Hidden Hiding Places: Detecting Concealed Information in a Mock Crime Scenario Using Pupillometry and RSVP

Tobias Kroner

S4257510

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 2223_1b_02

Supervisor: Dr. Robbert van der Mijn

Second evaluator: M.Sc. Ana Vilotijević

In collaboration with: Charlotte Bosch, Anouk Feijen,

Danie Hettema, Teresa Nguyen, and Janette Possul

April 14, 2023

Abstract

Concealed information testing (CIT) makes use of involuntary physiological reactions to detect hidden knowledge and is relevant to forensic applications. To curb common countermeasures, studies have used rapid serial visual presentation (RSVP), a method where stimuli are presented on the fringe of awareness. Since pupil dilation indicates stimulus salience and recognition, a novel and practical research paradigm for deception detection is the combination of RSVP with cognitive pupillometry. Measuring pupil size in a mock-crime scenario, previous studies have found participants' names to elicit a clear dilation reaction during an RSVP task, even when concealed and task-irrelevant. With a convenience sample of 57 Dutch psychology students, our study aimed to conceptually replicate these findings with another category of stimuli. Next to mean pupil size we also used mean rate of pupil size change (derivative) as additional measure. Using location names that were made salient during the experiment, we found a significant difference between a task-irrelevant real hiding location and controls for mean pupil derivative but not for size on a group level (in the predetermined time-window of 600-900 ms after stimulus onset). Classification analysis with these features revealed a hit-rate of 16% and 4% on an individual level respectively. To increase discriminability and reliability of this CIT method, future research could combine several ocular measures, use autobiographical stimuli, and enlarge the time-window to 600-1000ms.

Keywords: CIT, cognitive pupillometry, pupil dilation, RSVP, salience

Hidden Hiding Places: Detecting Concealed Information in a Mock Crime Scenario Using Pupillometry and RSVP

Concealed information is any knowledge someone tries to hide from others in their mind. Deception is a common, and perhaps necessary, feature of human conduct that may occur in a variety of different situations (Saxe, 1991). While everyone lies on occasion, in criminal contexts it can be of societal importance to figure out if a suspect is guilty because they possess crime-relevant information. To detect this hidden knowledge in forensic settings, researchers devised Concealed Information Testing (CIT), akin to what is colloquially known as polygraph testing or “lie detection”. In general, responses that occur when trying to deceive are different from those observed with truthful behavior (Block et al., 1952). Traditionally called the guilty knowledge test (Lykken, 1959), CIT is based on the assumption that if participants possess information related to a specific criminal activity, crime-relevant stimuli will elicit implicit responses (Patrick, 2011). Since its inception as subject of scientific study in the early 1940s (Geldreich, 1941), several different psychophysiological measures have been employed to uncover deception. These involuntary reactions include autonomic nervous system responses such as electrodermal activity (MacLaren, 2001), respiratory measures (Elaad & Ben-Shakhar, 2009), and cardiovascular measures (Gamer, 2011), and more recently stimulus-evoked P300 brain potentials (Rosenfeld, 2011). Measurements in all experiments reflect reflexive changes in arousal levels, which can be compared to reactions to neutral stimuli to assess whether participants possess hidden knowledge. Studies with mock-crime paradigms were described to have yielded promising results in an influential early review (Ben-Shakhar & Elaad, 2003). However, no single measure or test has been proven to be consistently reliable (Meijer et al., 2016).

Problematic for the validity of CIT is the use of countermeasures that can be purposefully employed by uncooperative participants to shroud their responses and invalidate the results (Peth et al., 2016). In P300-based tests for example, such countermeasures may consist of imagining exciting situations or pressing fingers together, resulting in physiological responses to the investigated stimuli that are indistinguishable from the ones to others (Rosenfeld et al., 2004). An effective counter to these ploys is rapid serial visual presentation (RSVP; Bowman et al., 2013). RSVP was originally conceived to study attention and working memory, such as with visual search in a word list (Lawrence, 1971) and conceptual short-term memory for pictures (Potter, 1976). In an RSVP task, sequences of stimuli are subsequently displayed in the same location, with each stimulus being shown for about 100 ms (Broadbent & Broadbent, 1987). The participant is usually tasked with identifying or detecting a salient *target* stimulus within a rapidly presented stream of non-target stimuli (the *distractors*). A task-irrelevant stimulus can be assigned as the *probe* and used to investigate the possession of concealed information. The probe may, for example, be chosen to be a stimulus that is relevant to a (mock-) crime. By comparing the response to the probe to the response to a non-salient control stimulus (the *irrelevant*) one may be able to detect concealed information. While participants are not instructed to search for the probe, it can still elicit an involuntary bodily response due to its intrinsic salience, even when it is not consciously perceived (Bowman et al., 2014). In fact, most items in an RSVP stream may not be consciously perceived by the participant. At a sufficiently high speed of presentation, information processing happens too fast for top-down cognitive control to be able to regulate the associated bodily response when encountering subliminally salient stimuli (Chen et al., 2022). Reflexive changes in participants' psychophysiological responses are thus exploited by presenting stimuli on the fringe of awareness, making this research paradigm highly robust against countermeasures (Bowman et al., 2014).

RSVP was first developed as a paradigm for CIT by Bowman et al. (2013). The researchers measured P300 event-related potentials (ERPs) during a fake-name search task, using electroencephalography (EEG) to detect concealed identity information. Participants were instructed to pretend that another name was theirs while looking at RSVP streams containing 15 names. Next to 14 distractors the streams each contained one critical item, namely either a participant's assumed fake name (the *target*), their real name (the *probe*), or one of two preselected unfamiliar control names (the *irrelevant-1/2*). For each trial participants had to indicate whether or not they saw their name. If it was the fake name they saw, they were instructed to answer 'yes'. The researchers found that probe trials, which contained the participants' real names and thus had to be answered 'no', yielded significantly increased P300 ERPs when compared to irrelevant trials. This finding held even when participants were expressly told to conceal responses to their own names in different ways. Versions of this subliminal salience search paradigm have recently been successfully applied to email addresses (Harris et al., 2020) and famous names (Alsufyani et al., 2021). A significant difference between probe and irrelevant ERPs on an individual level was found in 64% and 86% of participants respectively. While being quite reliable, most CIT with RSVP relies on expensive and impractical brain-imaging techniques such as EEG. To enable more widespread use outside the laboratory, eye movements may be a more feasible and effective independent measure.

