Eye Movements and Pupil Size Reveal Deception in Computer Administered Questionnaires

Andrea K. Webb¹, Douglas J. Hacker¹, Dahvyn Osher¹, Anne E. Cook¹,
Dan J. Woltz¹, Sean Kristjansson², and John C. Kircher¹

¹ University of Utah, 1705 E. Campus Center Dr. Rm 327, Salt Lake City, UT 84112

² Washington University, St. Louis, MO
andreakwebb@gmail.com

Abstract. An oculomotor test is described that uses pupil diameter and eye movements during reading to detect deception. Forty participants read and responded to statements on a computerized questionnaire about their possible involvement in one of two mock crimes. Twenty guilty participants committed one of two mock crimes, and 20 innocent participants committed no crime. Guilty participants demonstrated speeded and accurate reading when they encountered statements about their crime and increases in pupil size. A discriminant function of oculomotor measures successfully discriminated between guilty and innocent participants and between the two groups of guilty participants. Results suggest that oculomotor tests may be of value for pre-employment and security screening applications.

Keywords: Oculomotor measures, pupil size, deception.

1 Introduction

Many government agencies routinely conduct credibility assessments to screen applicants for positions in intelligence, security, law enforcement, immigration, and public transportation. Errors in classifying an individual as truthful or deceptive in these settings can have serious consequences for the individual and society. Current screening techniques rely primarily on the polygraph; however, other techniques have been used, such as self-report measures of integrity or personality, behavioral analyses, or speech content analyses [1-3]. Recently, questions have been raised about the validity of the polygraph for screening and its susceptibility to countermeasures. For example, the National Research Council (NRC) was critical of the polygraph for pre-employment screening and highlighted the need for “an expanded research effort directed at methods for detecting and deterring major security threats, including efforts to improve techniques for security screening...” ([4], p. 8). Similarly, self-report integrity tests for screening potential employees have been criticized due to questions about their effectiveness [2], and behavioral and content analyses have their own shortcomings [3].

We developed an alternative test for deception that relies on measures commonly used by cognitive psychologists to study the psychology of reading (e.g., [5]). Specifically, decisions about truth and deception were based on changes in pupil size and

eye movements that occur while participants read and responded to statements about their possible involvement in mock crimes.

The size of the pupil varies with processing load and cognitive resource capacity [6]. Task-evoked pupil responses provide a reliable and sensitive psychophysiological index of the momentary processing load during performance of a wide variety of cognitive activities [7-9]. Although early research suggested that "emotional factors are relatively unimportant as determinants of pupillary responses observed in carefully controlled information-processing tasks" ([10], p. 288), an association has been noted between pupil response and emotional arousal, with larger pupil diameters associated with greater arousal [11].

Regardless of whether the effects of deception on pupil size are associated with changes in cognitive effort or emotional arousal, we predicted that participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully. Consistent with our prediction, Bradley and Janisse [12] and others [13] found that pupil responses to statements on a concealed information polygraph test discriminated between truthful and deceptive subjects. Similarly, Webb and colleagues [14] found that pupil responses were as diagnostic and at least as useful as electrodermal responses for detecting deception on a comparison question polygraph test.

Results from reading studies have shown that when people experience difficulty in reading a word or phrase, their fixations on the text increase in frequency and duration, and they spend more time reading and rereading [5, 15]. Several investigators have used eye movements during reading or viewing images to detect deception. Baker, Stern, and Goldstein [16] presented test questions on a computer monitor and found that participants' fixation durations successfully discriminated between truthful and deceptive responses in 9 of 10 participants. Deceptive responses were associated with longer fixation durations than truthful responses. Other investigators were able to detect attempts to conceal information by analyzing patterns of eye movements while participants viewed images of crime locations [17] or familiar and unfamiliar stimuli [18]. Based on these findings, we predicted that deception would be associated with increased fixation frequency and reading time.

2 Method

2.1 Participants

Twenty-eight males and 12 females from the University of Utah community were recruited from fliers on campus to participate in a psychological experiment for pay and a possible bonus. Participants ranged in age from 18-38 years ($M = 22.32$), were predominantly Caucasian (87.5%), single (75%), and students at the University of Utah (97.5%).

