Detecting Concealed Location Names with Pupillometry Using Rapid Serial Visual Presentation in a Mock-Crime Scenario

Teresa Nguyen

s4272145

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number 002

Supervisor: dr. Robbert van der Mijn

Second evaluator: Ana Vilotijević

In collaboration with: Charlotte Bosch, Anouk Feijen,

Danie Hettema, Tobias Kroner, and Janette Possul

April 14, 2023

Abstract

Detecting concealed information can be useful to distinguish between guilty and innocent suspects in law enforcement. A reliable method to test that is the combination of concealed information testing (CIT) and rapid serial visual presentation (RSVP). While CIT relies on physiological responses to salient stimuli to detect concealed information, it is vulnerable to countermeasures that may conduce false negatives. This vulnerability can be reduced by adding an RSVP that presents stimuli so quickly that they cannot be consciously reacted to. Previous research using CIT and RSVP found P3 brain potentials to be significant indicators of concealed information. Alternative indicators may be ocular measurements, specifically pupil dilation and derivatives, since they are comparingly cheaper and user-friendly. In our study, we examine pupil dilation and derivative in response to location names presented in a CIT-RSVP framework as part of a mock-crime scenario. We propose that pupil size and derivative would be significantly larger and steeper in the real hiding location (probe) condition than in control (irrelevants) conditions. Forty-nine participants read a mock-crime scenario and reacted to a task-relevant location (target) in an RSVP stream of probe, irrelevants, and distractors. Applying linear mixed effects regression on a group level, we only found significant differences with pupil derivatives. Employing *t*-tests, we correctly identified 4% of guilty participants using pupil size, and 16% using pupil derivatives. Overall, this study showed that pupil derivatives are promising indicators for concealed information; however, further testing is required for this paradigm to be considered in law enforcement.

Keywords: Concealed Information Testing (CIT), Rapid Serial Visual Presentation (RSVP), pupillometry, mock-crime scenario

Detecting Concealed Location Names with Pupillometry Using Rapid Serial Visual Presentation in Mock-Crime Scenario

Mentioning lie detection will surely evoke a particular kind of image in one's mind: a suspect sitting in front of a detective, attached to a machine via multiple cables, measuring their physiological response to a series of questions. This technique is indeed commonly utilized in US-American police interrogation in combination with the Control Question Test (CQT) and is a decisive tool to determine guilt or innocence in suspects (National Research Council of the National Academy of Sciences, 2003). The CQT consists of crime-related, general misconduct-related, and irrelevant neutral questions (Ben-Shakhar, 2002). However, they have increasingly been criticized for their unreliability and vulnerability to countermeasures (National Research Council of the National Academy of Sciences, 2003; Iacono & Ben-Shakhar, 2019). The main points of criticism were its bias, lack of structure, non-scientific basis, and pressure on the suspect to confess to the point of false confessions (Verschuere et al., 2011). Hence, it was important to find methods with solid empirical foundation to ensure fair and effective interrogation.

A well-studied and evidence-supported alternative is the concealed information test (CIT) (Lykken, 1959). The approach is to present the suspect with crime-relevant stimuli (*probes*) along with other non-crime-relevant stimuli (*irrelevants*) while measuring their physiological reaction; e.g., skin conductance response (SCR), heart rate, eye movements, brain activity via EEG (Lykken, 1959; Suzuki et al., 2004; Verschuere et al., 2007; Bowman et al., 2013). It is assumed that if the suspect consistently reacts significantly differently to the crime-relevant stimulus as compared to the control stimuli, the suspect must have critical knowledge about the crime-relevant stimulus (klein Selle et al., 2018).

Unfortunately, the CIT by itself is also prone to simple countermeasures. Thinking of other exciting experiences or self-inflicted physical pain can induce a misleading significant response to neutral stimuli (Rosenfeld et al., 2004). For that reason, presenting stimuli in rapid serial visual presentation (RSVP) has been suggested as an improvement. It presents stimuli so quickly that they are on the fringe of awareness – so shortly that they cannot be consciously processed but long enough for salient items to be picked up and be recallable (Broadbent & Broadbent, 1987). Certain physiological responses accompany this recognition of salient stimuli that may serve as indicators of detection of concealed information, e.g., the P3 potential (Bowman et al., 2013) and perhaps pupil dilation (Chen et al., 2022). Therefore, we aim to further expand on the combination of CIT and RSVP in this research while using ocular measurements, specifically pupillometry, as the physiological response, similarly to Chen et al. (2022).