Eye movements such as pupil dilation, blinking and (micro-) saccades reflect a variety of cognitive processes. Among them are visual attention (Binda et al., 2014) and task-directed cognitive control (Eckstein et al., 2017), which are both relevant for CIT and RSVP. Pupils dilate in response to increased cognitive activity, such as recognition (Otero et al., 2011), emotional arousal (Bradley et al., 2008), or when encountering salient and unexpected stimuli that capture one's attention (Mathôt, 2018). Additionally, a constricted pupil may reflect inhibition of the processing of irrelevant distractors (Rondeel et al., 2015). However, the involuntary dilation in response to salient stimuli also occurs when these stimuli are task-irrelevant (Gilzenrat et al., 2010). Dilations evoked by motivationally significant stimuli are accompanied by a P300 ERP, which are both connected to the activation of phasic responses in the locus coeruleus-norepinephrine (LC-NE) system (Nieuwenhuis et al., 2011). This suggests that pupil size may be a useful measure in RSVP-based CIT. An early CIT study by Lubow and Fein (1996) already showed pupil size to be an effective measure in a mock-crime scenario. The researchers classified participants, randomly assigned to be part of a mock-crime or not, as guilty or innocent with a 50%-70% and 100% hit rate respectively by comparing pupil sizes in response to different items relevant or irrelevant to the hypothetical crime. This was achieved by assigning detection scores to participants if their pupil size in response to the crime-related item within a set was larger than to other items (Lykken, 1959). Later studies, including ours, focus on quantitative differences across trials rather than qualitative differences within trials for classification.

In a recent study, Chen et al. (2022) used the task design of Bowman et al. (2013) for a conceptual replication of their RSVP paradigm to investigate whether pupillometry could be used to detect concealed identity information, namely participants' (real) hidden names. Participants adopted a fake name for which they had to search in an RSVP task, with streams containing either this fake name (the *target*), their real name (the *probe*), or a control name (the *irrelevant*), and distractors. Comparing probe and irrelevant, the researchers found task-irrelevant (yet salient) real names to elicit a detectable difference in pupil dilation on the group level, but not on the individual level for most participants. The differences were especially large in the first half of the experiment and declined in effect size over time, suggesting a possible modulation by habituation, fatigue, or training over the course of the experimental session. The present study seeks to conceptually replicate Chen et al. (2022) with slight alterations in line with their proposed changes and with a different category of stimuli.

In order to use this paradigm in more practical settings, other types of probes should be considered. Our study uses location names as stimuli in an RSVP paradigm to detect hidden hiding places with pupillometry. Since location names are thought to be intuitively less salient than real names, this study aims to increase their salience before the experiment begins. In the current subliminal salience study, participants take part in a mock-crime scenario in which they are asked to imagine hiding a crime-related object at a location of their choosing from a predetermined list (the *probe*). They then are given another location (the *target*) at which they are instructed to pretend that they hid the object in order to “beat” the experiment. In the following RSVP task, word streams containing various location names are displayed, each containing either the probe, the target, or one of two irrelevant control locations (the *irrelevant-1/2*) among distractors. Participants will be instructed to indicate whether they saw the target location in each RSVP stream. We hypothesize that participants’ pupils will dilate more and increase in size at a faster rate for the probe trials than for irrelevant ones. With this evidence, measured via pupillometry, we aim to detect whether participants conceal information about their hiding location by comparing mean pupil size and slope measures between probe and irrelevant trials on a group and an individual level. Participants will be classified as guilty, innocent, or undetermined based on those features. Our research question is thus: Can concealed location information be detected in a mock crime scenario via RSVP and pupillometry?

Method

Participants

Fifty-seven participants (45 female, 12 male) took part in this study. Following the exclusion criteria described in the preprocessing section, eight participants were excluded, leaving 49 participants (37 female, 12 male) aged 17–35 (M = 20.7, SD = 3.1). There were five left-handed and 44 right-handed participants. All participants were first-year psychology students at the University of Groningen and received study credits as compensation, which are a requirement for passing the course. Participants were native Dutch speakers with normal or corrected-to-normal vision, and no self-reported dyslexia, however, three participants did not fill out the dyslexia question. Eye makeup, glasses, and contact lenses were removed prior to the experiment if they affected the apparatus. Participants gave their digital informed consent before participating in the study, which was approved by the ethical committee of the Psychology Department of the University of Groningen (approval number: PSY-2223-S-0166). The experiment was conducted in line with the recommendations of the World Medical Association Declaration of Helsinki (2013). After the experiment, oral debriefing was provided to all participants.

Apparatus & Stimuli

The experiment was conducted in a laboratory at the Behavioral and Social Sciences faculty building of the University of Groningen. Participants were seated at a height-adjustable desk in front of a 27″ LCD Iiyama PL2773H monitor with a display resolution of 1280 × 720 pixels and a refresh rate of 100 Hz. They placed their head on a chin rest that was attached to the desk at a distance of approximately 71 cm from the screen. The height of the chin rest was adjusted for each participant individually and the eye tracker was calibrated to the participants' eyes. Using PyGaze (Dalmaijer et al., 2014), an EyeLink 1000 eye tracker was used to record pupil size throughout the whole procedure at a rate of 1000 Hz. Stimuli were presented on a second computer using OpenSesame 3.3.14 (Mathôt et al., 2012) running on Windows 10 Enterprise. The set of Dutch locations (villages, towns, and cities) used as stimuli were taken from the Metatopos (Klein, 2022) database. Of the 2500 locations, 1395 were filtered out if they consisted of more than one word, contained diacritical marks, or were salient (e.g., capitals of the provinces; see Open Science Framework (OSF) [https://osf.io/q5cua/?view_only=4e9e63fabe394fb2a76206272ac113d1] for the full list of removed and used locations), leaving 1105 for the experiment. Location names had to consist of a minimum of three and a maximum of eight letters. All name stimuli were padded on both sides with '+' and '#' characters to even out their length, resulting in strings of eleven characters. They were presented starting with a capital letter, in the center of the screen, in a white (RGB: 255, 255, 255), Courier mono-spaced font, and on a dark background (RGB: 40, 40, 40). Fixation dots were of the same color as the stimuli. The visual angle for each stimulus was 9.26° in width and 1.60° in height. The illuminance of the ambient light was 40 lux, as measured from the perspective of the participant in front of the screen. Other materials used for this experiment included a physical map of 15 possible hiding locations (see Appendix A) on an A4 paper in black and white and a red Stabilo Pen marker.

