Journal of Experimental Psychology: Applied

Lyin' Eyes: Ocular-Motor Measures of Reading Reveal Deception

Anne E. Cook, Douglas J. Hacker, Andrea K. Webb, Dahvyn Osher, Sean D. Kristjansson, Dan J. Woltz, and John C. Kircher

Online First Publication, April 30, 2012. doi: 10.1037/a0028307

Journal of Experimental Psychology: Applied

1076-898X/12/$12.00 DOI: 10.1037/a0028307

Lyin' Eyes: Ocular-Motor Measures of Reading Reveal Deception

Anne E. Cook and Douglas J. Hacker
University of Utah

Andrea K. Webb
Draper Laboratory, Cambridge, Massachusetts

Dahvyn Osher
University of Utah

Sean D. Kristjansson
Washington University School of Medicine

Dan J. Woltz and John C. Kircher
University of Utah

Our goal was to evaluate an alternative to current methods for detecting deception in security screening contexts. We evaluated a new cognitive-based test of deception that measured participants' ocular-motor responses (pupil responses and reading behaviors) while they read and responded to statements on a computerized questionnaire. In Experiment 1, participants from a university community were randomly assigned to either a "guilty" group that committed one of two mock crimes or an "innocent" group that only learned about the crime. Participants then reported for testing, where they completed the computer-administered questionnaire that addressed their possible involvement in the crimes. Experiment 2 also manipulated participants' incentive to pass the test and difficulty of statements on the test. In both experiments, guilty participants had increased pupil responses to statements answered deceptively; however, they spent less time fixating on, reading, and rereading those statements than statements answered truthfully. These ocular-motor measures were optimally weighted in a discrimination function that correctly classified 85% of participants as either guilty or innocent. Findings from Experiment 2 indicated that group discrimination was improved with greater incentives to pass the test and the use of statements with simple syntax. The present findings suggest that two cognitive processes are involved in deception—vigilance and strategy—and that these processes are reflected in different ocular-motor measures. The ocular-motor test reported here represents a new approach to detecting deception that may fill an important need in security screening contexts.

Keywords: deception detection, ocular-motor measures, pupil size, cognitive load

Many government agencies and private corporations routinely conduct credibility assessments to screen applicants for positions in intelligence, national and private 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. A National Research Council (NRC) report 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 . . ." (National Research Council, 2003, p. 8). Although other techniques have been used, such as self-report measures of integrity or personality, behavioral analyses, or speech content analyses (Krapohl, 2002; Meesig & Horvath, 1995; Office of Technology Assessment, 1990; Sackett & Wanek, 1996; Vrij, 2008), their validity has also been questioned.

Traditional polygraph approaches rely on within-subject comparisons of psychophysiological responses to questions, and examinees' emotions are presumed to play a strong role in these responses. More recently, several new cognition-based tests for deception have been developed, all of which are based more or less on the notion that mentally it is more difficult to lie than to tell the truth (e.g., Johnson, Barnhardt, & Zhu, 2005; Seymour et al., 2000; Vendemia, Buzan, & Green, 2005; Vrij et al., 2009; Vrij, Mann, Kristen, & Fisher, 2007; Walczyk, Mahoney, Doverspike, & Griffith-Ross, 2009; Walczyk, Roper, Seemann, & Humphrey, 2003). The purpose of the current two studies was to assess an

Anne E. Cook, Douglas J. Hacker, Dahvyn Osher, Dan J. Woltz, and John C. Kircher, Educational Psychology Department, University of Utah; Andrea K. Webb, Draper Laboratory, Cambridge, Massachusetts; Sean D. Kristjansson, Department of Psychiatry, Washington University School of Medicine.

This research was funded by a seed grant from the University of Utah to John C. Kircher and Douglas J. Hacker. Portions of this research were conducted by Dahvyn Osher and Andrea K. Webb in partial fulfillment of the requirements for their PhD degrees at the University of Utah. Drs. Cook, Hacker, Woltz, and Kircher have financial interests in Credibility Assessment Technologies, which is attempting to develop and market the deception detection technology. This company was not formed until after the research was conducted.

Correspondence concerning this article should be addressed to Anne E. Cook, Educational Psychology Department, 1705 Campus Center Dr., Rm. 113, University of Utah, Salt Lake City, UT 84115. E-mail: Anne.Cook@utah.edu

ocular-motor test for deception that was based on the cognitive workload hypothesis, as well as other previous empirical findings from the literature in psychology of reading (e.g., Just & Carpenter, 1993; Rayner, 1998). Specifically, we based decisions about truth and deception on pupil responses and eye movements that occurred while participants read and responded to statements about their possible involvement in a mock crime.

The size of the pupil reflects within-task, between-task, and between-individual variations in processing load (Kahneman, 1973). Increased pupillary responses are associated with increased difficulty on a wide variety of cognitive tasks, including recall and transformation of digit strings (Kahneman & Beatty, 1966), mental multiplication (Ahern & Beatty, 1979; Hess & Polt, 1964), sentence processing (Just & Carpenter, 1993; Schluroff, 1982), letter processing (Beatty & Wagoner, 1978), and lexical translation (Hyona, Tommola, & Alaja, 1995).

Consistent with the cognitive workload hypothesis, deception researchers have also found increases in pupil size associated with deception (Berrien & Huntington, 1943; Dionisio, Granholm, Hillix, & Perrine, 2001; Heilveil, 1976; Lubow & Fein, 1996). Indeed, pupil responses to statements on polygraph tests can be used to discriminate between truthful and deceptive participants (Bradley & Janisse, 1981) and may be at least as diagnostic as electrodermal responses (Webb, Honts, Bernhardt, Kircher, & Cook, 2009). Thus, we predicted that participants would show greater increases in pupil diameter in response to statements answered deceptively than to statements answered truthfully.

In the literature on the psychology of reading, eye tracking technology has been used to record the location and duration of eye fixations as people read and respond to text (Just & Carpenter, 1980; McConkie, Hogaboam, Wolverton, Zola, & Lucas, 1979; Rayner, 1998). Results from these studies have shown that when people experience difficulty in reading a word or phrase, their fixations on the text increase in frequency and duration, intersaccade distances decrease, and they spend more time reading and rereading (Rayner, 1998; Rayner & Pollatsek, 1989). Reading researchers also have argued that the initial fixations on a passage of text reflect readers' first attempts to integrate and comprehend incoming information, whereas subsequent fixations may reflect readers' attempts to verify the information against other sources of knowledge (Cook & Myers, 2004; Garrod & Terras, 2000) or other more strategic processing goals (Hyona & Nurminen, 2006; Long & Lea, 2005). The effects of deception on response times to words presented visually on a computer monitor are well established (e.g., Seymour et al., 2000). However, if the reading researchers are correct, simple response time may reflect a combination of different reading processes that differ in sensitivity to the effects of deception. The present study measured response time but also distinguished between several measures of reading behavior that can occur within a single response: fixation frequency (number of fixations), initial reading time (first-pass duration), and subsequent rereading (second-pass duration).

Several investigators have used eye movements to detect deception or attempts by participants to conceal information. Baker, Stern, and Goldstein (1992b) presented test questions on a computer monitor and found that fixation durations were longer for deceptive than for truthful participants. Other investigators were able to detect attempts to conceal information by analyzing patterns of eye movements while participants viewed images of crime locations (Ellson, Davis, Saltzman, & Burke, 1952) or familiar and unfamiliar stimuli (Althoff & Cohen, 1999). Based on these findings, we predicted that deception would be associated with increases in number of fixations, first-pass duration, and second-pass duration.