In an earlier combination of CIT and RSVP (Bowman et al., 2013), a P3 event-related potential obtained through EEG was used as an indicator for the recognition of a salient stimulus. Participants saw a stream of names on a screen, while they were instructed to report if they saw their self-chosen fake name (*target*). Their own real name (probe), a naturally salient stimulus, could also instead be shown in the stream. Despite being instructed not to react to it, they still demonstrated a significant P3 response. That was remarkable because this implies that humans showed a measurable P3 reaction to salient information even on the fringe of awareness. In a subsequent study, adding RSVP also turned out to be resistant against cognitively controlled countermeasures, such as concentrating on not seeing their real name (probe) or counting the number of times an irrelevant frequent name appears (irrelevant) (Bowman et al., 2014).

Stimuli in an RSVP are presented one at a time, all placed on the same position on the screen, for about 100 ms each. They are categorized as either *critical stimuli* (probes, targets, and irrelevants) or non-critical stimuli (*distractors*). Probes are the “concealed information”: In the current study, they are crime-relevant and therefore assumed to be salient. They are task-irrelevant, though, as the participant is not supposed to actively look for it; accordingly, it is not part of the experiment's task instruction. The reaction to the probe is what is crucial because it provides evidence about whether the probe is still being registered and reacted to despite the fast-paced presentation of stimuli. A significant physiological reaction to it, e.g., a P3 signal, usually indicates intrinsic salience. Since the probe must be concealed as part of CIT, they are instructed to look for the target in the RSVP stream as pretence. Additionally, in order to maintain the participant's attention throughout all RSVP trials, they have to indicate whether they saw the target after each trial. Furthermore, a control condition is implemented by adding irrelevants: An irrelevant is a non-task relevant and non-salient stimulus shown at the same frequency as the probe. Since the probe and irrelevants are equalized in frequency, any difference in physiological reaction between them can be attributed to the probe's intrinsic salience rather than frequent exposure. Finally, distractors are non-task relevant non-salient stimuli that act as randomized fillers for the rest of the RSVP stream (Bowman et al., 2013). Bowman and colleagues' work (2013; 2014) set an example with potential for further research to test different variations of their paradigm, e.g., with different physiological measures.

Following the same CIT-RSVP paradigm, Chen et al. (2022) used pupil dilation instead of P3 signals (Bowman et al., 2013) as a physiological indicator for concealed information. They successfully found differences in pupil dilation between the fake name (target) and control name (irrelevants) conditions. While there was only a positive trend regarding the difference between the real name (probe) and the irrelevants in individuals, it was significant on a group level. The advantage of using pupil dilation instead of P3 signals (Bowman et al., 2013) is that the former is more practical and easier to measure since measuring EEG is time-intensive and invasive. Pupil measurements, on the contrary, can be administered with a comparably cost-efficient eye tracker. Eye trackers can also measure pupil size and eye movement without the participant being physically connected to the device (Chen et al., 2022), thus offering a less invasive alternative. Therefore, this method is more user-friendly and can easily be worked with, which is important for a smooth incorporation into practical usage in law enforcement.

Besides the practical advantages of pupillometry over EEG, the theory behind the relationship between pupillometry and concealed information is promising as well. It is widely known that pupils contract when exposed to light and when focusing on nearby objects (Backhaus, 2011; McDougal & Gamlin, 2008). Pupil dilation, on the other hand, can be an indicator of cognitive arousal and psychosensation, such as heightened mental effort (Hess & Polt, 1960; 1964) or as an orienting response (klein Selle, 2018). The orienting response (OR) is an instantaneous response in behavior and physiology to sudden stimuli, especially to unexpected and salient ones (klein Selle et al., 2022). An example of that is the "cocktail party effect": Despite a noisy environment and one's preoccupation, hearing one's own name will elicit an immediate reaction. Similarly, pupil dilation accompanies the OR, thus occurring when encountering novel or salient stimuli (Wang & Munoz, 2014; 2015). This underlying mechanism can therefore also be used to identify guilty suspects (Lieblich et al., 1970; Lykken, 1974; klein Selle et al., 2018) as, according to the OR, they will show physiologically different responses to the salient stimulus as compared to the non-salient control one.