Design

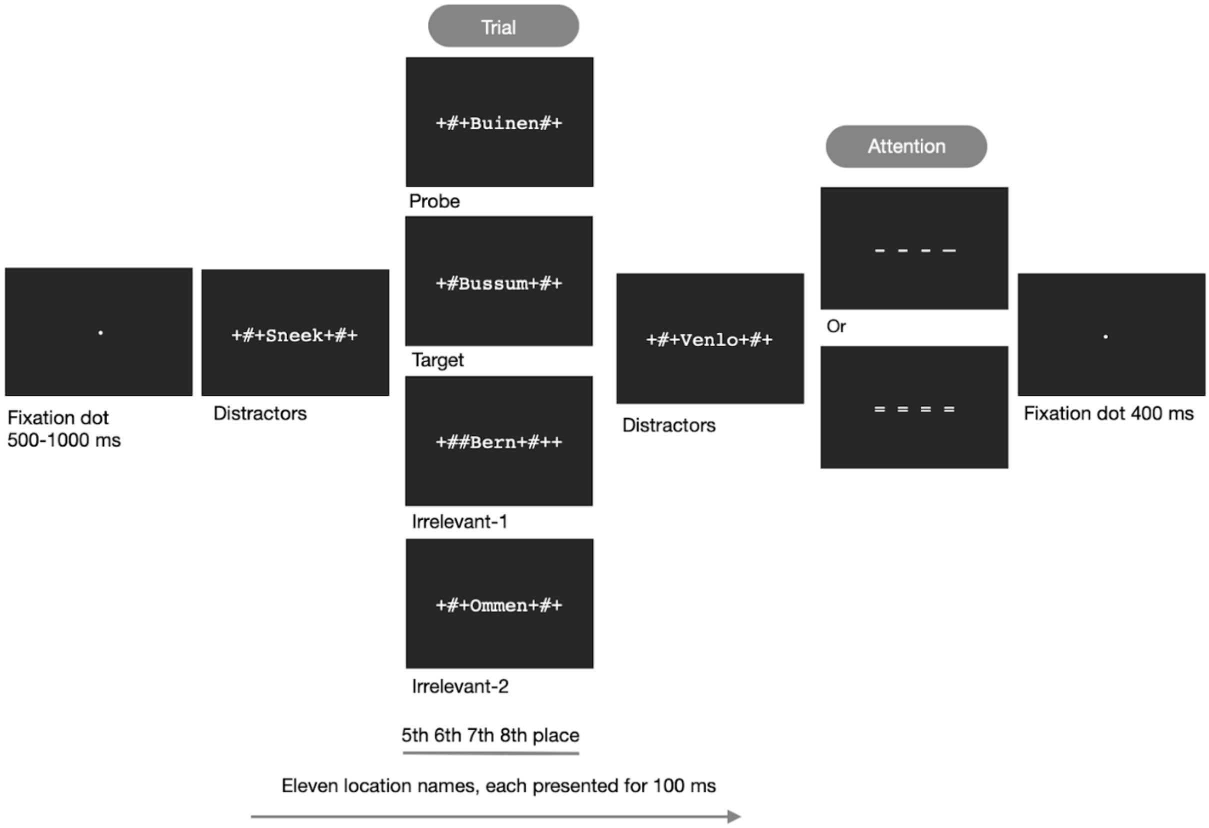
We used a within-participant experimental design. As part of a mock-crime scenario, an RSVP task was used, with a list of Dutch locations as stimuli (Klein, 2022; see OSF). From the list of 1105 Dutch locations, 15 were randomly chosen to be the potential real hiding location (probes) and removed from the rest of the word list. The list of potential probes was the same for each participant. To limit conscious or unconscious search strategies, such as focusing only on the first letter of the stimuli in the RSVP stream, the fake hiding location (target) was sampled from the list of locations with the same starting letter as the probe (Harris et al., 2021). The first control item (irrelevant-1) was also sampled from the locations that share the same starting letter as the probe and target. For the second control item (irrelevant-2), a new location name was chosen randomly for each trial. Irrelevant-1 and irrelevant-2 were unknown to the participant. The distractor list was constructed by removing the target, irrelevant-1, and irrelevant-2 from the location list, as were the locations starting with the same letter as probe, target, and irrelevant-1. The participant's hometown was also removed from the distractor list to control for its naturally high salience.

The experiment consisted of a practice block of 10 trials and an experimental block of 144 trials, summing up to a total of 154 trials per participant. The practice block was identical to the experimental block; however, it did not include the probe to avoid habituation as suggested by Chen et al. (2022). The experimental block consisted of trials with either probe, target, irrelevant-1, or irrelevant-2 at a ratio of 1:3:1:1 respectively to ensure an equal presentation of target and non-target stimuli.

Each trial started with a fixation dot that was presented for 750 ms with a 250 ms jitter during which a baseline pupil size was established. After the fixation dot, an RSVP stream of 11 items was shown with each stimulus being presented for 100 ms. The critical items (probe, target, irrelevant-1, irrelevant-2) presented in the experimental trials were randomly placed at the 5th, 6th, 7th, or 8th position in the RSVP stream (Figure 1). The 5th position was used as the first possible critical item position to accommodate for the pupil adjustment period at the start of the RSVP task. Having the last possible critical item at the 8th position allows time for the pupil to react after the critical item and before the end of the last fixation dot (Chen et al, 2022). Each stream ended with either equal signs (= = = = = = = = = =) or dashes (- - - - - - - - -) for 100 ms, followed by a fixation dot for 400 ms to allow for measurement of pupil size change after the end of the RSVP stream. To ensure that participants paid attention after each trial, the participants answered whether they saw equal signs or dashes and if they saw the target. The participant indicated their answer by pressing “C” if they saw the target and “M” if they did not see the target; response mapping was counterbalanced between participants. To keep the participants engaged, they either received or lost 5 or 10 points for (in)correctly identifying equal signs or dashes, and the target respectively. Points were granted or subtracted directly after giving the answers and accumulated points were shown during feedback.

Figure 1

RSVP trial sequence



Note. In this example, Buinen is the hiding location the participant chose, Bussum is the fake hiding location they were given, Bern and Ommen are the control locations.

The RSVP task was preceded by a short reaction task to increase the salience of the probe. It started with a countdown from three to one after which either probe or another location, randomly chosen from the probe list, was presented. The participant’s task was to indicate whether or not they saw the probe by pressing either “C” or “M”, which was counterbalanced between participants. If the participants answered correctly within 500 ms, they collected 10 points and if they answered incorrectly or took longer than 500 ms, they lost 10 points. The reaction task consisted of 60 trials, with the ratio of the critical items (probe vs. distractor) being 1:1.

Procedure

The experiment was carried out in Dutch, except for the verbal instructions and explanations, which were given in either Dutch or English. Each participant read the information sheet about the study and after a brief explanation of the process of the experiment, the participants received the rest of the instructions digitally. After giving their digital informed consent, participants provided information about their sex, handedness, age, hometown, dyslexia, and visual acuity. Participants read a story as a part of a mock crime scenario, in which a friend asked them to hide an incriminating suitcase somewhere in the Netherlands (the full story can be found in the OSF). After that, they could choose a hiding location (probe) from the previously mentioned map.