2.2 Apparatus

Participants' eye movements were monitored using an Applied Sciences Laboratory (ASL) Model 501 head-mounted eye tracker. The eye tracker was interfaced with two 1.8 GHz Hewlett Packard desktop computers: one ran the eye tracker and recorded

the data, and the other ran the experiment. Participants had freedom of head movement while wearing the eye tracker. Viewing was binocular, and eye movement was recorded from each participant's right eye 60 times per second. Participants' head movements and orientation were recorded with a magnetic head tracker, the output of which was stored with eye position and pupil diameter at 60 Hz.

2.3 Materials

Participants responded to 48 statements; 16 statements were neutral (e.g., "The sky is blue on sunny days."), 16 statements addressed the cash crime, and 16 statements addressed the card crime. Each statement type required an equal number of true and false responses, and each collection of true and false statements was further divided into equal numbers of statements with negation (e.g., "I did not take the \$20 from the secretary's purse.") and without negation (e.g., "I took the \$20 from the secretary's purse."). Eight statements were presented on eight rows on the computer monitor. The rows were 3.23 cm apart on a 54 cm monitor in portrait orientation positioned approximately 72 cm in front of the subject. Participants used a mouse to select one of two radio buttons (True or False) on the right edge of the monitor adjacent to each of the eight statements. When participants completed a page, they clicked a button at the bottom of the screen to advance to the next display that presented eight new statements. The three types of statements did not differ significantly by number of words, but they did differ by number of characters (neutral $M = 38.25$, cash $M = 45.88$, card $M = 50.88$; $p < .01$). To adjust for differences in statement length, number of fixations was converted to number of fixations per character, and first- and second-pass reading times were converted to ms per character.

2.4 Measures

Pupil diameter and three measures of reading were obtained for each statement and repetition. All measures were recorded only when participants had fixated within a rectangular region-of-interest that surrounded each statement. The region-of-interest was 32 mm in height, started with the first character, and ended with the last character of the statement.

Pupil diameter. Reading onset was defined as the first sample of the first of four consecutive fixations in the region-of-interest. The difference in pupil diameter between the first sample and each subsequent sample for a period of 4 s provided an evoked pupil response curve. In addition to the response curve, the area under the curve was computed to obtain a global measure of the magnitude of the pupil response. Area under the pupil response curve was the sum of positive differences between the low point that followed reading onset and each subsequent sample until the response recovered to the level of the low point or 4 s following reading onset, whichever occurred first.

Measures of reading behaviors. Number of fixations was the number of times a participant fixated in the region-of-interest. First pass duration was the time the participant spent fixating on the statement before leaving the region-of-interest and looking

elsewhere. Second pass duration was the total time the participant spent rereading the statement after once having left the region-of-interest. Total time was the sum of first pass and second pass durations.

Four criteria were used to define a fixation (Eyenal Manual, Applied Sciences Laboratory, Bedford, MA). First, a fixation began at the first of six consecutive samples that occurred within $.5^\circ$ of visual angle. Second, any three consecutive fixation samples farther than 1° of visual angle in the horizontal or vertical direction from the running mean position ended the fixation. Third, the final fixation position was the mean position of all fixation samples between the beginning and end of the fixation period, but any two or fewer consecutive fixation samples that were farther than 1.5 standard deviations from the mean position were excluded from the calculation of the final position. Finally, any fixation duration longer than 1 second was considered an artifact and deleted.

2.5 Procedure

Participants were recruited by placing fliers for the study at various locations around the University of Utah campus. The fliers provided contact information and indicated that participants would receive pay and a potential bonus. When they called, prospective participants were given a brief description of the study, screened for inclusion criteria (i.e., over 18 years old, proficient in English, and able to read without corrective lenses), and given an appointment. Participants were then emailed initial instructions and a map of campus with a description of the study location. Participants were called the day before their scheduled appointment, reminded of their appointment, instructed to get a good night's sleep and not to drink caffeine two hours before their appointment time.

Each participant reported alone to a room on campus, entered the room, closed the door, read and signed the consent form, and read the computer-administered instructions. No researcher was present at the initial study location. After reading the instructions, the participant was given the option to discontinue the study. For those who decided to continue, they were randomly assigned to an innocent condition ($n = 20$) or a guilty condition. Guilty participants were further subdivided into a "Cash" crime condition ($n = 10$) and "Card" crime condition ($n = 10$). Guilty participants were informed that they had no more than 30 min to complete their assigned crime. Participants in the "Cash" crime were instructed to steal \$20 from a secretary's purse. Participants in the "Card" crime were instructed to steal credit card information from a student's computer. Innocent participants were given general descriptions of the crimes but did not enact them. They were told not to report to the testing room until at least 20 minutes after the time they were scheduled to arrive for their appointment.

