# Detection of Concealed Information in a Mock-Crime Scenario Using Pupillometry and Rapid Serial Visual Presentation

Janette Possul

S4357779

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 2223_1b_02

Supervisor: Dr. Robbert van der Mijn

Second evaluator: MSc Ana Vilotijević

In collaboration with: Charlotte Boch, Anouk Feijen, Danie Hettema, Teresa Nguyen and Tobias Kroner

April 14, 2023

**Abstract**

Concealed information testing (CIT) has been researched extensively as an evidence-based method to identify individuals with crime-related knowledge. However, CIT relies on physiological responses that can be voluntarily manipulated, leading to confounded results. Combining CIT with rapid serial visual presentation (RSVP) can improve this issue, while pupillometry may provide a more feasible indicator of cognitive processes compared to other physiological measures. We combined CIT, RSVP, and pupillometry to test whether pupil size and the rate of change in pupil size (derivative) are larger when participants are presented with crime-relevant stimuli compared to crime-irrelevant stimuli. We designed a within-subjects experiment, where participants took part in a mock crime after which they completed an RSVP task while their pupil reactions were recorded. While we found no significant group-level differences for the pupil size ($\alpha < .05$), the derivative showed a significant difference ($\alpha < .05$). However, the control condition yielded a significant difference as well. At the individual level, using *t*-tests, two participants showed a significant mean pupil size difference in pupil size and 12 participants showed a significant difference in the derivative, resulting in hit rates of 4% and 16% respectively. These findings are different from the previous study using the same paradigm, however, we demonstrated that the derivative is potentially a more sensitive measure than pupil size. Research on combined CIT, RSVP, and pupillometry is scarce, and further investigation is needed to explore the optimal conditions suitable for field use.

*Keywords:* pupillometry, rapid serial visual presentation, mock crime, concealed information testing

**Detection of Concealed Information in a Mock-Crime Scenario Using Pupillometry and Rapid Serial Visual Presentation**

Polygraph testing to detect deception made its way to forensic settings as early as 1923 (Synnot et al., 2015). Although polygraphs have been continuously used in crime investigations, their scientific validity is still debated a century later, and the search for a reliable and applicable method continues. In 1959, David Lykken designed and experimentally tested a novel way to detect concealed information and called it Guilty Knowledge Test (GKT), which is presently known as Concealed Information Test (CIT). CIT has shown high discriminability in laboratory settings, especially in correctly identifying naive individuals. When used optimally, CIT offers a valid, evidence-based method to test for concealed knowledge (Icano, 2011; Ben-Shakar 2012)**.** In the CIT, an examinee is presented with a list of items in which one of the items, called the *probe*, is crime-relevant and the rest of the items are equally plausible yet crime-irrelevant, called *controls*. If the examinee knows about the crime, the crime-relevant item (probe) elicits an involuntary physiological response (Patrick, 2011). Systematic physiological reactions related to the activation of the autonomic nervous system, such as larger skin conductance responses (MacLaren, 2001), respiratory suppression (Elaad et al, 1992), and heart rate declaration (Gamer, 2011), have been correlated to recognizing crime-relevant items, compared to the neutral items.

Moreover, a distinct electrophysiological response observed in the electroencephalograph (EEG) pattern has been used as a reliable measure to detect concealed knowledge in CIT (Rosenfeld, 2011). The procedure combining CIT and EEG, also known as P300 oddball paradigm, presents subjects with frequent, irrelevant stimuli as controls and infrequent relevant “oddball” stimuli as probes. These oddballs elicit larger amplitude P300 in subjects to whom the relevant stimuli carry any meaning, such as crime-related knowledge. For the innocent subjects, on the other hand, these probes are no different from the neutral control stimuli (Rosenfeld et al., 1987). The requirement to test the subjects with the P300 paradigm, however, is that the subjects process each stimulus they are presented with. To ensure the subject's attention, a third type of stimuli can be added to the test - a task-relevant, but otherwise meaningless *target* stimuli, which subjects must find and react to (Farewell & Donchin, 1991). P300 response using P300 oddball paradigm has been widely tested as a valid indicator for detecting concealed information (Mejier et al., 2014).