In case these locations included the participant's hometown, they were instructed not to choose it as their hiding location. After seeing the map with possible locations on the screen, participants were also presented with a physical map on which they were instructed to circle their chosen location. Next, the participants indicated their selected location via multiple choice digitally. Probe selection was followed by the reaction task described earlier.

After the reaction task, the story continued: The participants were suspected by the police of being an accomplice to a crime their friend committed, and their knowledge of the crime was going to be tested using CIT. Participants received instructions for the RSVP task. After the RSVP practice block, the participant could – if necessary – ask further questions if the task was unclear, and then proceeded to the experimental block. During the experimental block, participants had a break after every 36 trials. Participants could take as much time as they needed for the breaks and were allowed to move their heads from the chin rest. After the experiment, we debriefed the participants about the objective of the study. The whole procedure took 30 to 45 minutes. The whole experiment is available in the OSF.

Data processing

The scripts that have been used for the experiment, processing, and analysis of the data are accessible in the OSF. This processing includes the removal of any missing or unusable data. Python (version 5.4.1) running via Anaconda was used for data analysis, with the python-eyelinkparser module (version 0.17.3; Mathôt, 2023) being used for (pre-) processing.

Analysis

Our hypotheses regarding differences on a group level were tested using linear mixed-effects regression (LMER) and analyzed with the statsmodels (version 0.14.0) package. With that, we determined whether the mean pupil size and derivative in the time window are significantly larger in the probe condition than in the irrelevant-2 condition. For the individual analyses, we conducted *t*-tests comparing each participant's mean pupil size and derivative in the probe condition to the irrelevant-2 condition. To test for discriminability, we classified participants as either guilty, innocent, or undetermined based on the features of mean pupil size and derivative using individual *t*-test results.

Preprocessing

First, the accuracy of responses to the prompt "If you saw [target], press [C/M]. If not, press [C/M]?" were assessed. Participants with an accuracy score below 50% in target trials were excluded from further analyses. Following the approach of Chen et al. (2022) and the recommendation of Mathôt & Vilotijević (2022), we used their algorithm to reconstruct pupil data that were missing due to blinks. Trials with over 20% of data missing were marked as bad trials. Participants with more than 10% bad trials were excluded.

We baselined pupil size to the average pupil size during 50 ms after the onset of the critical item (T1) from each trial (Wilschut & Mathôt, 2022). In each trial, the baseline pupil size was subtracted from subsequent pupil size measurements. Next, we time-locked the data to the presentation of T1. Mean pupil size and derivative were intended to be computed during the time window from 600-1000 ms based on the suggestion of Chen et al. (2022). However, due to a mistake of setting the fixation dot to 400 ms, the pupil tracing after the latest T1 position (8th) could not exceed 900 ms. Thus, we were limited to a time window of 600-900 ms. Pupil-size samples were downsampled to 100 Hz.

Group level analysis

An LMER model was estimated to investigate the difference in pupil size between the irrelevant-2 control condition and the probe, target, and irrelevant-1 conditions. Mean pupil size in the predefined time window was used as a dependent continuous variable, and the T1 condition was used as a categorical independent variable (fixed effect). The participant was used as a random factor.

The mean rate of change in pupil size was analyzed by calculating the derivative for each condition. Derivatives were then analyzed with LMER in the same way as was mean pupil size, using the same time window. This was to assess whether the differences in the rate of pupil size change can be explained by the T1 condition.

Individual analysis

Individual *t*-tests were computed for each participant for mean pupil size and derivative. We performed four *t*-tests for each participant: For both mean pupil size and derivative, the probe was compared to irrelevant-2, and irrelevant-1 to irrelevant-2. Probe and irrelevant-2 were compared to detect differences in mean pupil sizes/derivatives due to the possession of concealed information. Irrelevant-1 and irrelevant-2 were compared to detect differences due to presentation frequency.

Classification

To investigate the discriminability of our testing method, we conducted a classification analysis. We used the mean pupil size and pupil derivative in the window of

600-900 ms after T1 onset as predictors. The participant was marked as guilty if the *t*-test between probe and irrelevant-2 exceeded the critical threshold $t(48) = 1.677$, $p < .05$. Second, if the *t*-statistic was below the critical threshold, the participant was marked as innocent. We also checked the significance of the difference between irrelevant-1 and irrelevant-2 for each participant. If a participant marked as guilty showed a significant difference between irrelevant conditions, we marked this participant as undetermined.

Results

Task performance

Four participants were excluded due to low accuracy on the question of whether or not they saw the target. Additionally, three participants were excluded due to missing data. One participant was excluded due to both criteria. The remainder of the 49 participants had a mean accuracy score of 96.76% on the attention task and 84.68% on the search task, indicating above guessing rate accuracy.

Group level analysis

The mean pupil traces for the different combined T1 trials as well as pupil measures are summarized in Figure 2a and 2b for pupil size and in Figure 3a and 3b for pupil derivative. Specifically, Figure 2a and 3a illustrate the effects of the different conditions on the average pupil trace over time, while Figure 2b and 3b provide an overview of the overall differences between conditions.

Figure 2a

Traces of Mean Pupil Size

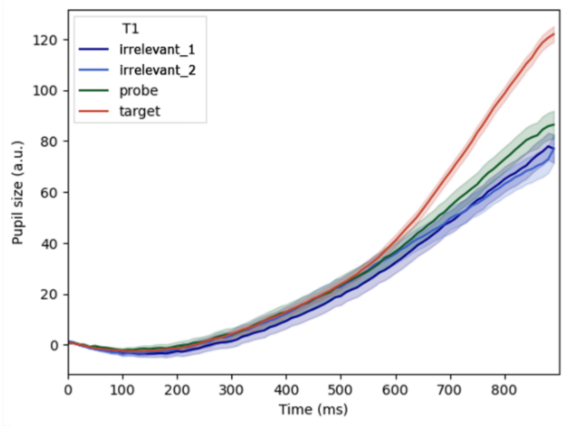
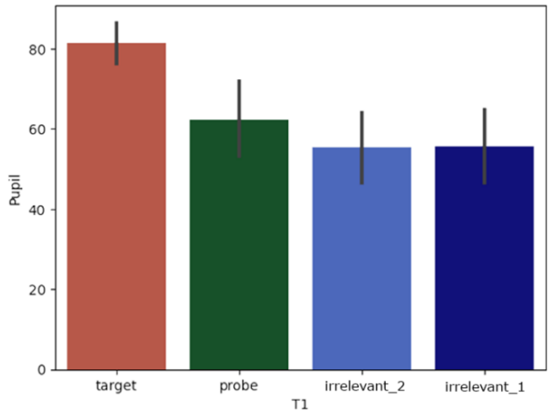


Figure 2b

Mean Pupil Size for Each Condition



Note. Time window from 600-900 ms.

