reported	0.583
		fake	0.073
		crime	0.071
		campus	0.604
	5	false	0.120
		report	0.583
		police	0.458
		made	0.684
	6	campus	0.604
		police	0.458
		received	0.646
		false	0.120
		crime	0.071
		report	0.583
	7	responsible	0.837
		false	0.120
		report	0.583
		campus	0.604
		police	0.458
	8	innocent	0.729
		reporting	0.583
		fake	0.073
		crime	0.071
		campus	0.604
		police	0.458
	9	part	0.663
		falsely	0.140
		reporting	0.583
		crime	0.071
		campus	0.604
	10	false	0.120
		report	0.583
		received	0.646
		campus	0.604
		police	0.458
	11	incorrect	0.140
		report	0.583
		received	0.646

Table 9 (continued)

Issue	Statement	Content Words	Valence
		campus	0.604
		crime	0.071
		doing	0.604
	12	false	0.120
		report	0.583
		campus	0.604
		crime	0.071
		someone	N/A
		else	N/A

Table 10. Results of valence calculations by issue and item.

Issue	Statement	# of Content Words	Valence Sum	Mean Valence per Word
Cash	1	5	3.099	0.620
	2	5	3.554	0.711
	3	7	4.266	0.609
	4	5	3.443	0.689
	5	5	3.176	0.635
	6	3	1.612	0.537
	7	4	2.022	0.506
	8	7	4.952	0.707
	9	4	2.022	0.506
	10	4	2.715	0.679
	11	4	1.895	0.474
	12	4	2.315	0.579
Card	1	5	3.401	0.680
	2	5	2.810	0.562
	3	6	2.729	0.455
	4	5	2.879	0.576
	5	5	3.280	0.656
	6	5	3.401	0.680
	7	5	2.873	0.575
	8	7	4.860	0.694
	9	6	3.400	0.567
	10	5	2.979	0.596
	11	5	2.640	0.528
	12	4	2.974	0.744
Vandalism	1	4	1.475	0.369
	2	5	2.390	0.478
	3	5	2.693	0.539
	4	5	2.128	0.426
	5	5	2.116	0.423
	6	4	1.964	0.491
	7	4	1.964	0.491
	8	5	2.669	0.534
	9	5	2.735	0.547
	10	3	2.012	0.671
	11	3	1.360	0.453
	12	4	2.071	0.518
False report	1	4	1.998	0.500

Table 10 (continued)

Issue	Statement	# of Content Words	Valence Sum	Mean Valence per Word
	2	6	2.200	0.367
	3	5	2.574	0.515
	4	5	1.977	0.395
	5	4	1.845	0.461
	6	6	2.482	0.414
	7	5	2.602	0.520
	8	6	2.518	0.420
	9	5	2.061	0.412
	10	5	2.411	0.482
	11	6	2.648	0.441
	12	4	1.378	0.345

Table 11. Summary table of valence results.

Issue	Total # of Content words	Total Valence Sum	Average Valence per word
Cash	57	35.071	0.604
Card	63	38.226	0.609
Vandalism	52	25.577	0.495
False report	61	26.694	0.439

REFERENCES

Abe, N. (2011). How the brain shapes deception: An integrated review of the literature. *The Neuroscientist, 17*(5), 560–574.

Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., & Fujii, T. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cerebral Cortex, 16*(2), 192–199.

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

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

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450.

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry, 46*(9), 1309–1320.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

Baghaei, P., & Tabatabaee, M. (2015). The C-test: An integrative measure of crystallized intelligence. *Journal of Intelligence, 3*, 46–58.

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

Berridge, C. W. (2008). Noradrenergic modulation of arousal. *Brain Research Reviews, 58*(1), 1–17.

Bouret, S., Duvel, A., Onat, S., & Sara, S. J. (2003). Phasic activation of locus coeruleus neurons by the central nucleus of the amygdala. *The Journal of Neuroscience, 23*(8), 3491–3497.

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, *28*(11), 574–582.

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.

Bradley, M. T., & Janisse, M. P. (1979). Pupil size and lie detection: The effect of certainty on detection. *Psychology: A Journal of Human Behavior*, *16*(4), 33–39.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*(2), 229–246.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22.

Chin, J., Payne, B., Gao, X., Conner-Garcia, T., Graumlich, J. F., Murray, M. D., Morrow, D. G., & Stine-Morrow, E. A. L. (2015). Memory and comprehension for health information among older adults: Distinguishing the effects of domain-general and domain-specific knowledge. *Memory*, *23*(4), 577–589.