The most well-established physiological indicators of OR are the skin conductance response (SCR) (Bradley et al., 2008) and the P3 signal (klein Selle et al., 2016; 2017; 2018; 2021; Nieuwenhuis et al., 2021). Pupil dilation positively correlates with SCR (Bradley et al., 2008), and they both show the same characteristic of habituation (klein Selle et al., 2018), a defining feature of the OR (Sokolov, 1963). The habituation process describes the effect of the OR decreasing with more frequent appearances of a stimulus, which also applies to the

SCR (klein Selle et al., 2018). The effect of the pupil dilation did decrease the more often the probe was shown (Chen et al., 2022), which could be in line with the habituation process. In order to address this issue, we will decrease the amount of experimental RSVP trials as Chen et al. (2022) suggested. Since the habituation process implies that the effect can already be observed in the beginning, we think that this is a reasonable modification.

Additionally, pupil dilation may be mediated by phasic activation of the locus coeruleus-norepinephrine (LC-NE) system, associated with alertness (Nieuwenhuis et al., 2011). The LC-NE system's phasic response activates for objects that reward the participant (e.g., a high score), especially when task-related attention is required (Gabay et al., 2010). This provides relevant biological background about the pupil dilation for the task-relevant target stimulus. Since task-relevant and salient stimuli (target) as well as task-irrelevant and salient stimuli (probe) have evidently been linked to the LC-NE system and the OR respectively, it is a reasonable assumption to use pupil dilation to detect concealed information.

In terms of concealed information, while names carry naturally high salience (Bowman et al., 2013; Chen et al., 2022), cities can also hold salience as possible hiding locations of crime-related evidence or as autobiographical information (Norman et al., 2020). In the current study, the probe, target, and one control stimulus (*irrelevant-1*) will start with the same letter to account for the effect of searching strategies, in which they may only read the starting letter (Harris et al., 2021). In terms of setting, mock-crime scenarios are the most ecologically valid method within ethically acceptable circumstances to simulate realistic crime-related salience (Ben-Shakhar & Elaad, 2003). In terms of physiological measurements as indicators for detecting concealed knowledge, pupil dilation along with how fast the pupil dilates (*pupil derivative*) (Lubow & Fein, 1996) have been shown to be reliable indicators within a mock-crime scenario (Seymour et al., 2013).

Therefore, taking the previous findings into account, it is reasonable to assume that pupil size and derivative can indicate the detection of concealed information in an RSVP-based task. While intending to replicate the findings of Chen et al. (2022), we also offer a novel extension by introducing a mock-crime scenario, using location names as stimuli, and adopting pupil derivatives as an additional measurement. Specifically, we hypothesize that the pupil size and derivative would be significantly larger and steeper in the real hiding location (probe) condition than in control (irrelevants) conditions.