Figure 3a

Traces of Mean Pupil Derivative

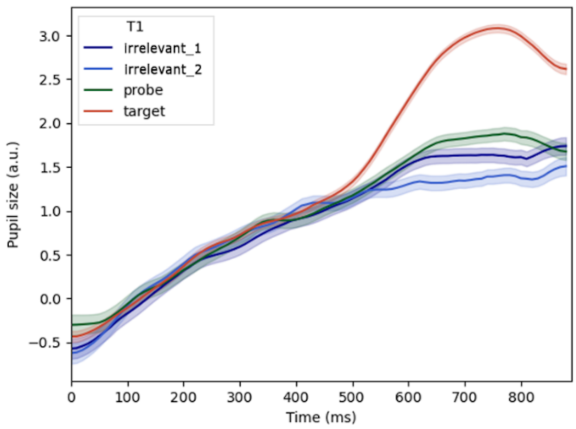
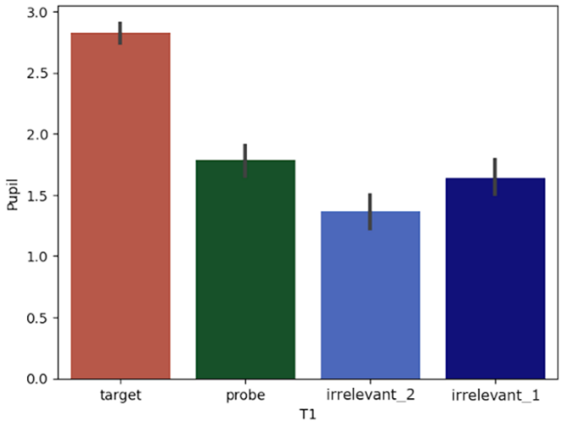


Figure 3b

Mean Pupil Derivative for Each Condition



Note. Time window from 600-900 ms.

Mean pupil size

Linear mixed effects analysis was carried out for the mean pupil size in the 600 to 900 ms time window, using irrelevant-2 as the reference condition in LMER. The results are presented in Table 1. The irrelevant-2 coefficient represents the mean pupil size in the irrelevant-2 condition. Coefficients of irrelevant-1, probe, and target represent the differences compared to the irrelevant-2 condition. The pupil dilation was significantly larger in the target condition than in the irrelevant-2 condition. Pupil dilation in the probe condition showed a difference from the irrelevant-2 condition; however, this was not statistically significant. There was no significant difference between the two irrelevant conditions.

Table 1

Linear Mixed Effects Analysis for Mean Pupil Size

Condition	Coef. (SE)	*z*	*Sig.*
Intercept	55.44 (10.24)	5.42	
Irrelevant-1	0.38 (5.94)	0.06	$p = .949$
Probe	7.15 (6.55)	1.09	$p = .275$
Target	26.18 (5.16)	5.08	$p < .001$

Note. N = 6876 (trials). α <.05.

Derivative

Linear mixed effects analysis was also carried out for the rate of change in pupil size in the same time window. The results are presented in Table 2. Irrelevant-2 was used as an intercept in LMER, and the corresponding coefficient represents the mean derivative in the irrelevant-2 condition. Coefficients of irrelevant-1, probe, and target represent the differences compared to the irrelevant-2 condition. The increase in pupil size was significantly larger in the target condition than in the irrelevant-2 condition. The rate of pupil dilation in the probe condition was statistically significant; however, contrary to our expectations, there was also a significant difference between irrelevant-1 and irrelevant-2 conditions.

Table 2

Linear Mixed Effects Analysis for Derivative

Condition	Coef. (SE)	*z*	*Sig.*
Intercept	1.36 (0.14)	9.60	
Irrelevant-1	0.27 (0.10)	2.66	$p = .008$
Probe	0.42 (0.14)	2.92	$p = .004$
Target	1.46 (0.14)	10.48	$p < .001$

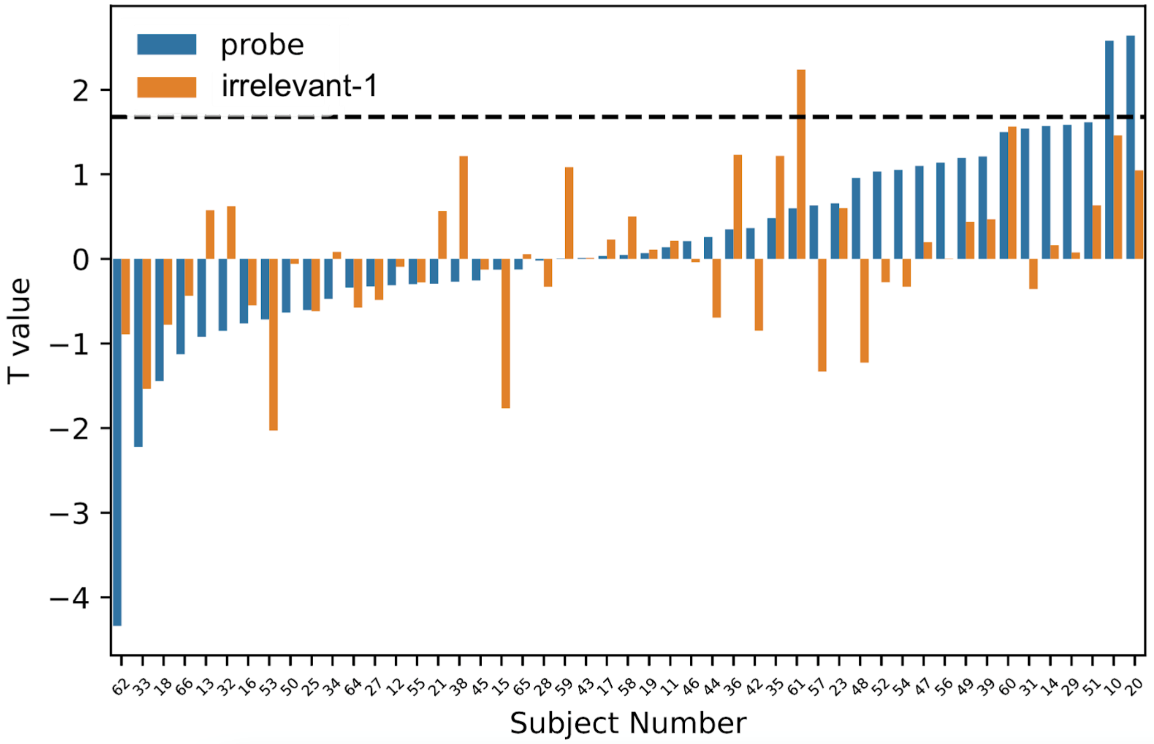
Note. $N = 6876$ (trials). $\alpha < .05$.