We tested these predictions in two experiments using a mock crime paradigm. In each experiment, participants were randomly assigned to guilty and innocent treatment conditions. Guilty participants committed a simulated theft, whereas innocent participants were informed of the thefts but did not commit them. We chose to inform innocent participants of the crimes because in field testing situations, innocent individuals often have knowledge of a crime even if they did not commit it. However, because the experimenter who administered the test was unaware of the participants' guilt status, all participants were suspected of the thefts and were offered a monetary bonus to convince the experimenter of their innocence. Participants were told to deny having committed either crime, and that they should respond to test statements as quickly and accurately as possible because delays in responding or errors could be taken as indicators of deception. Participants were fitted with a head mounted eye tracker and selected True or False in response to statements that were neutral or addressed each of two possible mock crimes (e.g., "I did not take the $20 from the secretary's purse"). The first experiment tested predictions that guilty participants would show greater increases in pupil diameter, increased fixations, and longer first pass and second pass reading times in response to statements answered deceptively compared with those answered truthfully. The second experiment was designed to replicate and assess the reliability of results from the first experiment and investigate the role of other variables that could influence the accuracy of an ocular-motor test for deception. All procedures in Experiment 1 and Experiment 2 were approved by the University of Utah Institutional Review Board.

Experiment 1

Method

Design. Equal numbers of participants were randomly assigned to one of three cells in a 3 × (3 × 3) mixed design. The between-subjects variable was guilt; participants were assigned to either an innocent condition or a guilty condition. Guilty participants were further divided into "Cash" or "Card" crime conditions. The two within-subjects factors were statement type (neutral, cash, and card) and repetition (three repetitions of each statement). The dependent variables were change in pupil diameter, response time, response errors, number of fixations, first pass reading time, and second pass reading time.

Participants. Fifty-five university students and staff were recruited from fliers on campus. For various reasons (e.g., pupil size, eye shape, etc.), the eye movements of 27% of prospective participants could not be tracked. Those individuals were paid $15 and excused from the experiment. The remaining 24 male and 16 female participants ranged in age from 18–36 years ($M = 22.35$, $SD = 4.3$), were predominantly Caucasian (82.5%), single (77.5%), and students at the University of Utah (92.5%). Participants were randomly assigned to innocent ($n = 20$), "Cash" ($n = 10$), and "Card" conditions ($n = 10$).

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 (i.e., 60 Hz). 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. The monitor was positioned approximately 63.5 cm from the participant's eyes.

Materials. The 48 test statements were repeated three times during the experiment in separate trial blocks. The statements were divided into three types: 16 statements were neutral, 16 addressed the cash crime, and 16 addressed the card crime. Each statement type required an equal number of true and false responses, and each group of True and False statements was subdivided 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."). The mean length in characters (with spaces) of the statements in the neutral, cash, and card conditions were 48.06 (SD = 9.33), 57.25 (SD = 9.01), and 65.56 (SD = 9.93) characters, respectively. The neutral statements were shorter than either the cash, $t(30) = 2.83$, $p < .01$, or the card statements, $t(30) = 4.26$, $p < .01$, but the cash and card statements did not differ in length, $t(30) = 1.58$, $p = .12$.

Measures. Response time, response accuracy, pupil diameter, and three measures of reading (number of fixations, first-pass duration, and second-pass duration) 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.

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° of visual angle. Second, any three consecutive fixation samples farther than 1° of visual angle in the horizontal or vertical direction from the mean location of the preceding samples 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 automatically deleted (see Rayner, 1998; Rayner & Pollatsek, 1989).

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 and represented changes in pupil response over time from stimulus onset. In addition to the response curve, the area under the curve was computed to obtain a single measure of the magnitude of the pupil response for reliability, bivariate, and discriminant function analyses described below. Area under the pupil response curve was calculated from the low point that followed reading onset until the pupil response fully recovered, or 4 s following reading onset, whichever occurred first. Area under this response curve was the sum of positive differences between the initial low point and each subsequent sample (Kircher & Raskin, 1988).

Measures of reading behaviors. Number of fixations was the number of times a participant fixated in the region-of-interest; this measure is a general indicator of processing difficulty. First-pass duration was the sum of durations of all fixations made from the time the participant first fixated on the statement until he or she fixated somewhere outside the region-of-interest; this measure reflects the amount of time initially spent reading a region. Second-pass duration was the sum of durations for all fixations that the participant made while rereading the statement after once having fixated outside the region-of-interest; this measure of re-reading may reflect lingering difficulty associated with a region and/or strategic processes. To adjust for differences in length as a function of statement type, number of fixations was converted to number of fixations per character, and response time and first- and second-pass reading times were converted to ms per characters. This is a common practice in controlling for differences in length in stimulus materials in psychological research on reading (see Rayner, 1998).

Procedure. Participants were recruited by placing fliers for the study at various locations around the university campus. In addition to providing contact information, the fliers indicated that participants would receive $30 in pay and a possible bonus of $30. 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 e-mailed 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 2 hours before their appointment time.

Per their instructions, 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. Those participants who decided to continue were randomly assigned to an innocent condition or a guilty condition. Guilty participants were further subdivided into "Cash" and "Card" crime conditions. Guilty condition 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. Innocent participants were told to report to the lab 20 to 35 min after the time they were scheduled to arrive for their appointment. This was to ensure that guilty and innocent participants arrived at the lab at approximately the same time so that the experimenter administering the test remained unaware of the conditions to which they had been assigned.

An attempt was made to motivate all participants to pass the test. Guilty and innocent participants were promised a $30 bonus in addition to their $30 in pay ($60 total) if they appeared truthful to all of the statements on the test. Prior to their arrival at the lab, all participants also were given the following instruction:

> [Y]ou must not make the examiner suspicious at any point during the test. The test is based on the idea that a person who committed a crime will have a difficult time answering quickly and honestly to questions about the crime. You could make the examiner suspicious if it takes you a long time to answer the questions or if you make lots of mistakes. To appear innocent, you should respond as quickly and as accurately as you possibly can.

Upon arriving at the lab, each participant completed a brief demographic questionnaire and was seated in front of the computer monitor. The ASL eye tracker was attached and calibrated. The participants then read a set of instructions on the computer screen. They were informed that statements would be presented individually on the computer screen, and they should indicate whether each statement was true or false.

Each statement was presented on a single line in the center of the computer monitor beginning at the left edge of the screen. To answer True or False, the participant used a mouse to click one of two radio buttons that appeared on the right side of the screen adjacent to the statement. The selected radio button showed the participant's response for 200 ms, and the statement was replaced by the next statement in the preprogrammed sequence 200 ms later.

The 48 statements were repeated in three blocks separated by an unrelated test of general knowledge that took 5–10 min 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 was removed and the participant was informed if they passed the test. At the conclusion of the test, each participant's guilt status automatically was retrieved via a local area network from the computer used to assign participants to treatments and administer the instructions. Because no algorithm had been developed to decide if the participant was truthful or deceptive on the test, participants assigned to the innocent condition were paid $60, and those assigned to one of the two guilty conditions were paid $30. Participants were debriefed about the study and payment procedures and were asked not to share this information with anyone who might participate in the study.

Results

The level of significance for all statistical tests was set at .05. Significance levels for tests involving within-subjects factors were conservatively assessed with Huynh-Feldt adjusted degrees of freedom.

Repetition effects. We tested if reading and response times for the statements would decrease across repetitions, and whether they differed as a function of statement type and guilt status. Repeated measures analysis of variance (RMANOVA) revealed no meaningful effects of repetitions on any of the outcome measures, and no interactions with statement type or guilt status. Because changes over repetitions were not diagnostic, means were computed across repetitions for all outcome measures.

Response errors. There were few errors overall. Analysis of variance revealed no significant main or interaction effects of guilt status on response errors.