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

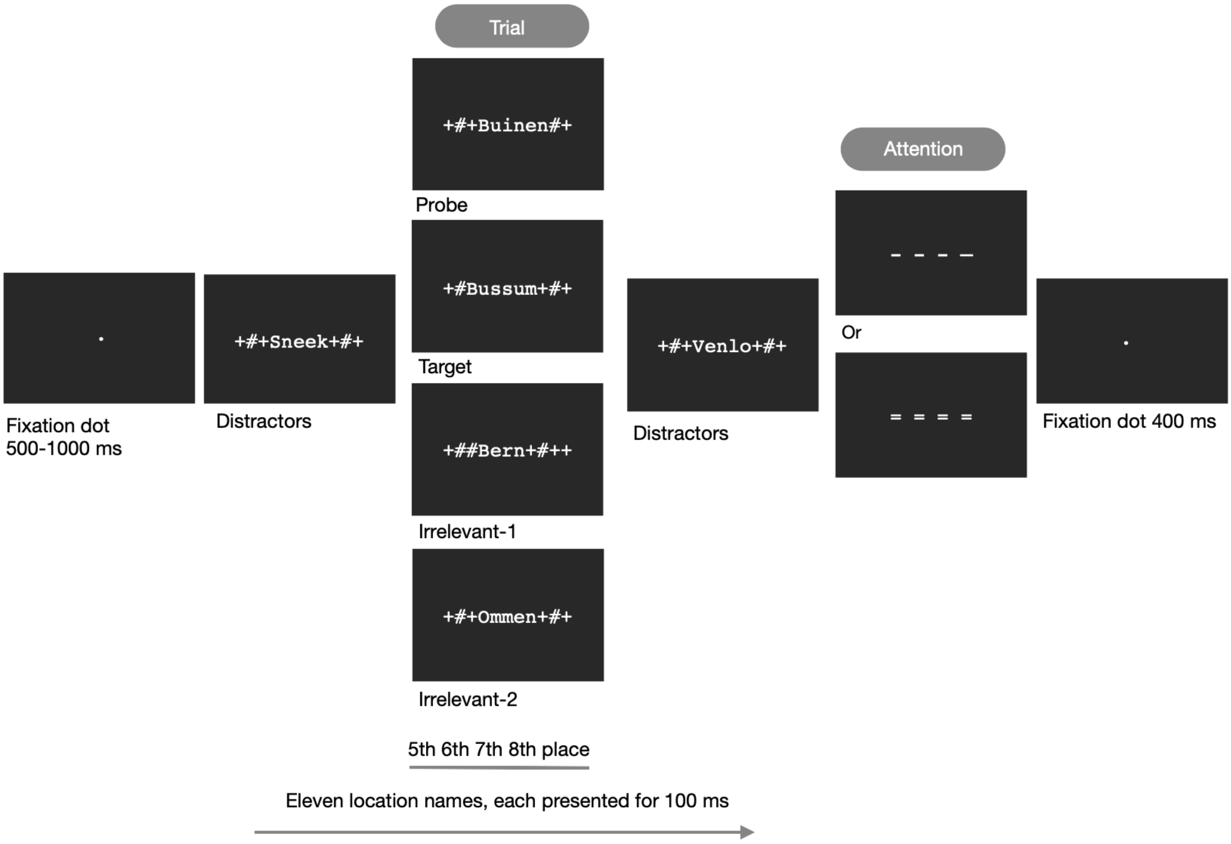
We used a within-participant experimental design. As part of a mock-crime scenario, an RSVP task was used, with a list of Dutch locations as stimuli (Klein, 2022; see OSF). From the list of 1105 Dutch locations, 15 were randomly chosen to be the potential real hiding location (probes) and removed from 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).

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 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.

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 (8^{th}) 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.

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.

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.

Figure 2a

Traces of Mean Pupil Size

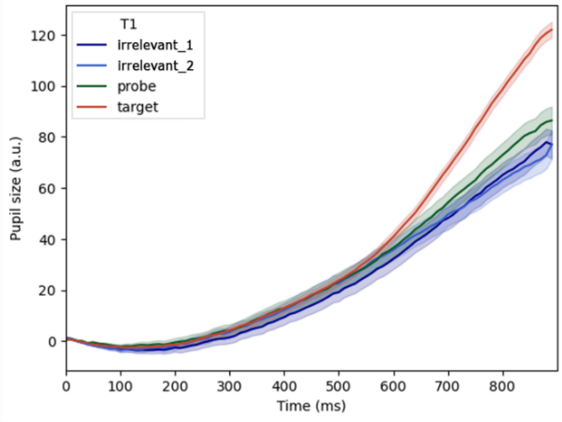
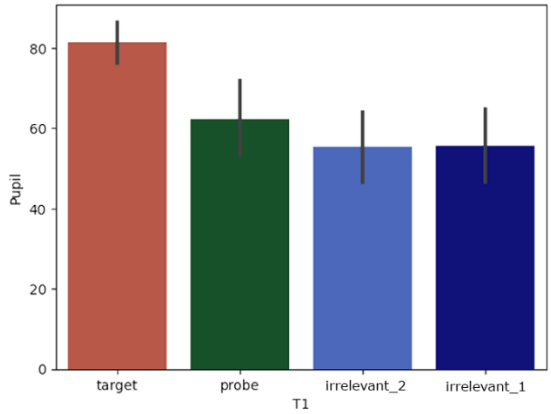


Figure 2b

Mean Pupil Size for Each Condition



Note. Time window from 600-900 ms.