Individual level analysis

In the individual level analyses of mean pupil size, although 28 out of 49 participants had a positive *t*-value, only two participants showed a significant mean pupil size difference between the probe and irrelevant-2 condition. One participant showed a significant difference between the irrelevant-1 and irrelevant-2 conditions for mean pupil size (Figure 4a). Figure 4b summarizes the pupil derivative analysis. 34 participants had a positive *t*-value, out of which 12 participants showed a significant difference between the probe and irrelevant-2 condition. Nine participants showed a significant difference between the irrelevant-1 and irrelevant-2 conditions for pupil derivative. Four out of these nine participants also had a significant difference between the probe and irrelevant-2 condition (Figure 4b). Appendix B summarizes all individual results.

Figure 4a

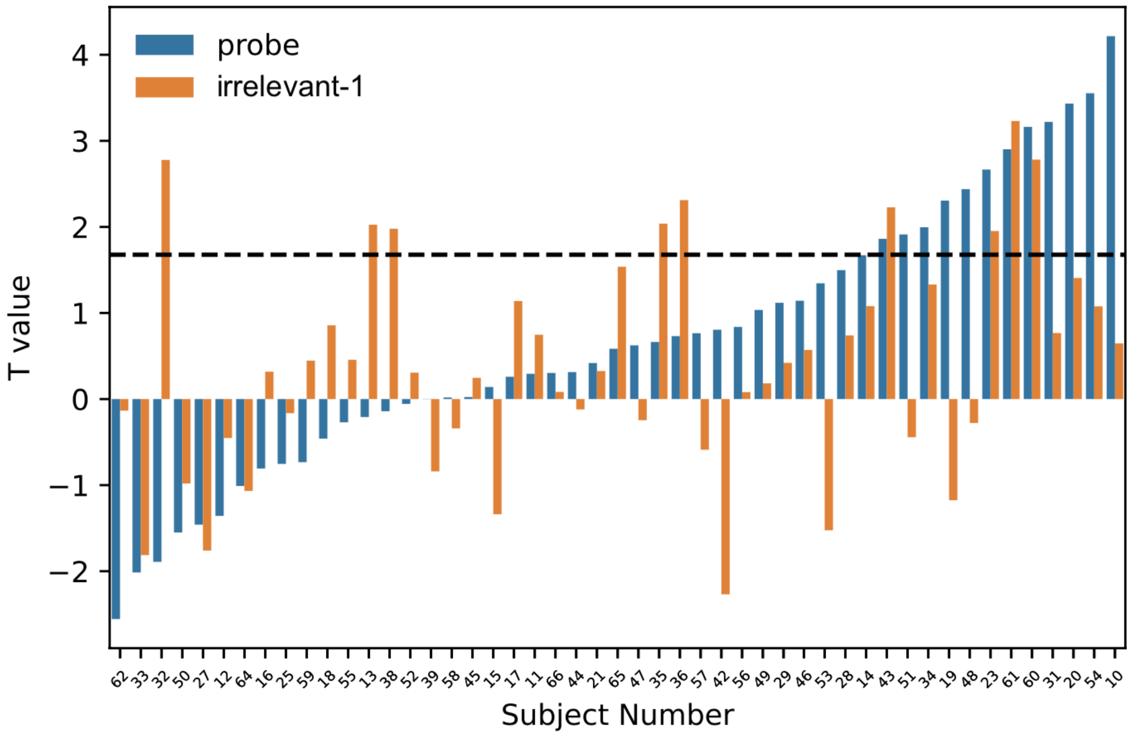
Individual Mean Pupil Size Differences (Probe and Irrelevant-1 vs. Irrelevant-2)



Note. *t*-tests were performed for each participant at $t(48) = 1.677$, $p < .05$ to compare mean pupil size in the probe and irrelevant-1 conditions to the irrelevant-2 condition. Subject numbers begin at 10 as the first participant.

Figure 4b

Individual Mean Pupil Derivative Differences (Probe and Irrelevant-1 vs. Irrelevant-2)



Note. *t*-tests were performed for each participant at $t(48) = 1.677$, $p < .05$ to compare mean pupil derivatives in probe and irrelevant-1 conditions to the irrelevant-2 condition. Subject numbers begin at 10 as the first participant.

Classification

To discriminate between guilty and innocent participants, we chose a *t*-value of $t(48) = 1.677$, $p < .05$ on individual *t*-tests as a cut-off for mean pupil size and pupil derivative factors separately. Using this method of classification resulted in marking two participants out of 49 as guilty based only on the mean pupil size, with a hit rate of .04. Eight participants out of 49 were marked as guilty based solely on pupil derivative, resulting in a hit rate of .16. Four participants were marked as undetermined based on pupil derivative.

Discussion

Cognitive pupillometry in combination with RSVP is a promising novel research method in the detection of concealed information (Chen et al., 2022; Mathôt & Vilotijević, 2022). This study aimed at conceptually replicating past findings in the research paradigm with another kind of stimulus - hidden hiding places. In general, our findings showed pupil response to be a valid measure for detecting familiarity with salient stimuli, even if task-irrelevant. This pattern is consistent with the previous literature (Bowman et al., 2013; Chen et al., 2022), even though it could not directly replicate their results, and highlights the psychosensory pupil response in reaction to an arousing stimulus (Mathôt, 2018). We hypothesized to find a difference in both mean pupil size and derivative in the predetermined time window of 600-900 ms between the probe and irrelevant-2 conditions on both an individual and group level. For mean pupil size, no significant differences at the group level (via LMER), nor at the individual level (via t-tests) for most participants, have been found. For mean pupil derivative we found a significant difference at the group level, with some participants also reflecting this finding on an individual level. It should be noted that for mean pupil derivative the irrelevant-1 condition also reached significance, which will be discussed below. Classification analysis yielded a hit rate of 4% and 16% when classifying participants as guilty based on mean pupil size or derivative respectively. We thus found mean pupil derivative to be a more accurate classification feature than mean pupil size. Because the derivative is more sensitive and pupil dilation highly reactive, pupil slope may be a promising measure for coming cognitive pupillometry studies.