Response time. The response time results, as well as the reading measures, are reported in Table 1. Analysis of response times revealed a significant main effect for group, $F(2, 37) = 4.24$, partial $\eta^2 = .19$. Guilty participants showed longer response times than innocent participants across all statement domains. In addition, the analysis revealed a significant Group $\times$ Statement Type interaction, $F(3.87, 71.63) = 3.18$, partial $\eta^2 = .15$. Participants who were guilty of stealing the cash took less time per character on the cash statements than on either the neutral or card statements. Conversely, participants who were guilty of stealing the credit card information took less time on the card statements than on either the cash or neutral statements. Innocent participants' response times did not differ as a function of statement type.

Pupil diameter. Changes in pupil size for 4 s following statement presentation are plotted at 10 Hz in Figure 1 for each group. The initial drop in pupil size is a common response to the onset of a visual stimulus (Reinhard, Lachnit, & Konig, 2006; Steinhauer & Hakerem, 1992). Innocent participants (Figure 1c) appeared to show greater increases in pupil size in response to crime-relevant statements than neutral statements, but the difference was relatively small, and their responses to cash and card statements were similar. Conversely, participants guilty of the cash

Table 1
Mean Response Time, Number of Fixations, and First- and Second-Pass Duration as a Function of Group and Statement Type in Experiment 1

	Guilty-cash						Guilty-card						Innocent					
	Cash		Card		Neutral		Cash		Card		Neutral		Cash		Card		Neutral	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Response time	50.13	1.21	54.81	1.46	51.79	1.22	51.24	.96	47.93	.82	52.97	1.03	42.92	.85	42.58	1.19	41.59	.89
Number of fixations	19.6	3.17	21.83	3.87	20.42	3.35	19.14	3.03	17.81	2.37	19.38	2.39	17.20	2.87	17.31	3.9	16.89	2.89
First-pass duration	42.20	1.15	42.17	1.2	43.93	1.16	40.97	.83	39.47	.62	46.02	.82	35.01	.69	34.16	.79	35.61	.75
Second-pass duration	8.0	.27	13.0	.53	8.0	.30	10.0	.40	8.0	.27	7.0	.31	8.0	.51	8.0	.64	6.0	.37

Note. Number of fixations are reported per character, whereas response time, first-pass duration, and second-pass duration are reported in ms per character. *SD*s for all measures appear in parentheses.

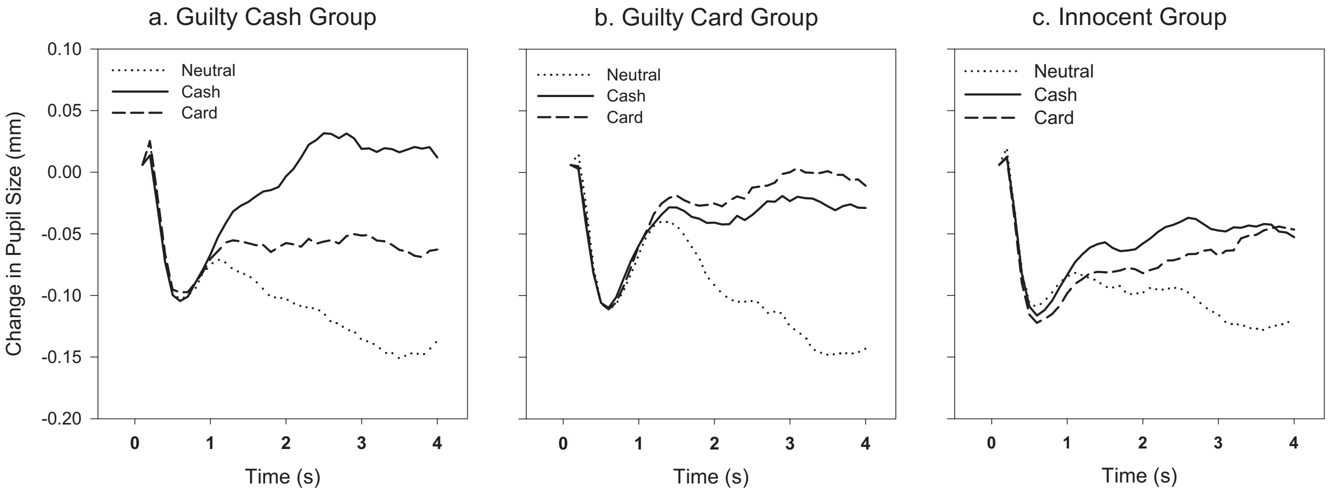

Figure 1. Mean evoked pupil response as a function of group and statement type for Experiment 1.

crime (Figure 1a) showed large increases in pupil size in response to cash statements and weaker responses to card and neutral statements, whereas participants guilty of the card crime (Figure 1b) showed their strongest responses to card statements.

Change in pupil size over the course of a response was analyzed with RMANOVA with two within-subjects factors and one between-groups factor. Time was a within-subjects factor with 40 levels (4 s at 10 Hz), statement type was a within-subjects factor with three levels (neutral, cash, and card statements), and guilt was a between-groups factor with three levels (innocent, guilty of cash crime, guilty of card crime). The Statement Type × Guilt interaction was significant, $F(4, 74) = 5.30$, partial $\eta^2 = .22$. The greatest increases in pupil diameter occurred when the participant was deceptive. The Statement Type × Guilt interaction is based on differences in mean response levels to each statement across groups, whereas the Time × Statement Type × Guilt interaction is based on differences in the shapes of the response curves for each statement type across groups. This three-way interaction also was significant, $F(19.25, 356.06) = 4.16$, partial $\eta^2 = .18$. The three types of statements evoked changes in the pupil over time that differed among the groups.

In Figure 1, it appeared that the absolute difference between cash and card statements was greater for participants who took the cash than for those who took the credit card information. This was tested with an RMANOVA using only guilty participants, where responses to relevant statements answered deceptively were compared with relevant statements answered truthfully. The analysis revealed no Deception × Statement Type interaction, $p = .14$. However, the three-way interaction of Time × Deception × Statement Type was significant, $F(3.31, 78.67) = 2.49$, partial $\eta^2 = .12$. The difference in pupil response associated with truthfulness and deception was greater for participants in the cash than card condition. In addition, innocent participants reacted more strongly to crime-related statements than to neutral statements. Follow-up tests within the innocent group revealed that the difference between crime-related and neutral statement types was significant, $F(1, 19) = 30.78$, partial $\eta^2 = .62$, as was the Time × Statement Type interaction, $F(5.38, 102.21) = 20.97$, partial $\eta^2 = .52$.

Reading measures. As expected, guilty participants made more fixations per character than did innocent participants; the main effect of guilt was significant, $F(2, 37) = 4.90$, partial $\eta^2 = .21$. Contrary to predictions, however, participants guilty of taking the cash made fewer fixations when reading statements about the cash than when reading about the credit card, and participants guilty of taking the credit card information made fewer fixations on statements about the credit card than the cash (Table 1). For innocent participants, the number of fixations was relatively consistent across statement types. This Statement Type × Guilt interaction was significant, $F(3.92, 72.61) = 4.09$, partial $\eta^2 = .18$.

Analysis of first-pass duration data revealed that guilty participants on average had longer first-pass time reading times than did innocent participants; the main effect of guilt was significant, $F(2, 37) = 4.03$, partial $\eta^2 = .18$. The Statement Type × Guilt interaction also was significant, $F(4, 74) = 3.17$, partial $\eta^2 = .15$. As seen in Table 1, this interaction reflected the fact that guilty participants' first pass reading times were shorter for statements about the two crimes than for statements about neutral content, whereas innocent participants' first-pass durations did not differ as a function of statement type.