Figure 3a

Traces of Mean Pupil Derivative

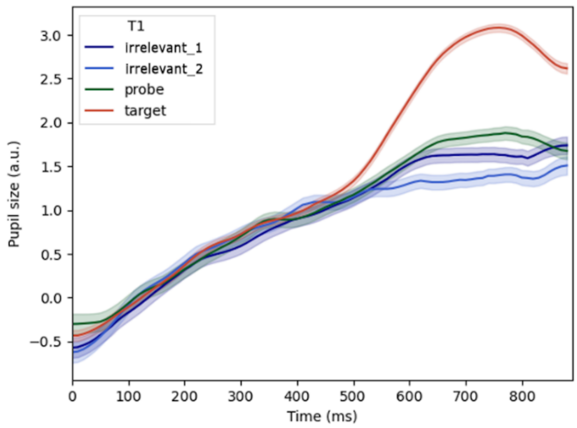
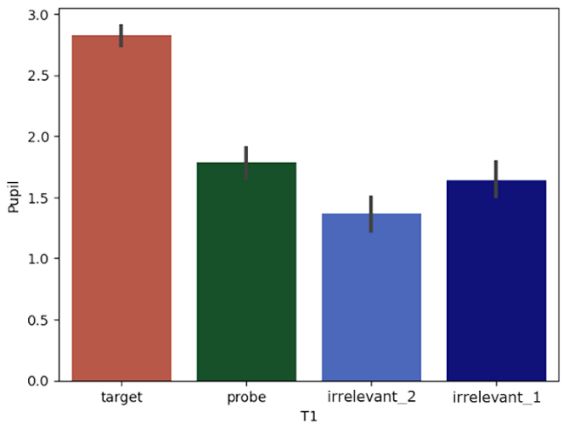


Figure 3b

Mean Pupil Derivative for Each Condition



Note. Time window from 600-900 ms.

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). $\alpha < .05$.

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	$p < .001$

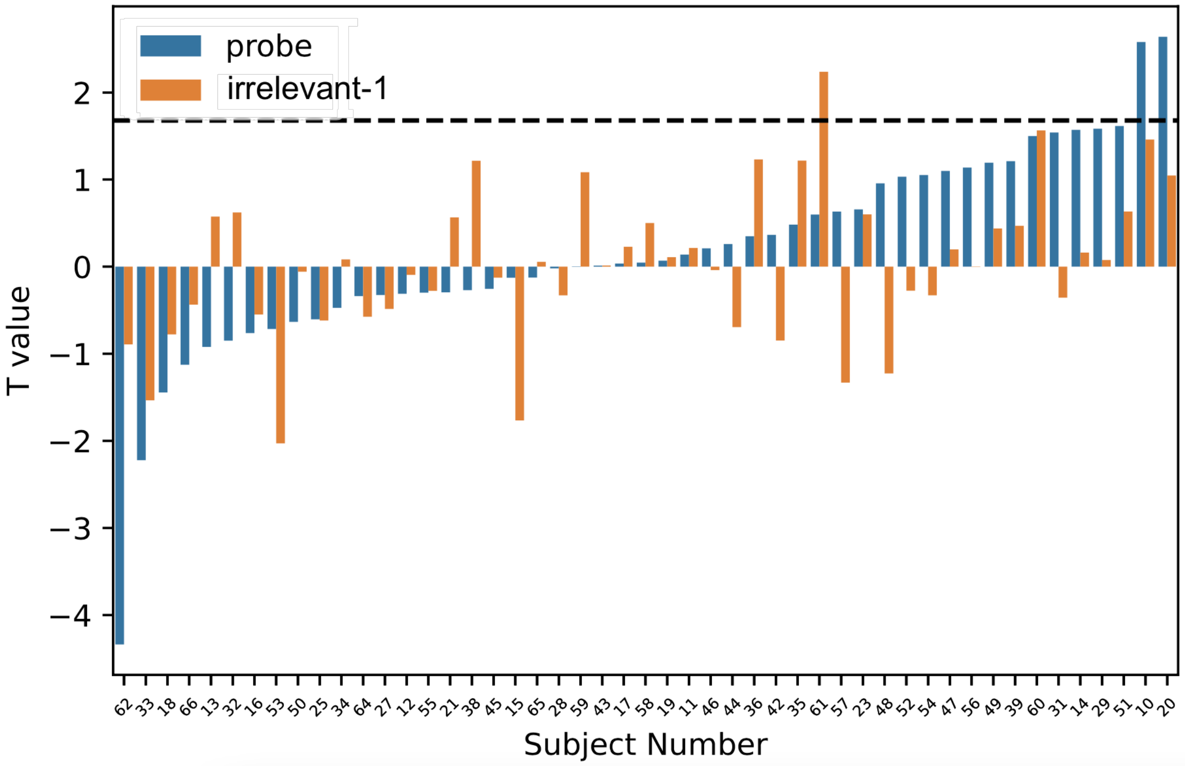
Note. $N = 6876$ (trials). $\alpha < .05$.