Although a reliable method under correct conditions, CIT is susceptible to countermeasures such as pressing fingers or imagining an excitable situation when seeing crime-irrelevant items. These countermeasures can lead to indistinguishable physiological responses between the crime-relevant and crime-irrelevant items (Rosenfeld, 2004). One solution to combat countermeasures is to use the rapid serial visual presentation (RSVP) paradigm in combination with CIT (Bowman et al., 2013). In RSVP, the stimulus is presented for a very brief time, about 100 ms, and immediately masked by the next stimulus. Presenting the relevant and irrelevant items in a stream of RSVP makes it nearly impossible for the examinees to use countermeasures because there is no time to consciously control the responses to the stimuli (Potter, 1976; Broadbent & Broadbent, 1987; Bowman et al., 2013).

Indeed, studies combining CIT with RSVP have reported promising results. Bowman et al. (2013) demonstrated that it is possible to detect identity deception by using RSVP and the P300 oddball paradigm described earlier. Participants' real name was used as a probe, a fake name, that participants chose for themselves, was used as a target, and a randomly chosen list of names was used as irrelevant control items. The names were presented to the subjects as RSVP streams and the participants were instructed only to look for, and to respond to their fake names. The results showed a larger P300 component of the EPR during probe trials compared to the irrelevant control trials. The authors called the technique “subliminal salience search” because the participants apply search mechanisms to notice salient stimuli while most information is not consciously processed (Bowman et al., 2013). The same technique has been further tested under different conditions and stimuli, including participants’ names (Rosenfeld et al., 2006), celebrity names (Alsufani et al., 2021), and email addresses (Harris et al., 2021).

The use of EEG to detect concealed information has been highly successful but the equipment to record EEG, or any other aforementioned physiological responses, can be expensive and procedures are somewhat invasive and time-consuming, often making it impractical. Pupil responses, on the other hand, are easy to measure and have been correlated to several cognitive processes that could indicate concealed knowledge. The most recognized pupil reaction is known as pupil light response (PLR) – constriction and dilation of pupils in response to changes in light (Mathôt, 2018). However, pupils are also known to dilate in response to emotional arousal induced by pleasant or unpleasant stimuli, correlating to sympathetic nervous system activity (Bradley et al., 2008). Additionally, pupil dilation has been related to cognitive efforts, such as arithmetic calculations (Hess & Polt, 1964) and recalling learned digits (Kahneman & Beatty, 1966). These findings have been confirmed in further studies (Van der Wel & van Steenbergen, 2018).

Moreover, pupil dilation has been shown to reflect recognition of stimuli. For example, Võ et al. (2008) demonstrated that pupils dilate more when participants were exposed to familiar words, compared to unfamiliar words, and called this “pupil old/new effect”. Heaver & Hutton (2011) used this effect in a deception detection setting and demonstrated that pupil old/new effect occurred even when participants were instructed to lie about recognizing the old words. This indicates the robustness of the pupillary response to familiar stimuli, even when consciously trying to hide recognition. Similar results have been reported by other studies (Mill et al., 2016; Mathot, 2018; Dobbins, 2021), suggesting that pupillary responses could be used to identify individuals with crime-relevant knowledge when presented with CIT.

Indeed, pupil dilation has been successfully used in experiments to discriminate between individuals with specific knowledge and individuals without given knowledge in CIT. Lubow and Fein (1996) were able to correctly identify all the innocent participants and 50% to 70% of guilty participants in their mock-crime experiments based on pupil dilation. Seymor et al. (2013) found that the speed of pupil dilation can also be used for identifying individuals hiding information. They were able to identify 83% of participants who were lying about recognizing familiar faces based on pupil size and 70% of the lying participants based on the speed of pupil dilation. Moreover, all the participants who were truthful about recognizing the familiar faces were correctly identified using either measure. Thus, in addition to pupil dilation, the speed of change in pupil size may serve as a measure for concealed knowledge.

Given the substantial amount of research providing support for CIT, RSVP, and pupillary responses as a marker for concealed information, combining these three paradigms could be a feasible solution to test crime-related knowledge in a forensic setting. Currently, only one study reports the use of combined RSVP, CIT, and pupillary response. Chen et al. (2022) conducted two pupillometry experiments using the RSVP paradigm as it was used by Bowman et al. (2013).