There was no main effect of guilt in the analysis of second-pass durations. The Statement Type × Guilt interaction was significant, however, $F(3.98, 73.66) = 3.65$, partial $\eta^2 = .16$. Participants guilty of the cash crime spent less time rereading statements about the cash, whereas participants guilty of the credit card crime spent less time rereading statements about the credit card. The findings for first- and second-pass reading measures were consistent with, though not independent of, those obtained for number of fixations; as noted by Rayner (1998), measures of fixation frequency and fixation duration tend to be highly correlated.

Bivariate and discriminant analyses. Bivariate and discriminant analyses were conducted to assess the degree to which area under the pupil response curve (pupil response), number of fixations, first-pass duration, and second-pass duration could be used to differentiate among the three treatment conditions. For each outcome measure, responses to neutral, cash, and card statements were used to derive three new variables. One variable was the mean response to neutral statements (*N*). The response to neutral statements provided a general measure of vigilance or cautiousness. The observed main effects of guilt on number of fixations and first-pass reading time suggested that guilty participants generally were more cautious than innocent participants. The second variable was the difference between the combined mean response to cash (R1) and card (R2) statements and the response to neutral statements (i.e., [R1 + R2]/2 – N). Guilty participants were expected to show greater differences between crime-related and neutral statements than innocent participants. The last variable was the difference between responses to cash and card statements (i.e., R1 – R2). We expected this difference to differentiate between the two guilty groups.

Two group membership indicator variables were created to distinguish among the three groups. One indicator variable (Guilty-Innocent) differentiated between the guilty (coded + 1) and innocent participants (−1). The second indicator variable (Cash-Card) distinguished between the Card (−1) and Cash groups (+1); innocent participants were coded 0. The correlations of the group indicator variables with the various outcome measures are presented in Table 2. The correlations indicate the extent to which the outcome measure discriminates between the groups and may be viewed as an index of predictive validity. For example, on the Guilty-Innocent indicator variable, a positive correlation would indicate that the guilty group had a higher mean score on the outcome measure than the innocent group.

For neutral statements, guilty participants had smaller pupil responses ($r = -.43$), fixated more often ($r = .48$), and spent more time reading the statements ($r = .49$) than did innocent participants. There were no significant differences between the two guilty groups in their responses to neutral statements. Differences between responses to crime-related and neutral statements distinguished between guilty and innocent groups on measures of pupil response ($r = -.51$) and first-pass duration ($r = .33$). Differences between responses to cash and card statements distinguished between the two guilty conditions on measures of pupil response ($r = .61$), number of fixations ($r = -.54$), and second-pass duration ($r = -.53$). Deceptive answers were preceded by greater increases in pupil diameter, fewer fixations, and shorter second-pass reading times.

Table 2
Correlations Between Ocular-Motor Measures and Group Membership Indicators in Experiment 1

Variable	Guilty vs. innocent	Cash vs. card
Neutral statements		
Pupil response (AUC)	$-.43^{**}$	.21
Number of fixations	$.48^{**}$	−.11
First-pass duration	$.49^{**}$	.08
Second-pass duration	.21	−.10
Crime vs. neutral statements		
Pupil response (AUC)	$.51^{**}$	.26
Number of fixations	−.17	−.21
First-pass duration	$-.33^{*}$	$-.33^{*}$
Second-pass duration	.03	−.01
Cash vs. card Statements		
Pupil response (AUC)	.27	$-.61^{**}$
Number of fixations	−.06	$.54^{**}$
First-pass duration	−.01	.12
Second-pass duration	−.09	$.53^{**}$

Note. AUC indicates area under curve.
$^{*}p < .05$. $^{**}p < .01$.

Stepwise discriminant analysis was used to select subsets of the available ocular-motor measures to classify cases into cash, card, and innocent groups (see Kircher & Raskin, 1988, for a detailed explanation of this procedure). The analysis produced two significant discriminant functions. The first discriminant function used the difference in pupil response between crime-related and neutral statements as well as number of fixations on neutral statements to discriminate between participants in guilty and innocent groups ($R^2 = .49$, $p < .01$). The second discriminant function used the difference in pupil response to cash and card statements and the difference in second-pass rereading of cash and card statements to discriminate between the two guilty groups ($R^2 = .41$, $p < .01$). The discriminant functions correctly classified nine of 10 cash crime participants, 8 of 10 card crime participants, and 17 of 20 innocent participants. Together, the selected ocular-motor measures yielded 85% correct classifications. Mean accuracy dropped from 85% to 80% with jackknifed classifications. Jackknifed classifications were obtained by removing the first participant from the data matrix and computing discriminant functions using the remaining N−1 participants. Those functions were used to classify only the first participant. Since the data for that one participant did not contribute to the solution of the discriminant functions, the classification of that individual was unbiased, or at least less biased, than the classification of the same individual by the discriminant functions based on all *N* participants. The data for the first participant were then added back to the data matrix, the data from second participant were removed from the matrix, and discriminant functions were derived using the remaining N−1 participants (1, 3, 4, . . . N). The second individual was classified by the functions based on Participants 1, 3, 4, . . . N. This procedure was repeated for each participant in the study. In the end, each participant was classified by functions that were independent of the data for that individual.

Discussion

In previous laboratory experiments, deception was associated with increased response time, increased errors, and decreased repetition effects (Baker et al., 1992b; Crosland, 1929; Seymour et al., 2000; Vendemia et al., 2005; Walczyk et al., 2009, 2003). Consistent with prior research and the cognitive workload hypothesis, guilty participants responded more slowly than innocent participants. However, we found no effects of deception on response accuracy, or in repetition effects on either response time or response accuracy.

As in previous research, deception in this experiment was associated with greater increases in pupil size (e.g., Bradley & Janisse, 1981; Webb, Honts, et al., 2009). Participants who stole $20 showed stronger pupil responses to statements about the $20, whereas participants who stole credit card information showed stronger pupil responses to statements about the credit card. In addition, innocent participants reacted more strongly to crime-related than neutral statements, even though their answers to all statements were truthful. Horowitz, Kircher, Honts, and Raskin (1997) reported a similar finding for innocent participants in a polygraph study. This effect for innocent participants may be attributed to recognition on their part of the significance of the crime-related questions.

The pupil data indicated that innocent participants responded more strongly to cash statements than credit card statements. Similarly, the absolute difference between cash and credit card statements was greater for participants who took the cash than for participants who downloaded credit card information. Why this occurred is unclear, since the two types of statements were equally important to the outcome. One possibility is that participants perceived the theft of $20 from a secretary's purse as a more egregious offense than downloading credit card information. Although we attempted to equate the two sets of crime-related statements, it also is possible that the cash statements were more semantically complex and required more cognitive effort to process. Statement complexity was manipulated in Experiment 2.

Sympathetically mediated electrodermal responses during polygraph examinations habituate rapidly over repetitions of test questions (Ben-Shakhar & Furedy, 1990; Kristjansson, Kircher, & Webb, 2007). However, no effect of repetitions was observed on pupil responses in Experiment 1 (for a similar finding, see Webb, Honts, et al., 2009) The present results may indicate that the pupil changes in these deception tests were mediated by the parasympathetic nervous system and mostly reflect changes in cognitive load (Bradley et al., 2008). Alternatively, the three repetitions of the statements may have provided insufficient opportunity to observe effects of trials on evoked pupil responses. This issue was investigated further in Experiment 2.

The effects on reading measures were partially consistent with expectations. As expected, guilty participants overall made more fixations and took longer to read the test statements than did innocent participants. Contrary to predictions, however, guilty participants made fewer fixations and spent less time reading and rereading when responding deceptively than when responding truthfully. Although these patterns of effects were evident in two independent samples of guilty participants, the samples were small, and the effects were inconsistent with predictions based on previous research in reading (Rayner, 1998) and the detection of deception (Baker et al., 1992b). Therefore, another goal of Experiment 2 was to determine whether the patterns of effects observed in Experiment 1 were reliable, especially those obtained for first- and second-pass reading times.