Since our study did not provide the same level of sensitivity as an EEG study of the same paradigm (Bowman et al., 2013), combining several ocular measures may be the way going forward. Even in the task-relevant target condition only 16% of participants reached individual significance for mean pupil size, compared to 80% for mean pupil derivative, which suggests that pupil responsiveness alone might not be discriminative enough for this CIT set-up. Measuring the number of blinks during trials in a concealed knowledge paradigm, Fukuda (2001) found a higher, earlier peaking average blink-rate in response to probe stimuli as compared to irrelevant ones. Seymour et al. (2012) combined pupil and pre-response blink-rate measures for a classification more accurate than using these individual features on their own. For an RSVP paradigm, the temporal distribution of blinking (TDB) after the critical item and before responding to the incriminating question about concealment could be used in combination with mean pupil size and/or derivative, for an effective detection of concealed information.

In the present study all participants were in possession of knowledge that made them ‘guilty’ of the mock-crime. This is far from real world applicability. Having innocent participants introduces the possibility of false positives and true negatives. Instead of using an arbitrary value based on maximum differences between probe and irrelevant responses, the optimal cut-off for classification could then be determined by analyzing receiver operating characteristics (ROC) curves, which depict the tradeoff between true positive and false positive rates (Hanley & McNeil, 1982).

The reasons behind some of the non-significant findings in the present study are various and uncertain, but we expect a combination of the following factors. Further research might benefit from taking these limitations into account. First, the salience of the probe (the chosen hiding location) is not intrinsic and may not be high enough for our experimental set-up. By providing participants with a novel probe and not using an already established autobiographical idea or concept, such as a participant’s name, sufficient salience might not be achieved. The immersive story and the reaction task, thought to counteract this discrepancy in the probe-learning phase, may not have been enough to increase salience sufficiently, leaving the participant unaroused in probe trials during the experimental phase. Having encountered the to-be concealed information right before the experiment also does not mirror real world scenarios for most cases. Even if the stimulus is not salient to the degree one's own name is, having consolidated it in long-term memory may be a crucial additional factor (Harris et al., 2021).

Second, having the irrelevant-1 item begin with the same starting letter as probe and target instead of a random one may have increased its salience to near probe levels. This is evident in the mean pupil slope trace, where one can at times observe a larger difference between irrelevant-1 and irrelevant-2 than between irrelevant-1 and probe conditions. In fact, the irrelevant-1 condition reached significance for mean pupil derivative. This shows that a search strategy might focus solely on the first few letters of a word (Su et al., 2011), and that participants thus have a harder time picking up on smaller differences between the conditions, decreasing discriminability.

Third, the time-window of 600-900 ms might not be adequate. Especially mean pupil size appears to keep increasing after the cut-off point; for mean pupil derivative the chosen time window seems to be sufficient. In line with prior literature, we thus suggest a time-window of 600-1000 ms for future studies in order to pick up on later increases in pupil size. However, exploratory post hoc analysis with this larger time window revealed no increased significance for our study. Still, for mean pupil derivative the probe condition remained significant on the group level, while the irrelevant-1 condition did not, suggesting a better discriminability between the T1 conditions for this time-window (see Appendix C for post hoc analysis tables).

Despite these limitations, the present study has enhanced our understanding of the relationship between pupil measures and concealed information. We are confident that the current research will stimulate further investigation of this important area to improve the predictive power of this promising paradigm.

References

Alsufyani, A., Harris, K., Zoumpoulaki, A., Filetti, M., & Bowman, H. (2021). Breakthrough percepts of famous names. *Cortex*, *139*, 267–281. https://doi.org/10.1016/j.cortex.2021.02.030

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *The Journal of Applied Psychology*, *88*(1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131

Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*(12), 3046–3052. https://doi.org/10.1152/jn.00502.2014

Block, J. D., Rouke, F. L., Salpeter, M. M., Tobach, E., Kubis, J. F., & Welch, L. (1952). An Attempt at Reversal of the Truth-lie Relationship as Measured by the Psychogalvanic Response. *The Journal of Psychology*, *34*(1), 55–66. https://doi.org/10.1080/00223980.1952.9916105

Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering Countermeasures: Detecting Identity Lies by Detecting Conscious Breakthrough. *PLOS ONE*, *9*(3), e90595. https://doi.org/10.1371/journal.pone.0090595

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal Salience Search Illustrated: EEG Identity and Deception Detection on the Fringe of Awareness. *PLOS ONE*, *8*(1), e54258. https://doi.org/10.1371/journal.pone.0054258

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Chen, I. Y., Karabay, A., Mathôt, S., Bowman, H., & Akyürek, E. G. (2022). Concealed identity information detection with pupillometry in rapid serial visual presentation. *Psychophysiology*, *n/a*(n/a), e14155. https://doi.org/10.1111/psyp.14155

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91. https://doi.org/10.1016/j.dcn.2016.11.001

Elaad, E., & Ben-Shakhar, G. (2009). Countering Countermeasures in the Concealed Information Test Using Covert Respiration Measures. *Applied Psychophysiology and Biofeedback*, *34*(3), 197–208. https://doi.org/10.1007/s10484-009-9090-5

Rondeel, E. W. M., van Steenbergen, H., Holland, R. W., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00494

Fukuda, K. (2001). Eye blinks: New indices for the detection of deception. *International Journal of Psychophysiology*, *40*(3), 239–245. https://doi.org/10.1016/S0167-8760(00)00192-6

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test.* Cambridge University Press.

Geldreich, E. W. (1941). Studies of the Galvanic Skin Response as a Deception Indicator. *Transactions of the Kansas Academy of Science (1903-)*, *44*, 346–351. https://doi.org/10.2307/3624902

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Harris, K., Miller, C., Jose, B., Beech, A., Woodhams, J., & Bowman, H. (2021). Breakthrough percepts of online identity: Detecting recognition of email addresses on the fringe of awareness. *European Journal of Neuroscience*, *53*(3), 895–901. https://doi.org/10.1111/ejn.15098

Klein, P. (2022). *Plaatsen met een woonplaatscode*. Metatopos noemt alle gemeenten en plaatsen in Nederland. https://www.metatopos.eu/Wpnr.php

Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, *10*(2), 85–89. https://doi.org/10.3758/BF03214320

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

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. https://doi.org/10.1037/h0046060

MacLaren, V. V. (2001). A quantitative review of the Guilty Knowledge Test. *Journal of Applied Psychology*, *86*(4), 674–683. https://doi.org/10.1037/0021-9010.86.4.674

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), Article 1. https://doi.org/10.5334/joc.18

Mathôt, S. (2023). Python EyeLinkParser. https://pypi.org/project/python-eyelinkparser/

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01957-7

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, *53*(5), 593–604. https://doi.org/10.1111/psyp.12609

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x

Patrick, C. J. (2011). Science on the rise: Birth and development of the concealed information test. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test.* Cambridge University Press.