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

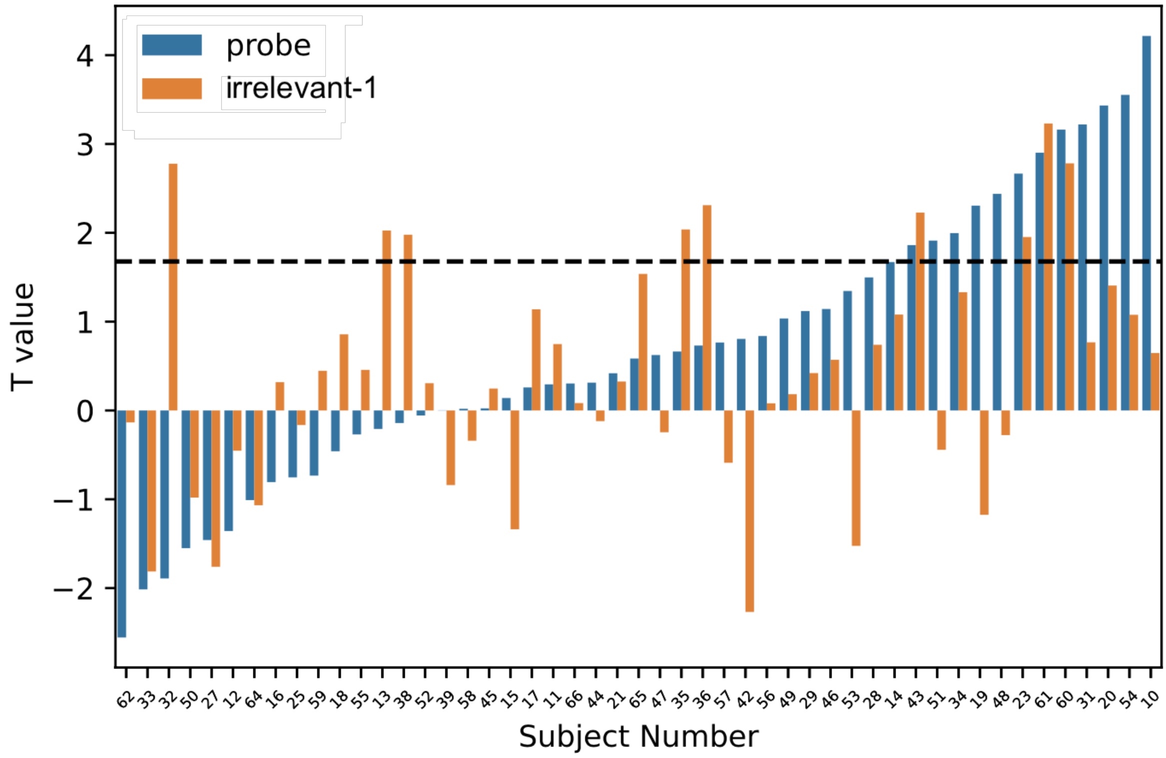
Individual Mean Pupil Size Differences (Probe and Irrelevant-1 vs. Irrelevant-2)



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

Our results provided little evidence for pupil dilation (hit rate of 4%) and moderate evidence for pupil derivatives (hit rate of 16%) as suitable indicators of concealed information. There was no significant difference between the probe and irrelevant-2 for pupil size on a group level but there was a positive trend. The findings do not support our hypothesis that pupil size would be significantly larger in the real hiding location (probe) trials than in the control trials (irrelevants). This is not in line with previous studies that found pupil dilation (Chen et al., 2022; Lubow & Fein, 1996) to be a significant indicator of concealed information. Using pupil derivative as the other indicator of concealed information, we found a significantly faster pupil dilation to the probe than to the irrelevant-2 on a group level. However, this was confounded by a significantly faster pupil dilation in the irrelevant-1 trials than the irrelevant-2 trials. That finding supports our hypothesis that pupil derivatives would be significantly steeper in the real hiding location (probe) trials than in the control condition (irrelevant-2). It is also consistent with earlier research finding pupil derivatives to be significant indicators of concealed information (Seymour et al., 2013).