The researchers observed a group-level significant pupil size difference in fake and real name conditions compared to the irrelevant name condition. However, on the individual level, the differences were statistically significant for three and six participants out of 26 and 31 participants, respectively. This is different from the study by Bowman et al (2013) who reported significant differences for all their participants in their EEG-based study. Chen et al. (2022) suggested the low proportion of individual differences could be attributed to the low power of the statistical test. While Harris et al. (2021) also reported a low proportion of individual differences in their EEG-based study, they proposed that participants might, for example, consciously or unconsciously search for the first letter of the target word and only process the rest of the word when the first letter matches the target word letter. It is possible that this kind of search mechanism was also used by the participants in the experiment by Chen et al. (2022). While combining CIT, RSVP and pupillometry offers great potential, further research and refinement are clearly needed.

The current study aims to generalize the findings from Chen et al. (2021) to non-personal stimuli in a mock crime scenario, providing higher ecological validity. Additionally, a meta-analysis by Ben-Shakar, & Elaad (2003) suggests that mock crime scenarios combined with CIT have the largest average effect size compared to CIT using card-test procedures or personal-item paradigm. The participants will take part in a mock crime where they will hide a crime-relevant item in a Dutch location, which will serve as the probe. The participants will be given another, crime-irrelevant location to focus on the following RSVP task; this location will serve as the target. The target and probe will be presented amongst other crime-irrelevant locations. We will measure participants' pupil response while they complete the RSVP task. We expect to detect probe recognition in the RSVP task by the change in size and speed of pupil dilation. We hypothesize that a) participants' pupil dilation will be larger when they see the probe compared to seeing an irrelevant word, b) participants' pupil dilates faster when they see the probe compared to seeing an irrelevant word and c) combining pupil size and pupil-slope will provide a better prediction of identifying whether participants recognize the probe or not, compared to using either of the pupil responses alone. We will use a time window of 600-1200 ms after the onset of the critical items as suggested by Chen et al. (2022) to measure the mean pupil size and the rate of change in the pupil size (derivative). We expect to observe these differences at both the group and individual levels. Furthermore, we aim to test the discriminability of our testing method by classifying participants based on individual analysis.

# 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**

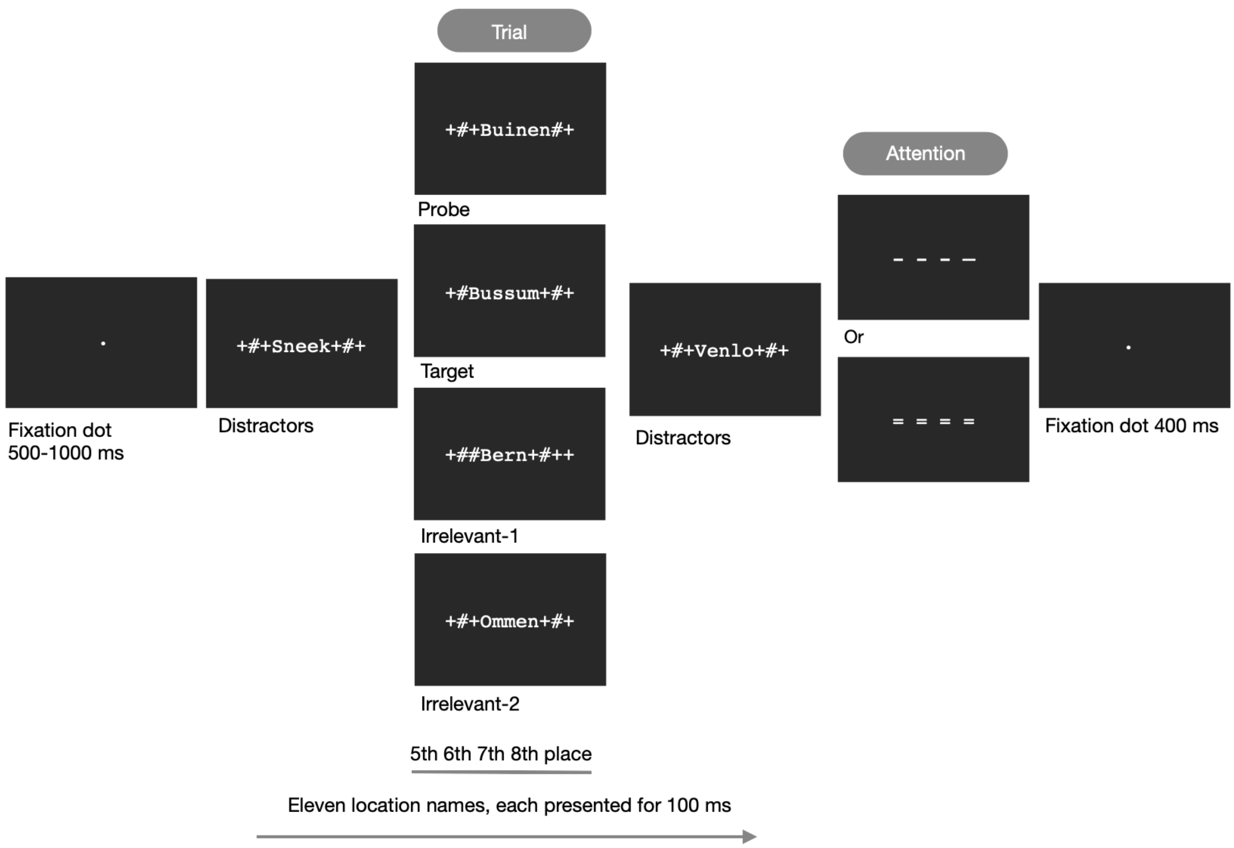
We used a within-participant experimental design. As part of a mock-crime scenario, an RSVP task was used, with a list of Dutch locations as stimuli (Klein, 2022; see OSF). From the list of 1105 Dutch locations, 15 were randomly chosen to be the potential real hiding location (probes) and removed from 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 ($8^{th}$) 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