Experiment 2

Meta-analyses of mock crime research on polygraph techniques (Kircher, Horowitz, & Raskin, 1988) and deception studies (DePaulo et al., 2003) suggest that motivation to deceive increases the diagnostic validity of measures of deception. Although the underlying psychological bases for polygraph and ocular-motor tests for deception may differ, it is possible that motivation to appear innocent will affect examinees' behavior. To test for effects of motivation in Experiment 2, half of the participants were motivated with a $30 bonus (as in Experiment 1), and for the remaining participants, this bonus was decreased to $1.

Statement content in Experiment 1 was counterbalanced for true/false responses, and for positive or negative wording. In some cases, the combination of these factors resulted in simple, easy-to-understand statements (e.g., *I did not take cash from the secretary's purse.*), and in other cases, it resulted in complex and difficult-to-understand statements (e.g., *The claim that I did not take the $20 from the purse is incorrect.*). It may be that the cognitive effort required to comprehend and respond correctly to the complex statements confounded and thus diminished our ability to distinguish among the groups. If so, differences between deceptive and truthful individuals should be more pronounced when only simple statements are included in the test. Alternatively, Vrij and colleagues found that by increasing the cognitive load associated with a recall task, they were better able to distinguish between deceptive and nondeceptive responses (Vrij et al., 2007; 2009). If those findings apply to the ocular-motor test, differences between deceptive and truthful participants should be greater when the test contains complex statements than when it contains only simple statements. To test these two accounts, half of our participants were presented with a mixed set of simple and complex statements, and half were presented with only simple statements.

Because the general pattern of responses to the committed-crime and non-committed-crime statements was consistent across the cash and card groups in Experiment 1, we simplified the design and used only one crime in Experiment 2; participants were either innocent or guilty of stealing $20 from a secretary's purse. However, in order to retain a non-neutral comparison condition for statements about the committed crime, participants were led to believe that some participants had stolen an exam from a professor's office, and they would be questioned about that crime as well. Thus, three statement types were presented to participants: neutral statements, statements about a crime they committed, and statements about a crime they did not commit.

There were three additional differences between Experiment 1 and Experiment 2. First, prior research has shown that eyeblinks are suppressed under conditions of cognitive load and occur when focused attention ends (Siegle, Ichikawa, & Steinhauer, 2008; Stern, Walrath, & Goldstein, 1984). In our test, eyeblinks may be suppressed while individuals read and respond to statements answered deceptively compared with statements answered truthfully. To test this prediction, we used a different eye tracker that allowed for the measurement of eyeblinks. If deceptive individuals suppress eyeblinks while they read and respond deceptively to a statement, they also may recover on the subsequent statement. That is, they may blink more on statements that follow a deceptive response than on statements that follow a truthful response (cf. Baker, Goldstein, & Stern, 1992a). We recorded eyeblinks for each statement, and for the subsequent statement, as a function of both group and statement type. We also included participant gender as a variable in our design and analyses, because the group sizes were too small to test for effects of gender in Experiment 1. Finally, we increased the number of repetitions of statements to five.

Method

Design. Equal numbers of male and female participants were randomly assigned to one of eight cells in a $2 \times 2 \times 2 \times 2 \times (3 \times 5)$ mixed design. The between-subjects variables were guilt (guilty vs. innocent), motivation ($30 vs. $1), statement difficulty (mix of simple and complex statements vs. simple statements only), and gender (male vs. female). The two within-subject factors were statement type (neutral, cash, and exam) and repetition (five repetitions of each statement). The dependent variables were change in pupil diameter, response time, response errors, number of fixations, first pass reading time, second pass reading time, blink rate, and next item blink rate.

Participants. One hundred thirty-six university students and staff were recruited via fliers posted on campus. Of these 136 participants, eight chose not to participate after learning of their experimental condition, five did not follow instructions, and two were lost because of experimenter error. An additional nine participants' (6.6%) eye movements were not able to be properly tracked. The remaining 56 male and 56 female participants ranged in age from 18–67 years ($M = 25.90$, $SD = 7.18$), were predominantly Caucasian (79.5%), single (72.3%), and students (85.7%).

Apparatus. An Arrington ViewPoint Eye Tracker (Arrington Research, Inc, Scottsdale, AZ) was used to record eye movements and pupil diameter at 30 Hz. The eye tracker was affixed to a pair of lens-less plastic safety goggles. Viewing was binocular, but eye movement and pupil diameter were recorded only from the right eye. Test statements were presented in a single line in the center of a 19-inch NEC MultiSync FE950 + flat screen CRT monitor beginning on the left side of the screen. The monitor was positioned approximately 50.8 cm from the participant's eyes.

Materials. Participants responded to 48 statements on five occasions. Sixteen statements pertained to the theft of the $20, 16 pertained to the theft of the exam, and 16 were neutral. There were equal numbers of correct True and False statements within each type. Half of the participants received a mixed set of statements that contained both simple and complex statements, and half received only simple statements. Complex statements included a relative clause (e.g., "The 20 dollars *that was in the office* is not in my possession.").

Mixed difficulty statements were longer than simple statements, $F(1, 95) = 6.15$, $p = .02$. Within the mixed difficulty set, the mean lengths in characters for the neutral, cash, and exam statements were 48.06 ($SD = 9.33$), 55 ($SD = 10.21$), and 53.81 ($SD = 11.63$), respectively. The neutral statements were marginally shorter than the cash statements, $t(30) = 2.01$, $p = .05$, but no other contrasts were significant, $ps > .13$. Within the set of simple statements, the mean lengths in characters for the neutral, cash, and exam statements were 44.31 ($SD = 9.39$), 49.5 ($SD = 7.44$), and 48.38 ($SD = 8.42$), respectively. Again, the neutral statements were marginally shorter than the cash statements, $t(30) = 1.73$, $p = .09$, but no other contrasts were significant, $ps > .2$.

Measures. Experiment 2 included all of the outcome measures in Experiment 1 and added two measures of blink rate: *Blink rate* was the number of blinks per second for each statement, and *next statement blink rate* was the number of blinks per second for the statement that followed.

Procedures. The procedures were the same as those in Experiment 1 with the following exceptions. Participants were told that some individuals had to download an exam from a professor's computer onto a disk, but in actuality, no one committed that crime. Rather than using the mouse to answer true or false to test statements, participants pressed the '1' or '3' key on the keypad of the computer keyboard, and these keys were labeled T and F, respectively. Participants completed five repetitions of the test statements rather than three. Last, participants completed a post-session questionnaire to rate the importance of the monetary bonus on a scale from 1 (*not at all*) to 5 (*extremely*).

Results

A main effect of guilt and a Statement Type × Guilt interaction are two sources of diagnostic information that may be used to decide if the individual was deceptive on the test. A goal of Experiment 2 was to determine whether either of these effects was moderated by gender, motivation to pass the test, statement complexity, or repetitions. Given the multifactor nature of the design for Experiment 2, each RMANOVA had over 60 sources of variance. To simplify the presentation of the results, only tests of the main effect of guilt, the interaction of Statement Type × Guilt, and three-way interactions that included statement type and guilt are reported. A complete listing of significant effects may be found in Webb (2008). Descriptive statistics for response time, number of fixations, first- and second-pass duration, blink rate, and next statement blink rate are presented in Table 3.

Manipulation check. The monetary bonus was rated as more important to participants promised $30 for a truthful outcome ($M = 2.87$, $SE = .11$) than to participants promised only $1 for a truthful outcome ($M = 1.75$, $SE = .11$), $F(1, 96) = 49.61$, partial $\eta^2 = .34$. Participants' self-reports were consistent with our intention to manipulate levels of motivation to pass the test.