Citron, F. M. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, *56*, 79–89.

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S. D., 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.

Correa, E. I., & Adams, H. E. (1981). The validity of the preemployment polygraph examination and the effects of motivation. *Polygraph*, *10*(3), 143–155.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.

Davson, H. (1990). *Physiology of the eye* (5th ed.). Pergamon Press.

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.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper,

H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.

Draper Laboratory. (2008, July 29). *Blink extraction*.

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology*, *26*(4), 442–451.

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, *16*(8), 1146–1153.

García-Madruga, J. A., Elosúa, M. R., Gil, L., Gómez-Veiga, I., Vila, J. Ó., Orjales, I., Contreras, A., Rodríguez, R., Melero, M. Á., & Duque, G. (2013). Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Reading Research Quarterly*, *48*(2), 155–174.

Gustafson, L. A., & Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology*, *47*, 408–411.

Gustafson, L. A., & Orne, M. T. (1965). The effects of perceived role and role success on the detection of deception. *Journal of Applied Psychology*, *49*, 412–417.

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* (pp. 159–216). Academic Press.

Handler, M. D., Honts, C. R., Krapohl, D. J., Nelson, R., & Griffin, S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Police and Criminal Psychology*, *24*(2), 69–86.

Handler, M. D., & Nelson, R. (2007). Polygraph terms for the 21st century. *Polygraph*, *36*(3), 157–164.

Honts, C. R., & Amato, S. (2007). Automation of a screening polygraph test increases accuracy. *Psychology, Crime & Law*, *13*(2), 187–199.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (2002). The scientific status of research on polygraph techniques: The case for polygraph tests. In D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Vol. 2, pp. 446–483). West Publishing.

Honts, C. R., & Thurber, S. (2019, March). *A comprehensive meta-analysis of the comparison question polygraph test*. Annual Meeting of the American Psychology Law Society, Portland, Oregon.

Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, *97*(Pt 1), 31–50.

Iacono, W. G., & Patrick, C. J. (1987). What psychologists should know about lie detection. In I. B. Weiner & A. K. Hess (Eds.), *Handbook of forensic psychology* (pp. 460–489). John Wiley & Sons.

Janisse, M. P., & Bradley, M. T. (1980). Deception, information and the pupillary response. *Perceptual and Motor Skills*, *50*(3), 748–750.

Johnson, R., 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.

Jonassen, D. H., Grabowski, B. L., & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and instruction* (1st ed.). Routledge.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kelley, C. M., & McLaughlin, A. C. (2012). Individual differences in the benefits of feedback for learning. *Human Factors*, *54*(1), 26–35.

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. (2018). Ocular-motor deception test. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 187–212). Elsevier.

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*, *117*(12), 79–90.

Krapohl, D. J. (2002). The polygraph in personnel screening. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 217–236). Academic Press.

Krapohl, D. J., Handler, M. D., & Sturm, S. (2012). *Terminology reference for the science of psychophysiological detection of deception* (Third). American Polygraph Associatoin.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(1), 18–27.

Lanius, R. A., Rabellino, D., Boyd, J. E., Harricharan, S., Frewen, P. A., & McKinnon, M. C. (2017). The innate alarm system in PTSD: Conscious and subconscious processing of threat. *Current Opinion in Psychology*, *14*, 109–115.

Lee, G., Ojha, A., Kang, J.-S., & Lee, M. (2015). Modulation of resource allocation by intelligent individuals in linguistic, mathematical and visuo-spatial tasks.

International Journal of Psychophysiology, *97*(1), 14–22.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (1st ed., pp. 55–75). Elsevier.

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*(2), 164–177.

Lumley, T. (2020). *Regression subset selection* (3.1) [Computer software].

Martin, S. L., & Terris, W. (1991). Predicting infrequent behavior: Clarifying the impact on false-positive rates. *Journal of Applied Psychology*, *76*(3), 484–487.

Megalakaki, O., Ballenghein, U., & Baccino, T. (2019). Effects of valence and emotional intensity on the comprehension and memorization of text. *Frontiers in Psychology*, *10*, 179.

Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 174–184. http://aclweb.org/anthology/P18-1017

National Research Council. (2003). *The polygraph and lie detection*. National Academies Press.

Nelson, R. (2015). Scientific basis for polygraph testing. *Polygraph*, *44*(1), 28–61.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4.