**Figure 2a**

*Traces of Mean Pupil Size*

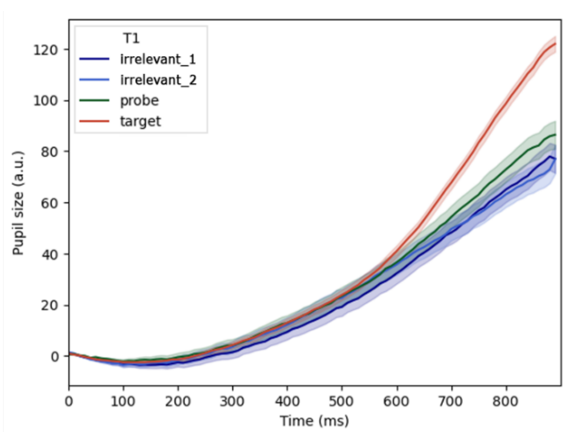
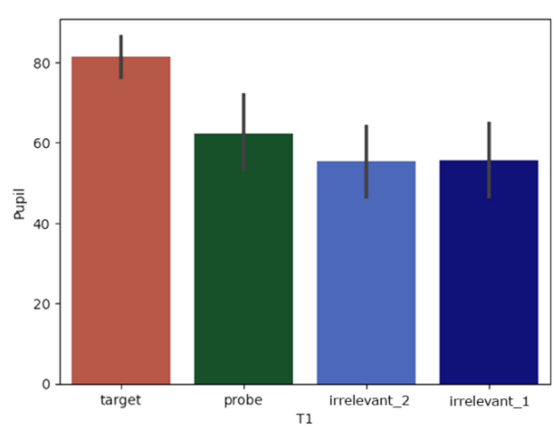



**Figure 2b**

*Mean Pupil Size for Each Condition*




*Note*. Time window from 600-900 ms.