Peth, J., Suchotzki, K., & Gamer, M. (2016). Influence of countermeasures on the validity of the Concealed Information Test. *Psychophysiology*, *53*(9), 1429–1440. https://doi.org/10.1111/psyp.12690

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test.* Cambridge University Press.

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*(2), 205–219. https://doi.org/10.1111/j.1469-8986.2004.00158.x

Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist*, *46*(4), 409–415. https://doi.org/10.1037/0003-066X.46.4.409

Seymour, T. L., Baker, C. A., & Gaunt, J. T. (2012). Combining blink, pupil, and response time measures in a concealed knowledge test. *Frontiers in Psychology*, *3*, 614. https://doi.org/10.3389/fpsyg.2012.00614

Su, L., Bowman, H., & Barnard, P. (2011). Glancing and Then Looking: On the Role of Body, Affect, and Meaning in Cognitive Control. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00348

Wilschut, T., & Mathôt, S. (2022). Interactions Between Visual Working Memory, Attention, and Color Categories: A Pupillometry Study. *Journal of Cognition*, *5*(1), 16. https://doi.org/10.5334/joc.208

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191. https://doi.org/10.1001/jama.2013.281053

Appendix A

Figure 5

Map of Potential Probes

Appendix B

Table 3

Individual t*-values for Each Condition*

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
10	4.22**	0.65	3.70**	2.58**	1.46	2.69**
11	0.29	0.75	4.05**	0.14	0.21	1.57
12	-1.36	-0.45	1.46	-0.31	-0.09	0.32
13	-0.21	2.03*	4.28**	-0.92	0.57	0.39
14	1.67	1.08	7.23**	1.57	0.16	2.04*
15	0.14	-1.34	3.90**	-0.13	-1.77*	1.55
16	-0.81	0.32	0.04	-0.76	-0.55	0.34
17	0.26	1.14	4.15**	0.04	0.23	2.11*
18	-0.46	0.86	2.26*	-1.44	-0.78	-1.16
19	2.31*	-1.17	3.01**	0.07	0.11	1.08
20	3.43**	1.41	6.49**	2.64**	1.05	2.72**
21	0.42	0.33	2.02*	-0.29	0.57	1.92*

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
23	2.67**	1.95*	4.44**	0.66	0.60	1.48
25	-0.75	-0.16	4.48**	-0.61	-0.62	0.09
27	-1.46	-1.76	4.06**	-0.33	-0.49	0.55
28	1.50	0.74	6.15**	-0.02	-0.33	2.02*
29	1.12	0.42	0.94	1.58	0.08	-0.50
31	3.22**	0.77	7.03**	1.54	-0.36	1.41
32	-1.89	2.78**	2.22*	-0.85	0.62	1.89*
33	-2.01	-1.81	0.95	-2.22	-1.53	-0.35
34	2.00*	1.33	3.00**	-0.47	0.08	0.66
35	0.66	2.04*	4.35**	0.48	1.22	0.99
36	0.73	2.31**	3.60**	0.35	1.23	0.99
38	-0.14	1.98*	4.01**	-0.27	1.22	1.80*
39	0.00	-0.84	2.13*	1.21	0.47	0.89
42	0.81	-2.27	3.32**	0.36	-0.85	0.84

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
43	1.86*	2.23*	2.74**	0.01	0.01	0.45
44	0.31	-0.12	-0.10	0.26	-0.69	-1.16
45	0.02	0.25	2.37*	-0.26	-0.13	-0.76
46	1.14	0.57	4.27**	0.21	-0.04	1.50
47	0.62	-0.25	0.34	1.10	0.20	0.80
48	2.44**	-0.28	4.16**	0.96	-1.23	0.70
49	1.04	0.18	2.56**	1.19	0.44	1.69*
50	-1.55	-0.98	0.09	-0.63	-0.06	0.47
51	1.91*	-0.44	3.29**	1.61	0.63	0.68
52	-0.06	0.31	2.89**	1.03	-0.28	0.40
53	1.35	-1.52	3.27**	-0.72	-2.03	-0.35
54	3.55**	1.08	4.44**	1.05	-0.33	1.22
55	-0.27	0.46	0.77	-0.30	-0.28	0.02
56	0.84	0.08	1.88*	1.14	0.00	0.49

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
57	0.76	-0.59	2.36*	0.63	-1.33	0.06
58	0.02	-0.34	2.09*	0.05	0.50	1.29
59	-0.73	0.45	4.86**	0.00	1.08	2.61**
60	3.16**	2.78**	4.39**	1.50	1.56	2.00*
61	2.90**	3.23**	5.19**	0.60	2.24*	2.10*
62	-2.56	-0.13	1.83*	-4.34	-0.89	-0.55
64	-1.01	-1.07	-0.45	-0.34	-0.58	0.42
65	0.58	1.54	5.25**	-0.13	0.06	0.48
66	0.30	0.08	0.70	-1.13	-0.44	-0.66

Note. t-values for the *t*-test comparing mean pupil size and derivative in probe, target, and irrelevant-1 condition to irrelevant-2 condition. Subject numbers begin at 10 as the first participant. *$p < .05$. **$p < .01$.

Appendix C

Table 4a

Linear Mixed Effect Analysis for Mean Pupil Size (Time Window 600-1000 ms)

	Coef. (SE.)	*z*	p
Intercept	60.41 (11.02)	5.48	< 0.001
Irrelevant-1	-6.28 (7.33)	-0.86	0.391
Probe	4.44 (7.52)	0.59	0.555
Target	29.04 (6.17)	4.71	< 0.001

Note. Exploratory post-hoc analysis. *N* = 46. Irrelevant-2 is the intercept. The data no longer contains trials where the T1-position was the 8th place. 11 participants excluded: 6 for missing data, 5 for accuracy.

Table 4b

Linear Mixed Effect Analysis for Mean Pupil Slope (Time Window 600-1000 ms)

	Coef. (SE.)	*z*	p
Intercept	1.41 (0.15)	9.59	< 0.001
Irrelevant-1	0.15 (0.11)	1.32	0.188
Probe	0.36 (0.15)	2.38	0.017
Target	1.32 (0.14)	9.32	< 0.001

Note. Exploratory post-hoc analysis. *N* = 46. Irrelevant-2 is the intercept. The data no longer contains trials where the T1-position was the 8th place. 11 participants excluded: 6 for missing data, 5 for accuracy.