At the appointed time, participants reported individually to the reading lab. To motivate participants to appear truthful, they were told that in addition to a \$30 payment for participating, a bonus of \$30 would be paid if the participant appeared truthful to all of the statements on the test. All participants were informed that they should do their best to appear truthful on the test, and that the test was based on the idea that a person who committed a crime would have a difficult time answering. Participants' written instructions included the statement, "To appear innocent, you should respond as quickly and as accurately as you possibly can."

Participants completed a demographic questionnaire and were seated in front of the computer monitor. The ASL eye tracker was attached and calibrated. For various reasons (e.g., pupil size, eye shape, corrective lens glare), the eye movements of some participants could not be tracked (27% of prospective participants). Those individuals were paid \$15 and excused from the experiment. Participants then read a set of instructions on the computer screen. The instructions told them that statements would be presented on the computer display and they should indicate if the statement is true or false. The 48 statements were presented three times in different orders separated by an unrelated task that took 5-10 minutes to complete. The presentation of the statements was randomized across participants with the provision that a statement of one type was followed an equal number of times by a statement from the other two types and never was followed immediately by a statement of the same type.

After completion of the testing, the eye-tracker apparatus was removed. Innocent participants were paid \$60, and guilty participants were paid \$30. The participant then was debriefed about the study and asked not to share this information with anyone who might participate in the study.

3 Results

We tested two predictions: Participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully, and deception would be associated with increased fixation frequency, reading time, and rereading time. Also, because statements were presented three times in different orders, we expected that deceptive participants would demonstrate less attenuation of response times across repetitions than innocent participants for statements on which they were required to lie. However, repeated-measures analysis of variance (RMANOVA) showed no meaningful effects of repetitions on any of the outcome measures.

3.1 Pupil Diameter

Pupil diameter was analyzed with RMANOVA. Time was a within-subjects factor with 40 levels (4 seconds at 10 Hz); statement type was a within-subjects factor with three levels (neutral, cash, card); and treatment condition was a between-subjects factor with three levels (innocent, cash-crime, card-crime). Huynh-Feldt corrected degrees of freedom were used for significance testing. There was a significant interaction between statement type and treatment condition, $F(4, 74) = 4.04$, $p = .005$, $\eta^2 = .18$. In addition, there was a significant three-way interaction between statement type, condition, and time, $F(76, 1406) = 2.25$, $p < .001$, $\eta^2 = .11$. Plots of the three-way interaction for each treatment condition are shown in Figure 1. Results supported the prediction that participants' pupil diameters were greatest when they read statements that they answered deceptively. Participants in the cash condition showed the greatest increase in pupil diameter in response to cash items followed by card and neutral items. Participants in the card condition showed the greatest increase in response to

Fig. 1. Evoked pupil responses to *neutral*, *cash*, and *card* statements. *Cash* statements evoked greater increases in pupil size for participants who stole cash (*cash group*), and *card* statements evoked greater increases in pupil size for participants who took credit card information (*card group*).

card items, followed by cash and neutral items. For innocent participants, there was little difference between cash and card items, but pupil diameter was larger to those items than to the neutral items.

3.2 Reading Behaviors

There was no main effect of guilt on number of fixations per character, first pass duration (reading time), second pass duration (rereading time), or total reading time. The interaction between guilt and type of statement was significant for number of fixations per character, $F(4, 74) = 4.38, p = .004, \eta^2 = .19$; first pass duration, $F(4, 74) = 2.55, p = .047, \eta^2 = .12$, and total reading time, $F(4, 74) = 4.19, p = .005, \eta^2 = .18$, but not second pass duration. Effects on total reading time are presented in Figure 2. As expected, number of fixations and first pass duration were correlated with total reading time, and the pattern of results obtained for those three measures were similar.

Contrary to predictions, participants in the cash condition made *fewer* fixations and spent *less* time on cash than card items. Likewise, participants in the card condition made *fewer* fixations and spent *less* time on card items than cash items. Innocent participants showed little difference among cash, card, and neutral items.