**Figure 3a**

*Traces of Mean Pupil Derivative*

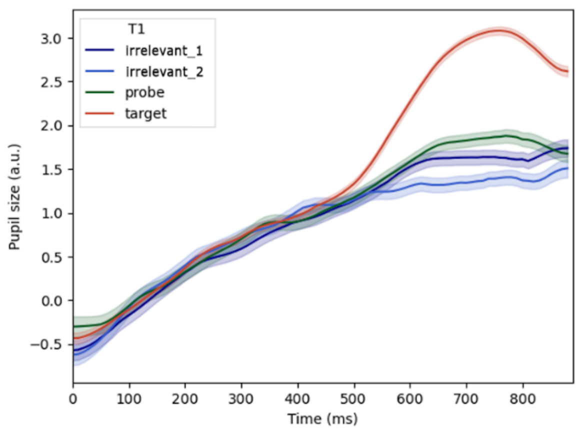
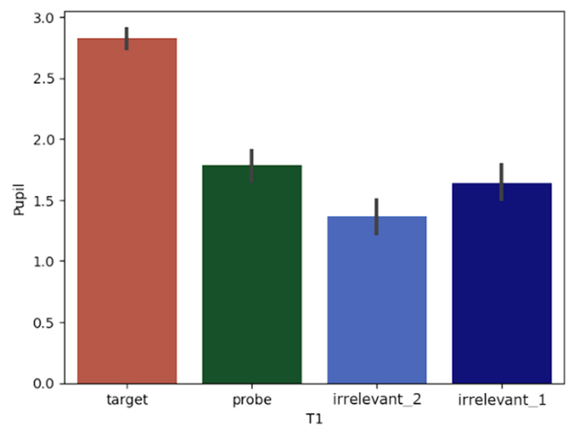



**Figure 3b**

*Mean Pupil Derivative for Each Condition*




*Note*. Time window from 600-900 ms.

summarized in Figures 2a and 2b for pupil size and in Figures 3a and 3b for pupil derivative. Specifically, Figures 2a and 3a illustrate the effects of the different conditions on the average pupil trace over time, while Figures 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

**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). $\alpha < .05$.

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.

**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**

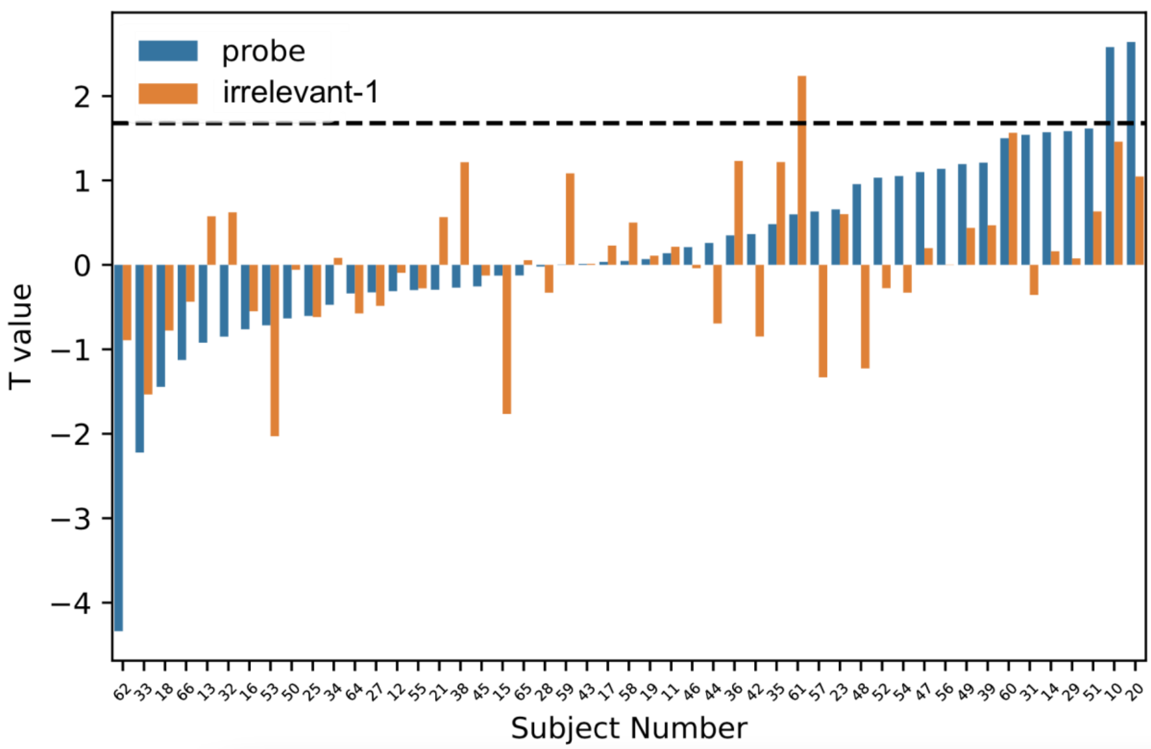
In the individual level analyses of mean pupil size, although 28 out of 49 participants had a positive *t*-value, only two participants showed a significant mean pupil size difference between the probe and irrelevant-2 condition. One participant showed a significant difference between the irrelevant-1 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.