The significantly faster pupil dilation to both probe and irrelevant-1 implies that irrelevant-1 elicited a similar pupil reaction as the probe, which further suggests that it cannot be certainly concluded that the significantly different reaction to the probe was solely because of its intrinsic salience. Since irrelevant-1 was shown at the same frequency and with the same starting letter as the probe, it could have been mistaken as the probe or target; hence explaining the significant difference between the two control conditions on both group level and individually for nine participants. It is noteworthy, however, that the same difference was not significant for pupil derivatives in the originally intended time window of 600-1000 ms in a post-hoc analysis (Appendix C). This finding suggests that this time window may be compatible with this paradigm, as recommended by prior research (Chen et al., 2022); and that pupil derivatives might be reasonably reliable indicators for concealed information due to the higher discriminability between critical conditions. Additionally, compared to pupil size, it is a more sensitive measurement as evident by its higher hit rate. Thus, it may be a more suitable fit for the delicate and fast fluctuation in pupil size in an RSVP.

Interestingly, despite setting the same starting letter for the real hiding location (probe), the target, and irrelevant-1 to equalize the effect of searching strategies over conditions (Harris et al., 2021), it can be assumed that participants adapted to this by changing their searching strategy accordingly. For example, one participant (25) mentioned in their spontaneous feedback that they instead focused on the middle part of the word (e.g., “melu” in Hemelum) instead of the starting letter. This supports the findings that search strategies emerge instinctively as part of an endogenous orienting response (Berger et al., 2005; klein Selle, 2018). That is also in line with the glance-look model of cognitive control (Su et al., 2011): Stimuli are first being “glanced” at. If the presented stimulus coincides with the search criterion (e.g., starting letter “B”), it will be further “looked” at and penetrates the deeper consciousness. In this case, it means that the participant disregards any stimuli that do not match the stimulus they are actively looking for (target) on a self-determined search criterion. However, since the participant knew that the target and probe have the same starting letter, they would subsequently choose another unique search criterion that enables them to quickly identify the target (e.g., “melu” in Hemelum) instead of reading the whole stimulus. Given the nature of the RSVP, these explanations seem plausible as it is difficult to fully read and comprehend every stimulus presented in such a short time span. Still, also taking individual differences in perception and processing in visual search strategies into account (Dreneva & Krichevets, 2021), different results in significance seem reasonable.

In order to explain why some individuals showed a significant effect and some did not, we propose that fatigue could have been a moderator. Two participants (33, 64) indicated in their spontaneous feedback that they took long breaks in between trial blocks, which could have been due to pre-existing fatigue or a growing strain on their eyes. That is congruent to previous research finding that pupil dilation to stimulus-evoked stimuli decreased with increasing (mental) fatigue (Hopstaken et al., 2015), which can also be applied to the task-irrelevant probe in the current study.

Summarizing the previous points, we firstly recommend future research to use pupil derivatives as indicators of concealed information due to their sensitivity and stable discrimination in the time window of 600-1000ms. Secondly, we suggest including a third irrelevant stimulus as a control condition, presented at the same frequency as the probe but without the same starting letter, in order to determine whether the frequency or similarity caused the confounding significant difference between the two control conditions (irrelevant-1 and -2). Furthermore, we suggest adding a short questionnaire at the end asking the participant how often they saw each critical stimulus, similarly to Chen et al. (2022). Expanding on this, it might prove useful, too, to instruct the participant to type out their real hiding location (probe) after the RSVP task to ensure that they remembered it until the very end. These two implementations may serve as effective control mechanisms to roughly estimate what the participant actually consciously perceived and remembered.

Thirdly, future research may employ a more active and involved mock-crime scenario similar to Lubow et al. (1996) in combination with an RSVP task. This might reflect real life scenarios more accurately, which may positively influence the salience of incriminating stimuli (Ben-Shakhar & Elaad, 2003). Additionally, implementing an “innocent” condition is vital to account for the probability of false positives as false convictions are a valid concern in real life (National Research Council of the National Academy of Sciences, 2003).

In conclusion, this study showed that pupillary measurements, especially pupil derivatives, are promising indicators for concealed information. However, further extensive testing is required, for which the aforementioned suggestions should be taken into account, in order for this paradigm to be considered for practical application in law enforcement.

References

Backhaus, S. (2011). Pupillary Light Response, Pupillary Response. *Encyclopedia of Clinical Neuropsychology*, 2086–2086. https://doi.org/10.1007/978-0-387-79948-3_272

Ben-Shakhar, G. (2002). A Critical Review of the Control Questions Test (CQT). *Handbook of Polygraph Testing*.