3.3 Discriminant Analysis

The outcome measures included pupil diameter, response time from first fixation, number of fixations, first pass, second pass, and total time. For each of the six outcome measures, we generated three potential predictor variables. One was the participant's mean for neutral items. Another was the difference between the mean for crime-related items (cash and card) and the mean for neutral items. The last variable was the difference between the mean for cash items and the mean for card items. To find an optimal subset of variables to discriminate among innocent, cash, and card groups, two preliminary step-wise discriminant analyses were conducted. One analysis selected variables to maximize the separation between guilty and innocent participants. The second analysis selected variables to maximize the separation between the two guilty groups (cash and card). The selected variables then were included in a final multiple group discriminant analysis. This analysis yielded two discriminant functions.

Five predictor variables were included in the two discriminant functions. The first function discriminated card participants from cash and innocent participants. It relied on the difference between the cash and card statements in total reading time, the difference between the crime-related and neutral statements in pupil diameter, and the difference between cash and card statements in pupil diameter. The second function discriminated between the cash and innocent participants. It relied on the difference between cash and card statements in total reading time, the difference between the crime-related and neutral statements in rereading time, and the difference between crime-related and neutral statements in pupil diameter.

The mean classification accuracy was 78.3%. In the present study, chance accuracy was 37.5%. Ninety-five percent of innocent participants, 80% of card participants, and 60% of cash participants were classified correctly. The discriminant functions had the most difficulty discriminating between cash and innocent participants. An internal validation of the discriminant analysis (jackknife) yielded 58.3% correct classifications. Seventy-five percent of innocent participants, 70% of card participants, and 30% of cash participants were classified correctly.

Fig. 2. Total time per character for participants guilty of taking cash (*Cash Group*), downloading credit card information (*Card Group*), or innocent of both mock crimes (*Innocent Group*). Participants in the *Cash Group* spent more time on *card* statements than cash statements. Conversely, participants in the *Card Group* spent more time on *cash* statements than *card* statements.

4 Discussion

Our purpose was to develop and evaluate a new test for detecting deception. We obtained oculomotor measures from guilty and innocent participants while they read and responded to neutral or crime-related statements on a computerized questionnaire. We predicted that participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully, and deception would be associated with increased fixation frequency, reading time, and rereading time. Only the first prediction was supported. Consistent with other research [12-14], deception was associated with increased pupil size. The increase in pupil size may have been due to increased cognitive load or emotional arousal. Contrary to predictions, deception was not associated with increased fixation frequency or reading time. Rather, guilty participants made *fewer* fixations and spent *less* time reading crime-related statements answered deceptively than crime-related statements answered truthfully. These findings were obtained in the present experiment for both the cash-crime group and the card-crime group, and the findings have since been replicated in two other experiments [19-20].

Innocent participants spent about the same amount of time on all test items. In contrast, guilty participants spent significantly less time on the statements about the crime they had committed than about the crime they did not commit. The pupil data suggest that guilty participants invested more mental effort in the processing of statements answered deceptively than truthfully, and the behavioral data indicate that they were successful in reducing the time they spent reading the statement. In fact, they were too successful; they overcorrected. If overcorrection was a consequence of a conscious

decision to read the incriminating items faster, it should be possible to train individuals to avoid detection by adopting a more general strategy to respond to all test items similarly.

On the other hand, overcorrection may be a consequence of a cognitive distortion known as a contrast effect or a salience effect [21]. Because the person's goal is to avoid detection when they lie, items answered deceptively stand out against a background of items answered truthfully. It has been argued that cognitive biases have evolved to facilitate information processing and protect the organism from harm (e.g., [22]). The cognitive distortion hypothesis is consistent with our findings, and it explains those of others (e.g., [23]). If it is correct, then the addition of multiple categories of items answered truthfully should increase the salience of items answered deceptively and the behavioral effects. In addition, training to defeat the test may have little effect on reading behaviors.

Oculomotor and reading measures were used to develop discriminant functions that discriminated among the three treatment conditions. The analysis yielded a mean correct classification percentage of 78.3% and a mean jackknifed classification percentage of 58.3%. The classification accuracies for the original discriminant analysis are comparable to those obtained for polygraph testing in screening contexts, but the jackknifed classification accuracies are not. The lukewarm results may have been due to the presentation of test items in a traditional questionnaire format. We presented eight statements on the computer monitor at once. Participants were free to reread statements, read statements out of order, and change their answers. Subsequent experiments in which individual statements were presented serially have produced larger effects and classification accuracies that exceed 85% [19-20]. It appears that the lack of constraint inherent in a more traditional questionnaire format introduced error in our measurements and attenuated the behavioral and physiological effects.