## 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.

**Figure 4a**

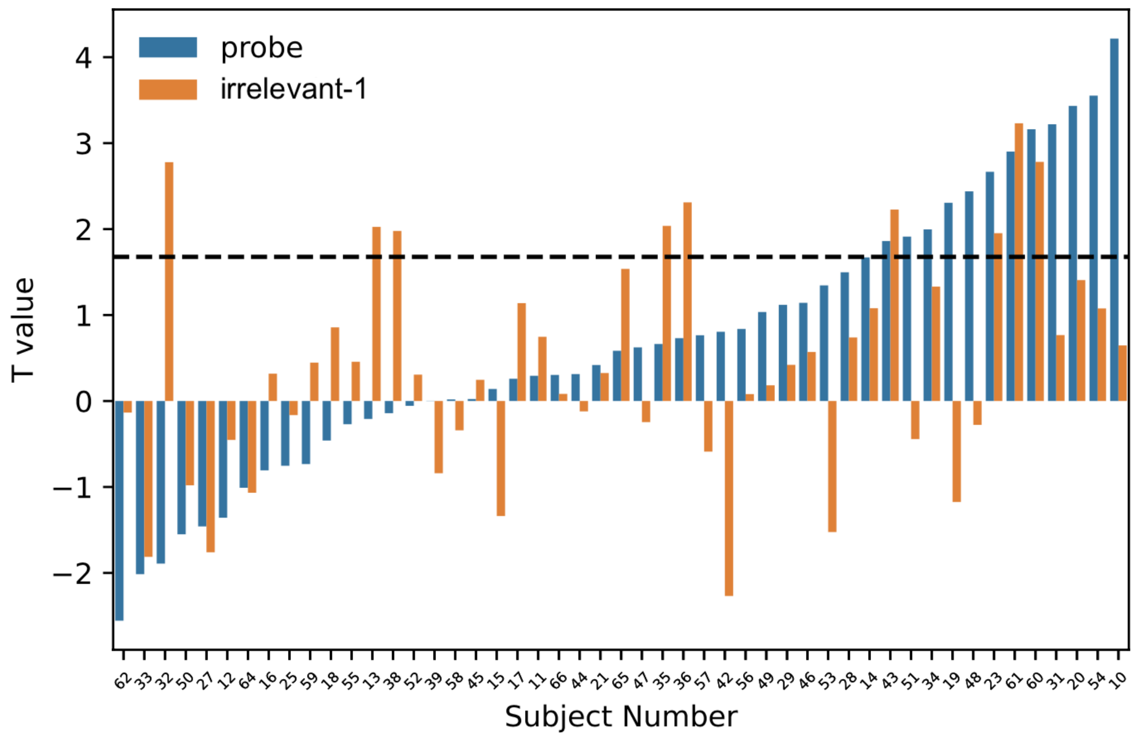
*Individual Mean Pupil Size Differences (Probe and Irrelevant-1 vs. Irrelevant-2)*



*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.

## Discussion

The goal of our study was to test the use of pupillometry in combination with CIT and RSVP to detect concealed knowledge, and to expand the findings from Chen et al. (2022) to non-personal stimuli. We expected there would be greater and faster pupil dilation in response to crime-relevant stimuli (probe) compared to crime-irrelevant stimuli (irrelevant-2) at both the group and individual levels. The group-level analysis revealed a significant difference between the probe and irrelevant conditions for the derivative but not for the pupil size. Unexpectedly, there was a significant difference in the derivative between the two irrelevant conditions, which confounds the findings between the probe and irrelevant-2 condition. Two participants showed a significant pupil size difference, and nine participants showed a significant slope difference. We classified two participants as guilty using mean pupil size, and eight participants were classified as guilty using the derivative. This results in .04 and .16 hit rates respectively. Additionally, four participants were classified as undetermined. We decided not to combine the pupil size and derivative factors for classification analysis since the discriminability of the pupil size was very low, compared to the derivative, and would not add much value to the combined predictor.