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

Berger, A., Henik, A., & Rafal, R.D. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology. General*, *134 2*, 207-21. https://doi.org/10.1037/0096-3445.134.2.207

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

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*, 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

Dreneva, A. A., & Krichevets, A. N. (2021). Analysis of individual categorial visual search strategies. *Neuroscience and Behavioral Physiology*, *51*(9), 1317–1322. https://doi.org/10.1007/s11055-021-01196-x

Gabay, S., Pertzov, Y., & Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception & Psychophysics*, *73*(1), 123–129. https://doi.org/10.3758/S13414-010-0015-4

Harris, K., Miller, M., 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

Hess, E. H., & Polt, J. M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. *Science*, 349–350. https://doi.org/10.1126/SCIENCE.132.3423.349

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/SCIENCE.143.3611.1190

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, *110*, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Iacono, W. G., & Ben-Shakhar, G. (2019). Current status of forensic lie detection with the comparison question technique: An update of the 2003 National Academy of Sciences report on polygraph testing. *Law and Human Behavior*, *43*(1), 86–98. https://doi.org/10.1037/LHB0000307

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

klein Selle, N., Gueta, C., Harpaz, Y., Deouell, L. Y., & Ben-Shakhar, G. (2021). Brain-based concealed memory detection is driven mainly by orientation to salient items. *Cortex*, *136*(2), 41–55. https://doi.org/10.1016/j.cortex.2020.12.010

klein Selle, N., Verschuere, B., & Ben-Shakhar, G. (2018). Concealed Information Test: Theoretical Background. *Detecting Concealed Information and Deception: Recent Developments*, 35–57. https://doi.org/10.1016/B978-0-12-812729-2.00002-1

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2016). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*, *53*(4), 579–590. https://doi.org/10.1111/psyp.12583

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. *Psychophysiology*, *54*(4), 628–639. https://doi.org/10.1111/psyp.12825

Lieblich, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR: Detection of Information as a Function of Stimulus Set Size. *Psychophysiology*, *6*(5), 601–608. https://doi.org/10.1111/J.1469-8986.1970.TB02249.X

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, 385 –388.

Lykken, D. T. (1974). Psychology and the lie detector industry. *The American Psychologist*, *29*(10), 725–739. https://doi.org/10.1037/H0037441

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/JOC.18/METRICS/

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

McDougal, D. H., & Gamlin, P. D. R. (2008). Pupillary Control Pathways. *The Senses: A Comprehensive Reference*, *1*, 521–536. https://doi.org/10.1016/B978-012370880-9.00282-6

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

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

Norman, D. G., Gunnell, D. A., Mrowiec, A. J., & Watson, D. G. (2020). Seen this scene? Scene recognition in the reaction-time Concealed Information Test. *Memory & Cognition*, *48*(8), 1388–1402. https://doi.org/10.3758/s13421-020-01063-z

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

Seymour, T. L., Baker, C. A., & Gaunt, J. T. (2013). Combining Blink, Pupil, and Response Time Measures in a Concealed Knowledge Test. *Frontiers in Psychology*, *0*, 614. https://doi.org/10.3389/FPSYG.2012.00614

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, *25*(1), 545–580. https://doi.org/10.1146/annurev.ph.25.030163.002553

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*(12), 348. https://doi.org/10.3389/FPSYG.2011.00348

Suzuki, R., Nakayama, M., and Furedy, J. (2004). Specific and reactive sensitivities of skin resistance response and repiratory apnea in a Japanese Concealed Information Test (CIT) of criminal guilt. *Canadian Journal of Behavioral Science*, *36*, 202 –209.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). Memory Detection: Theory and Application of the Concealed Information Test. *Memory Detection: Theory and Application of the Concealed Information Test*, 1–319. https://doi.org/10.1017/CBO9780511975196

Verschuere, B., Crombez, G., Koster, E., Van Bockstaele, B., and De Clercq, A. (2007). Startling Secrets: startle eye blink modification by concealed crime information. *Biological Psychology*, *76*, 52 –60.

Wang, C., & Munoz, D. P. (2014). Modulation of stimulus contrast on the human pupilorienting response. *European Journal of Neuroscience*, 1–11. https://doi.org/10.1111/ejn.12641

Wang, C., & Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134–140. https://doi.org/10.1016/j.conb.2015.03.018

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

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*
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*

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
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
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
57	0.76	-0.59	2.36*	0.63	-1.33	0.06

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
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 4

Linear Mixed Effect Analysis for Mean Pupil Slope

	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. $N = 46$. 11 participants excluded: 6 for missing data, 5 for accuracy. Exploratory post-hoc analysis. Irrelevant-2 is the intercept. The data no longer contains trials where the T1-position was the 8th place and the time window is 600-1000 ms.