In conclusion, the present findings suggest that reading behaviors may be used to detect deception and may supplement or provide an alternative to the polygraph or self-report. Additional research is needed to test for effects of countermeasures, assess the effects of adding items to cover multiple issues, test if similar effects can be observed in real-world screening contexts, and determine if changes in pupil size in this context reflect changes in cognitive load or emotional arousal.

References

1. Krapohl, D.J.: The Polygraph in Personnel Screening. In: Kleiner, M. (ed.) *Handbook of Polygraph Testing*, pp. 217-236. Academic Press, San Diego (2002)
2. Office of Technology Assessment: *The Use of Integrity Tests for Pre-employment Screening* (OTA-SET-442). U. S. Government Printing Office, Washington, DC (1990)
3. Vrij, A.: *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn. John Wiley & Sons Ltd, West Sussex (2008)
4. National Research Council: *The Polygraph and Lie Detection*. The National Academies Press, Washington, DC (2003)
5. Rayner, K.: Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol. Bull.* 124, 372-422 (1998)
6. Kahneman, D.: *Attention and Effort*. Prentice-Hall, Englewood Cliffs (1973)

7. Beatty, J., Wagoner, B.L.: Pupillometric Signs of Brain Activation Vary with Level of Cognitive Processing. *Science* 199, 1216–1218 (1978)
8. Kahneman, D., Beatty, J.: Pupil Diameter and Load on Memory. *Science* 154, 1583–1585 (1966)
9. Schluroff, M.: Pupil Responses to Grammatical Complexity of Sentences. *Brain Lang.* 17, 133–145 (1982)
10. Beatty, J.: Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychol. Bull.* 91, 276–292 (1982)
11. Stern, R.M., Ray, W.J., Quigley, K.S.: *Psychophysiological recording*, 2nd edn. Oxford University Press, New York (2001)
12. Bradley, M.T., Janisse, M.P.: Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures. *Psychophysiology* 18, 307–315 (1981)
13. Lubow, R.E., Fein, O.: Pupillary Size in Response to a Visual Guilty Knowledge Test: New Technique for the Detection of Deception. *J. Exp. Psychol. Appl.* 2, 164–177 (1996)
14. Webb, A.K., Honts, C.R., Bernhardt, P., Kircher, J.C., Cook, A.E.: Effectiveness of Pupil Diameter in a Probable-lie Comparison Question Test for Deception. *Legal Criminol. Psychol.* (in press)
15. Rayner, K., Pollatsek, A.: *The psychology of reading*. Prentice-Hall, Englewood Cliffs (1989)
16. Baker, L., Stern, J.A., Goldstein, R.: *The Gaze Control System and the Detection of Deception*. Final report to the U.S. Government (Contract #90-F131400). Washington University, Department of Psychology, St. Louis (1992)
17. Ellson, D.G., Davis, R.C., Saltzman, I.J., Burke, C.J.: *A Report of Research on Detection of Deception*. Technical report prepared for Office of Naval Research (Contract N6onr-18011). Indiana University (1952)
18. Althoff, R.R., Cohen, N.J.: Eye-Movement-Based Memory Effect: A Reprocessing Effect in Face Perception. *J. Exp. Psychol. Learn.* 25, 997–1010 (1999)
19. Osher, D.: *Multimethod Assessment of Deception: Oculomotor Movement, Pupil Size, and Response Time Measures*. Dissertation Abstracts International (2007)
20. Webb, A.K.: *Effects of Motivation, and Item Difficulty on Oculomotor and Behavioral Measures of Deception*. Unpublished dissertation, University of Utah, Department of Educational Psychology (2008)
21. Guttel, E.: Overcorrection. *Georgetown Law J.* 93, 241–283 (2004)
22. Haselton, M.G., Nettle, D., Andrews, P.W.: The Evolution of Cognitive Bias. In: Buss, D.M. (ed.) *Handbook of Evolutionary Psychology*, pp. 724–746. Wiley, Hoboken (2005)
23. Meijer, E.H., Smulder, F.T.Y., Johnston, J.E., Merckelbach, H.L.G.J.: Combining Skin Conductance and Forced Choice in the Detection of Concealed Information. *Psychophysiology* 44, 814–822 (2005)