A noteworthy finding from our experiment is the significant difference in derivative between the probe and irrelevant conditions, however, it is important to bear in mind the differences were significant also between control conditions. One possible explanation for this could be the presentation of the probe, target, and irrelevant-1 items with the same starting letter. Although we aimed to counteract the use of a search mechanism proposed by Harris et al. (2021) it may have led to the effect of the irrelevant-1 items on the derivative. Additionally, our analysis was limited to a time window of 600-900 ms after the critical item due to the use of a short fixation dot at the end of the RSVP trials. Based on previous research, the preferred time window would range from 600 to 1200 ms (Mathôt, 2018; Chen et al., 2022). In the post-hoc analysis, we removed all trials with T1 presented in the $8^{th}$ position to analyze the effect of expanding our time window to 600 to 1000 ms. Using the extended time window resulted in a non-significant derivative difference between the irrelevant conditions, while the derivative difference between the probe and irrelevant conditions remained significant (Appendix C). This finding provides additional support for the sensitivity of the derivative.

The results of our experiment are less impressive than the results from Chen et al. (2022) who found a significant difference in pupil size between probe and control conditions, and a higher proportion of differences in individual analysis. Chen et al. (2022) and other previous studies (Rosenfeld et al., 2006; Bowman et al., 2013; Harris et al., 2021) combining CIT with RSVP have used more naturally salient stimuli, such as biographical information. The Dutch location names chosen for our study may not have reached that level of relevance for the participants. Despite our inclusion of a reaction task aimed at increasing the salience of the probe, it is possible that the information may have only reached the significance of incidental information. As was demonstrated by Rosenfeld et al. (2006), even when participants have been repeatedly exposed to information that is not personally relevant yet well-rehearsed the information did not elicit the same P300 response as did highly relevant information. With this in mind, the selection of the stimuli for CIT needs to be carefully considered in laboratory settings as well as in the field settings.

Our experiment was based on the theory that cognitive processing is reflected in pupil size, and relevant stimulus will result in larger pupil dilation (van der Wel & Steenbergen, 2018; Mathot, 2018), therefore, we also expected to see a difference in pupil size and slope between the task-relevant target condition and irrelevant-2 condition. This was not our main interest but serves as an auxiliary condition for the experiment. The analysis revealed a group-level significant difference in mean pupil size in the target condition, but individual results for the target conditions showed significant differences in pupil size only for 12 participants (Appendix E). Although pupil size is generally correlated to cognitive task performance, the direction and strength of this correlation vary between individuals (van der Wel & Steenbergen, 2018). If the pupil size to the task-relevant target stimuli did not change significantly, it is reasonable to assume that no effect would be present for the probe stimulus either, because of the individual differences in cognitive pupil reaction.

Neither pupil size nor derivative provided a satisfactory hit rate in our classification analysis. A combination of pupil measurements with other eye movements could result in a better discrimination factor for identifying participants with crime-relevant knowledge versus participants without such knowledge (Seymor et al., 2013). To improve classification further, future studies could add a control group to the experiments and test the balance of type I and type II error rates using the receiver operator characteristic (ROC) curve to ensure high discriminability between guilty and innocent participants (Hanley & McNeil, 1982).

Using pupillometry in combination with CIT and RSVP to detect concealed knowledge has not been researched extensively and the results remain non-conclusive. While the results of Chen et al. (2022), as well as other studies, support the use of pupil dilation to detect concealed information or recognition (Lubow & Fein, 1996; Võ et al., 2008; Seymor et al., 2013; Dobbins, 2021; Selle et al., 2022), the effect of the stimuli and the individual differences in pupil reaction using RSVP paradigm need more investigation. It may be that pupillometry combined with CIT and RSVP is suitable for detecting certain types of concealed information, for example, autobiographical information, but perhaps not for others. Moreover, modifications to the RSVP stimuli, such as incorporating multiple probes, could lead to better results. Lastly, we recommend investigating the derivative as a predictor for identifying concealed information in combined CIT and RSVP tests in future studies as the derivative might be more sensitive to the RSVP stimuli than pupil size.