Paas, F., Tuovinen, J. E., Tabbers, H., & Gerven, P. W. M. V. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Patnaik, P. (2015). *Ocular-motor methods for detecting deception: Effects of practice feedback and blocking*. University of Utah.

Patnaik, P., Woltz, D. J., Hacker, D. J., Cook, A. E., Francke ramm, M. de 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.

Phan, K. L., Liberzon, I., Welsh, R. C., Britton, J. C., & Taylor, S. F. (2003). Habituation of rostral anterior cingulate cortex to repeated emotionally salient pictures. *Neuropsychopharmacology*, *28*(7), 1344–1350.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of

emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*(2), 331–348.

Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual differences and cognitive load theory. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (1st ed., pp. 65–87). Cambridge University Press.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin*, *84*(4), 782–799.

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of Applied Psychology*, *89*(6), 1099–1105.

R Core Team. (2017). *R: A language and environment for statistical computing.* (3.4.2) [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Rajkowski, J., Majczynski, H., Clayton, E., & Aston-Jones, G. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. *Journal of Neurophysiology*, *92*(1), 361–371.

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* (pp. 587–605). Lawrence Erlbaum.

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

Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 1–48). Academic Press.

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* (pp. 63–129). Academic Press.

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*. Prentice Hall.

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function part I: Principles of functional organisation. *Current Neuropharmacology*, *6*, 235–253.

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2012). Emotion words affect eye fixations

during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 783–792.

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2014). Emotion words and categories: Evidence from lexical decision. *Cognitive Processing*, *15*(2), 209–215.

Seigneuric, A., Ehrlich, M.-F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*(1), 81–103.

Senter, S., Weatherman, D., Krapohl, D. J., & Horvath, F. (2010). Psychological set or differential salience: A proposal for reconciling theory and terminology in polygraph testing. *Polygraph*, *39*(2), 109–117.

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

Sladky, R., Höflich, A., Atanelov, J., Kraus, C., Baldinger, P., Moser, E., Lanzenberger, R., & Windischberger, C. (2012). Increased neural habituation in the amygdala and orbitofrontal cortex in social anxiety disorder revealed by fMRI. *PLoS ONE*, *7*(11), e50050.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, *12*(3), 2849–2853.

Swanson, H. L. (2011). Intellectual growth in children as a function of domain specific and domain general working memory subgroups. *Intelligence*, *39*(6), 481–492.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285.

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

Tillman, C. M., Bohlin, G., Sørensen, L., & Lundervold, A. J. (2009). Intelligence and specific cognitive abilities in children. *Journal of Individual Differences*, *30*(4), 209–219.

Tully, K., & Bolshakov, V. Y. (2010). Emotional enhancement of memory: How norepinephrine enables synaptic plasticity. *Molecular Brain*, *3*(1), 15.

U.S. Department of Defense Polygraph Institute. (1995a). *A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats* (No. DoDPI94-R-0008). U.S. Department of Defense Polygraph

Institute. https://fas.org/sgp/othergov/polygraph/tes.html

U.S. Department of Defense Polygraph Institute. (1995b). *Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage* (No. DoDPI94-R-0009). U.S. Department of Defense Polygraph Institute.

Vendemia, J. M. C., Buzan, R. F., & Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *The American Journal of Psychology*, *118*, 413–429.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*(4), 141–142.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, *24*, 33–49.

Walczyk, J. J., Roper, K. S., Seeman, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*, 755–774.

Webb, A. K., Honts, C. R., Kircher, J. C., Bernhardt, P., & 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.

Woltz, D. J. (1990). Repetition of semantic comparisons: Temporary and persistent priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 392–403.

Wright, C. I., Fischer, H., Whalen, P. J., McInerney, S. C., Shin, L. M., & Rauch, S. L. (2001). Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. *Neuroreport*, *12*(2), 379–383.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). Academic Press.

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Andrew Carlos Potts**

has been approved by the following supervisory committee members:

John C. Kircher	, Chair	**04/29/2020** Date Approved
Andrea Webb	, Member	**04/30/2020** Date Approved
Anne E. Cook	, Member	**04/30/2020** Date Approved
Dan J. Woltz	, Member	**05/01/2020** Date Approved
Douglas Hacker	, Member	**04/29/2020** Date Approved
Seung-Hee Son	, Member	**04/29/2020** Date Approved

and by **Jason Burrow-Sánchez** , Chair/Dean of

the Department/College/School of **Educational Psychology**

and by David B. Kieda, Dean of The Graduate School.