Gender effects. Gender moderated the Statement Type × Guilt interaction on blink rate (partial $\eta^2 = .04$). Because the effect was small and gender did not interact with guilt or moderate

Table 3
Mean Response Time, Number of Fixations, First- and Second-Pass Duration, Blink Rate, and Next Statement Blink Rate as a Function of Group and Statement Type in Experiment 2

	Guilty-cash						Innocent					
	Cash		Exam		Neutral		Cash		Exam		Neutral	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Response time	56.93	1.51	65.24	1.71	64.54	1.53	54.44	1.64	61.0	2.77	55.01	1.37
Number of fixations	16.30	3.63	18.45	4.16	18.40	3.74	16.77	4.5	17.18	4.95	16.46	3.6
First-pass duration	42.78	1.06	48.53	1.06	48.93	1.08	43.13	1.31	44.14	1.35	44.07	1.16
Second-pass duration	4.19	4.19	5.61	.42	5.02	.45	3.23	.39	4.15	.39	2.5	.24
Blink rate	.06	.05	.07	.05	.07	.05	.07	.05	.07	.05	.07	.05
Next statement blink rate	.07	.05	.07	.05	.06	.05	.07	.05	.07	.05	.07	.05

Note. Number of fixations are reported per character, whereas response time, first-pass duration, and second-pass duration are reported in ms per character. Blink rate and next item blink rate are presented in blinks per second.

the Statement Type × Guilt interaction for any other outcome measure, it is not discussed further.

Repetition effects. RMANOVA revealed no interactions of guilt with repetitions or interactions of Guilt by Statement Type × Repetition for any of the outcome measures. As a result, all subsequent analyses are based on means computed across repetitions.

Response errors. One innocent participant scored 7.01 standard deviations from the grand mean error rate and was dropped from the analysis. The main effect of guilt on response errors was significant, $F(1, 95) = 5.73$, partial $\eta^2 = .06$. Guilty participants had a higher error rate ($M = .06$, $SD = .01$) than did innocent participants ($M = .05$, $SD = .01$). The effect of guilt interacted with statement difficulty, $F(1, 95) = 4.37$, partial $\eta^2 = .04$. Guilty participants made more mistakes with simple statements than with mixed statements, whereas innocent participants made fewer mistakes with simple statements than with mixed statements. The Statement Type × Guilt interaction was not significant.

Response time. Although the main effect of guilt was not significant for response time, $F(1, 96) = 3.25$, $p = .07$, there was a small but significant Statement Type × Guilt interaction, $F(1.47, 141.29) = 3.57$, partial $\eta^2 = .04$. The interaction was due primarily to the finding that guilty participants responded more quickly when they lied to statements concerning the theft of the $20 than when they answered truthfully to neutral statements or statements about the other crime.

Pupil diameter. Changes in pupil diameter following the presentation of cash, exam, and neutral statements are plotted at 10 Hz in Figure 2 for guilty and innocent participants, respectively. The Statement Type × Guilt interaction was significant, $F(1.87, 179.18) = 20.62$, partial $\eta^2 = .18$, as was the three-way interaction of Time × Statement Type × Guilt, $F(4.24, 406.80) = 13.13$, partial $\eta^2 = .12$. For guilty participants, the pupil responses to cash statements were greater than responses to exam statements, $F(1, 55) = 20.13$, partial $\eta^2 = .27$. Although no difference between responses to cash statements and exam statements was expected for innocent participants, pupil responses to exam statements were greater than responses to cash statements, $F(1, 55) = 11.54$, partial $\eta^2 = .17$. The difference between pupil responses to crime-related and neutral statements was significant for guilty, $F(1, 55) = 118.71$, partial $\eta^2 = .68$, and innocent participants, $F(1, 55) = 65.36$, partial $\eta^2 = .66$.

Reading measures. Although the main effect of guilt on number of fixations was not significant, there was a significant Statement Type × Guilt interaction, $F(1.81, 174.65) = 19.86$, partial $\eta^2 = .17$. This interaction again was because of deceptive individuals making fewer fixations while reading statements about the crime they committed than statements about another crime or neutral content. The Statement Type × Guilt × Motivation interaction also was significant, $F(1.82, 174.65) = 3.34$, partial $\eta^2 = .03$. For number of fixations, there was less difference between the guilty and innocent participants in the low motivation condition than the high motivation condition. It is interesting to note that the incentive manipulation had a greater effect on innocent participants than guilty participants. The innocent participants in the low motivation condition made more fixations than innocent participants in the high motivation condition. In addition, mixed statements were fixated more often than simple statements, $F(1, 96) = 4.60$, partial $\eta^2 = .05$. However, statement difficulty did not interact with guilt or moderate the Statement Type × Guilt interaction.

Although there was no significant main effect of guilt on first-pass duration, the Statement Type × Guilt interaction was significant, $F(1.95, 187.21) = 18.71$, partial $\eta^2 = .16$. Guilty participants had shorter first-pass durations when they were deceptive than when they were truthful. For innocent participants, first-pass durations were similar across statement types.

The main effect of guilt on second-pass durations was significant, $F(1, 96) = 4.93$, partial $\eta^2 = .05$, as was the Statement Type × Guilt interaction, $F(2, 192) = 6.24$, partial $\eta^2 = .06$. Across statement types, guilty participants did more rereading than did innocent participants. Guilty participants did less rereading when they were deceptive than when they were truthful, whereas innocent participants' rereading durations varied less across statement types.

As predicted, blink rates were lower for guilty participants when they read statements about the crime they had committed, $F(2, 192) = 3.12$, but the effect was small, partial $\eta^2 = .03$. Effects of interest on next statement blink rate were not significant.

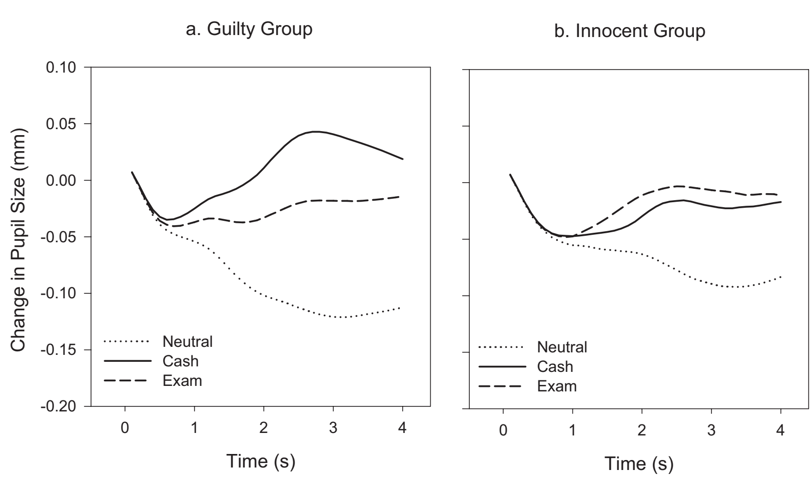

Figure 2. Mean evoked pupil response as a function of group and statement type in Experiment 2.

Bivariate and discriminant analyses. A group membership indicator was created that distinguished between guilty (+1) and innocent (−1) participants. The indicator variable was correlated with measurements for neutral statements, the difference between crime-related and neutral statements, and the difference between cash and exam questions. The correlations for each ocular-motor measure are presented in Table 4. Twelve of 18 measures were significantly correlated with group membership. The most diagnostic measure was the difference between the mean first-pass duration for cash and exam statements ($r = .53$).

Stepwise linear discriminant analysis was performed with the 12 measures that correlated with group membership. Four variables were selected for the discriminant function. They included the difference between first-pass duration for cash and exam statements, the difference between pupil responses to cash and exam statements, the difference between second-pass durations for crime and neutral statements, and the difference between next statement blink rates for crime and neutral statements. The function correctly classified 46 of the 56 guilty participants (82.2%) and 50 of the 56 innocent participants (89.3%). The same results were obtained with jackknifed classifications (85.7% correct overall). When the four variables selected in Experiment 1 were used to classify the cases in Experiment 2, accuracy dropped by about 5% to 78.6% for guilty participants and 82.1% for innocent participants (80.4% overall).

Discussion

The patterns of results from Experiment 2 replicate and extend those of Experiment 1. Compared with innocent participants, guilty participants had overall longer response times, more fixations, and longer reading and rereading times. Guilty participants also had larger increases in pupil diameter, made fewer fixations, spent less time reading and rereading statements, and made fewer eyeblinks while they read statements answered deceptively than statements answered truthfully.

We also tested whether motivation influences the ability to distinguish between guilty and innocent participants. High-motivated innocent participants made fewer fixations than low-motivated innocent participants. Conversely, there was no difference between high- and low-motivated guilty groups in number of fixations. This suggests that guilty participants may be intrinsically motivated to avoid detection, whereas innocent participants invested more effort to earn a large reward than a small one.

The only evidence that statement difficulty interacted with group and statement type was on response accuracy. Contrary to predictions based on the findings of Vrij et al. (2007, 2009), the difference between guilty and innocent participants was larger when the test contained only simple statements than when it contained a mix of simple and complex statements.

Table 4
Correlations Between Ocular-Motor Measures and Group Membership Indicators in Experiment 2

Variable	Guilty-innocent
Neutral statements	
Pupil response (AUC)	.10
Number of fixations	$-.26^{**}$
First-pass duration	$-.21^{*}$
Second-pass duration	$-.34^{**}$
Blink rate	.03
Next statement blink rate	.09
Crime vs. neutral statements	
Pupil response (AUC)	$-.29^{**}$
Number of fixations	$.32^{**}$
First-pass duration	$.25^{**}$
Second-pass duration	$-.28^{**}$
Blink rate	$.25^{**}$
Next statement blink rate	$-.22^{*}$
Cash vs. exam statements	
Pupil response (AUC)	$-.46^{**}$
Number of fixations	$.53^{**}$
First-pass duration	$.53^{**}$
Second-pass duration	.12
Blink rate	.07
Next statement blink rate	−.08

Note. AUC = area under the response curve.
$^{*} p < .05$. $^{**} p < .01$.

General Discussion

Our purpose was to assess a new test for detecting deception that was based on the assumption that deception is more cognitively demanding than being truthful. Guilty and innocent participants in two separate experiments responded either true or false to statements presented on a computer. The statements were repeated in different orders three or five times. Measures of response accuracy, response time, pupil diameter, reading behavior, and blink rates were obtained to assess hypothesized differences in cognitive workload. The results of the present research generally were consistent with the cognitive workload hypothesis.

Guilty participants exhibited clear differences from innocent participants. Guilty participants, given syntactically simple test statements, made more errors than did innocent participants, took longer to respond, made more fixations on the text, had longer reading times, and in Experiment 2, also had longer rereading times. These findings are consistent with and extend prior research on the effects of deception on response errors and response time (Baker et al., 1992b; Crosland, 1929; Seymour et al., 2000; Vendemia et al., 2005; Walczyk et al., 2003, 2009). In addition, guilty participants blinked significantly less often as they processed statements answered deceptively than when they processed statements answered truthfully (see Siegle et al., 2008; Stern et al., 1984). Although previous researchers reported differences in repetition effects for guilty and innocent individuals in response time (Baker et al., 1992b) and response accuracy (Vendemia et al., 2005), there was no evidence of these effects in either Experiment 1 or 2.

Consistent with previous research, guilty participants in both experiments showed greater increases in pupil diameter for statements answered deceptively than for statements answered truthfully (e.g., Berrien & Huntingdon, 1943; Bradley & Janisse, 1981; Dionisio et al., 2001; Heilveil, 1976; Kircher, Podlesny, Bernhardt, Bell, & Packard, 2000; Lubow & Fein, 1996; Webb, Honts, et al., 2009). The reading behaviors of guilty participants were more surprising, however. We had predicted that guilty participants would experience the greatest difficulty reading statements subsequently answered deceptively and that difficulty would be indicated by relatively long response times and first- and second-pass reading times. Instead, guilty participants responded faster, made fewer fixations, and spent less time reading and reading statements about the crime they committed than statements about another crime or neutral statements. Although these effects were not expected, they have since been replicated in two other independent groups of guilty participants in a mock crime experiment (Webb, Hacker, et al., 2009).

The pupil data and the guilt main effects in the reading data support the hypothesis that deception is more cognitively demanding than being truthful. However, the finding that guilty participants actually reread less when they encountered incriminating statements suggests that two cognitive processes are associated with deception. First, deception is a cognitively demanding task that requires planning, comparison, and execution of purposefully incorrect responses (Vrij, 2008). The extra vigilance required by deception results in increased cognitive workload, which is reflected in increased response errors, increased response time, and increased overall reading time. The second type of processing involved in deception is more strategic. All participants in our experiments were informed that they should respond as quickly and accurately as possible or they would appear guilty and fail the test. Deceptive participants appeared to heed this instruction when they encountered statements about the crime they committed. To avoid detection, they probably attempted to read these statements quickly, and they may have even deliberately suppressed rereading behaviors to avoid appearing deceptive. This finding is consistent with the view that participants can exert some conscious control over their reading behaviors to implement specific reading strategies (Hyona & Nurminen, 2006). The differential processing of incriminating test statements by guilty participants came at a cost, however. The extra effort required to control reading of statements answered deceptively may have contributed to the observed increase in pupil diameter for those statements. The idea that deception requires both caution and strategies is consistent with other theoretical accounts of deception (Johnson et al., 2005; Kircher, 1981; Steller, 1987) and research on cognitive strategies in general (Nelson & Narens, 1990). Our findings thus build on previous cognitive-based tests of deception to show that it is important to include not only general indices of processing difficulty (e.g., response time, pupil diameter, and number of fixations), but also measures that separate initial and delayed processing difficulty (e.g., first and second-pass rereading time, respectively).

When measures were combined in a discriminant function analysis, they yielded accuracy rates that are comparable to those of the polygraph (Krapohl, 2002). The overall classification accuracy rates were approximately 85% for Experiment 1 and 86% for Experiment 2. The accuracy achieved in Experiment 1 becomes even more salient when compared against the chance probability of a correct decision of 37.5% (vs. 50% in Experiment 2). Given that stepwise discriminant analysis capitalizes on chance, it is expected that the classification accuracy of these functions would drop if tested on a new sample of cases. When the discriminant function developed in Experiment 1 was applied to the data from Experiment 2, the classification accuracy dropped to approximately 80%.

Several factors could influence the accuracy of the ocular-motor test for deception. First, we found in Experiment 2 that innocent participants who were offered incentives to appear truthful responded faster and read and reread less, making them easier to discriminate from guilty participants. By providing adequate incentives for innocent individuals to pass the test, it may be possible to reduce the percentage of false positive decisions made by the test. Second, the eye tracking equipment could influence accuracy of the test. When we switched to an eye tracker that was simpler to calibrate and allowed for the calculation of blink rates, fewer participants were lost because of equipment problems and we gained a possibly valuable outcome measure of blink rate. Although it could be argued that the eye trackers used in both experiments were invasive, the attachment of a monitoring device to the examinee may add to the general arousal experienced, and thus to the general motivation to appear innocent during a test. If this hypothesis is correct, then use of such overt monitoring devices may not only be more effective than covert monitoring with more remote eye trackers, but it would also mitigate risks and ethical concerns of data collection from uninformed individuals. Finally, although tests that are based on the cognitive workload hypothesis may be predominantly cognitive, these tests probably also include an emotional component. Researchers have noted associations between the pupil response and emotional arousal, with larger pupil diameters associated with greater arousal (Bradley, Micolli, Escrig, & Lang, 2008; Partala & Surakka, 2003; Stern, Ray, & Quigley, 2001). The finding that innocent participants showed pupil enlargement in response to crime-relevant statements is consistent with the idea that threatening (i.e., potentially incriminating) stimuli evoke stronger autonomic responses than neutral stimuli even when responses are truthful (Horowitz et al., 1997). Future research should continue to evaluate the influence of these and other variables on the accuracy of the test.

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 measures in some field settings. A preliminary combination of response measures with pupil diameter, number of fixations, and first and second-pass reading times allowed for differential classification of guilty and innocent participants with accuracies at least as high as those achieved by screening polygraph and self-report techniques. Additional research is needed (a) to assess the relative contributions of cognitive and emotional processes to the effects observed on pupil size and reading behaviors, (b) to determine whether the effects obtained in these laboratory experiments generalize to field settings, and (c) to determine whether countermeasures may be used to defeat the ocular-motor test.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205,* 1289–1292. doi:10.1126/science.472746

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 997–1010. doi: 10.1037/0278-7393.25.4.997

Baker, L., Goldstein, R., & Stern, J. A. (1992a). *Saccadic eye movements in deception*: *Final report to the U.S. Government (Contract #91-P-0003).* St. Louis, MO: Washington University, Department of Psychology.

Baker, L., Stern, J. A., & Goldstein, R. (1992b). *The gaze control system and the detection of deception: Final report to the U.S. Government (Contract #90-F131400).* St. Louis, MO: Washington University, Department of Psychology.

Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science, 199,* 1216–1218. doi:10.1126/science.628837

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective.* New York, NY: Springer-Verlag Publishing. doi:10.1007/978-1-4612-3282-7

Berrien, F. K., & Huntington, G. H. (1943). An exploratory study of papillary responses during deception. *Journal of Experimental Psychology, 32,* 443–449. doi:10.1037/h0063488

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18,* 307–315. doi:10.1111/j.1469-8986.1981.tb03040.x

Bradley, M. M., Micolli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45,* 602–607. doi:10.1111/j.1469-8986.2008.00654.x

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language, 50,* 268–288. doi:10.1016/j.jml.2003.11.003

Crosland, H. R. (1929). *The psychological methods of word-association and reaction-time as tests of deception.* Eugene, OR: University of Oregon Publications.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbrook, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118. doi:10.1037/0033-2909.129.1.74

Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology, 38,* 205–211. doi:10.1111/1469-8986.3820205

Ellson, D. G., Davis, R. C., Saltzman, I. J., & Burke, C. J. (1952). A report of research on detection of deception. *Technical report prepared for Office of Naval Research (Contract N6onr-18011).* Bloomington, IN: Indiana University.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42,* 526–544. doi:10.1006/jmla.1999.2694

Heilveil, I. (1976). Deception and pupil size. *Journal of Clinical Psychology, 32,* 675–676. doi:10.1002/1097-4679(197607)32:3<675::AID-JCLP2270320340>3.0.CO;2-A

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143,* 1190–1192. doi:10.1126/science.143.3611.1190

Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of control questions in physiological detection of deception. *Psychophysiology, 34,* 108–115. doi:10.1111/j.1469-8986.1997.tb02421.x

Hyona, 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,* 31–50. doi:10.1348/000712605X53678

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

Johnson, R., Jr., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research, 24,* 386–404. doi:10.1016/j.cogbrainres.2005.02.011

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354. doi:10.1037/0033-295X.87.4.329

Just, M. A., & Carpenter, P. A. (1993). The intensity of dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47,* 310–339. doi:10.1037/h0078820

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154,* 1583–1585. doi:10.1126/science.154.3756.1583

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

Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12,* 79–90. doi:10.1007/BF01064275

Kircher, J. C., Podlesny, J. A., Bernhardt, P., Bell, B., & Packard, T. (2000). Blood pressure and pupil diameter measures of deception [abstract]. *Psychophysiology, 37,* S19.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291–302. doi:10.1037/0021-9010.73.2.291

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

Kristjansson, S. D., Kircher, J. C., & Webb, A. K. (2007). Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. *Psychophysiology, 44,* 728–736. doi:10.1111/j.1469-8986.2007.00544.x

Long, D. L., & Lea, R. B. (2005). Have we been searching for meaning in all the wrong places? Defining the "search after meaning" principle in comprehension. *Discourse Processes, 39,* 279–298.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied, 2,* 164–177. doi:10.1037/1076-898X.2.2.164

McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D. W., & Lucas, P. A. (1979). Toward the use of eye movements in the study of language processing. *Discourse Processes, 2,* 157–177. doi:10.1080/01638537909544463

Meesig, R., & Horvath, F. (1995). A national survey of practices, policies and evaluative comments on the use of pre-employment polygraph screening in police agencies in the United States. *Polygraph, 24,* 57–136.

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). New York, NY: Academic Press.

Office of Technology Assessment. (1990). The use of integrity tests for pre-employment screening (OTA-SET-442). Washington, DC: U.S. Government Printing Office.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59,* 185–198. doi:10.1016/S1071-5819(03)00017-X

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

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

Reinhard, G., Lachnit, H., & Konig, S. (2006). Tracking stimulus processing in Pavlovian pupillary conditioning. *Psychophysiology, 43,* 73–83. doi:10.1111/j.1469-8986.2006.00374.x

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49,* 787–829. doi:10.1111/j.1744-6570.1996.tb02450.x

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17,* 133–145. doi:10.1016/0093-934X(82)90010-4

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". *Journal of Applied Psychology, 85,* 30–37. doi:10.1037/0021-9010.85.1.30

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45,* 679–687. doi:10.1111/j.1469-8986.2008.00681.x

Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. In D. Friedman & G. Bruder (Eds.), *Psychophysiology and experimental psychopathology: A tribute to Samuel Sutton* (Vol.658, pp. 182–204). New York, NY: Annals of the New York Academy of Sciences.

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

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording* (2nd ed.). New York, NY: Oxford University Press.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology, 21,* 22–33. doi:10.1111/j.1469-8986.1984.tb02312.x

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. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). West Sussex, England: John Wiley & Sons Ltd.

Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior, 33,* 159–166. doi:10.1007/s10979-008-9143-y

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior, 31,* 599–518. doi:10.1007/s10979-006-9066-4

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. doi:10.1007/s10869-009-9090-8

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology, 17,* 755–774. doi:10.1002/acp.914

Webb, A. K. (2008). *Effects of motivation and item difficulty on oculomotor and behavioral measures of deception.* University of Utah. Unpublished doctoral dissertation.

Webb, A. K., Hacker, D. J., Osher, D., Cook, A. E., Woltz, D. J., Kristjansson, S., & Kircher, J. C. (2009). Eye movements and pupil size reveal deception in computer administered questionnaires. In D. D. Schmorrow, I. V. Estabrooke, & M. Grootjen (Eds.), *Foundations of augmented cognition: Neuroergonomics and operational neuroscience* (pp. 553–562). Berlin, Germany: Springer-Verlag.

Webb, A. K. Honts, C. R., Bernhardt, P., Kircher, J. C., & Cook, A. E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminological Psychology, 14,* 279–292. doi:10.1348/135532508X398602

Received October 21, 2011
Revision received March 7, 2012
Accepted March 23, 2012 ■