In conclusion, we have discovered derivative as a potential highly sensitive pupil reaction for identifying concealed information in combination with RSVP. Using CIT with RSVP and pupillometry is a novel idea and while promising, the current paradigm is not efficient enough to be used in forensic investigations. More research is needed to discover which conditions, measurements, and analyses provide the most reliable results.

**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. *Journal of Applied Psychology*, *88*(1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131

Ben-Shakhar, G. (2012). Current research and potential applications of the Concealed Information Test: An overview. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00342

Berrien, F. K., & Huntington, G. H. (1943). An exploratory study of pupillary responses during deception. *Journal of Experimental Psychology*, *32*(5), 443–449. https://doi.org/10.1037/h0063488

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. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18*(3), 307–315. https://doi.org/10.1111/j.1469-8986.1981.tb03040.x

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/BF0321049

Chen, I. Y., Karabay, A., Mathot, S., Bowman, H., Akyürek, E., G. (2022). Concealed identity information detection with pupillometry in rapid serial visual presentation. *Psychophysiology, 60*(1), 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

Dobbins, I.G. Pupil dilation signals recognition salience. *Psychonomic Bulletin & Review, 28*, 565–573 (2021). https://doi.org/10.3758/s13423-020-01866-w

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (“lie detection”) with event-related brain potentials. *Psychophysiology, 28*(5), 531–547. https://doi.org/10.1111/j.1469-8986.1991.tb01990.x

Gardner, R. M., Mo, S. S., & Borrego, R. (1974). Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory. *Bulletin of the Psychonomic Society, 3*(3), 237–238. https://doi.org/10.3758/BF03333458

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, 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

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19*(4), 398–405. https://doi.org/10.1080/09658211.2011.575788

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

Iacono, W. G. (2011). Encouraging the use of the guilty knowledge test (GKT): What the GKT has to offer law enforcement. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test.* Cambridge University Press

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

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

klein Selle, N., Suchotzki, K., Pertzov, Y., & Gamer, M. (2022). Orienting versus inhibition: The theory behind the ocular-based Concealed Information Test. *Psychophysiology, 6*(3), e14186. https://doi.org/10.1111/psyp.14186

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

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

Mathot, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 16. https://doi.org/10.5334/joc.18

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

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

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology, 51*(9), 879–904. https://doi.org/10.1111/psyp.12239

Mill, R. D., O'Connor, A. R., & Dobbins, I. G. (2016). Pupil dilation during recognition memory: Isolating unexpected recognition from judgment uncertainty. *Cognition*, *154*, 81–94. https://doi.org/10.1016/j.cognition.2016.05.018

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.

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., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience, 34*(1-2), 125–129. https://doi.org/10.3109/00207458708985947

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

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*(3), 251–259. https://doi.org/10.1016/j.ijpsycho.2005.06.002

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.

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*, *3*. https://doi.org/10.3389/fpsyg.2012.00614

Synnott, J., Dietzel, D., & Ioannou, M. (2015). A review of the polygraph: History, methodology and current status. *Crime Psychology Review, 1*(1), 59-83. https://doi.org/10.1080/23744006.2015.1060080

Van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review 25*(6), 2005-2015. https://doi.org/10.3758/s13423-018-1432-y

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2007). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45*(1), 130-140. https://doi.org/10.1111/j.1469-8986.2007.00606.x

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 |

| Subject_nr | Probe derivative | Control derivative | Target derivative | Probe mean pupil size | Control mean pupil size | Target mean pupil size |
|---|---|---|---|---|---|---|
| 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* |
| 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 |

| Subject_nr | Probe derivative | Control derivative | Target derivative | Probe mean pupil size | Control mean pupil size | Target mean pupil size |
|---|---|---|---|---|---|---|
| 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 |

| Subject_nr | Probe derivative | Control derivative | Target derivative | Probe mean pupil size | Control mean pupil size | Target mean pupil size |
|---|---|---|---|---|---|---|
| 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 |
| 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 |

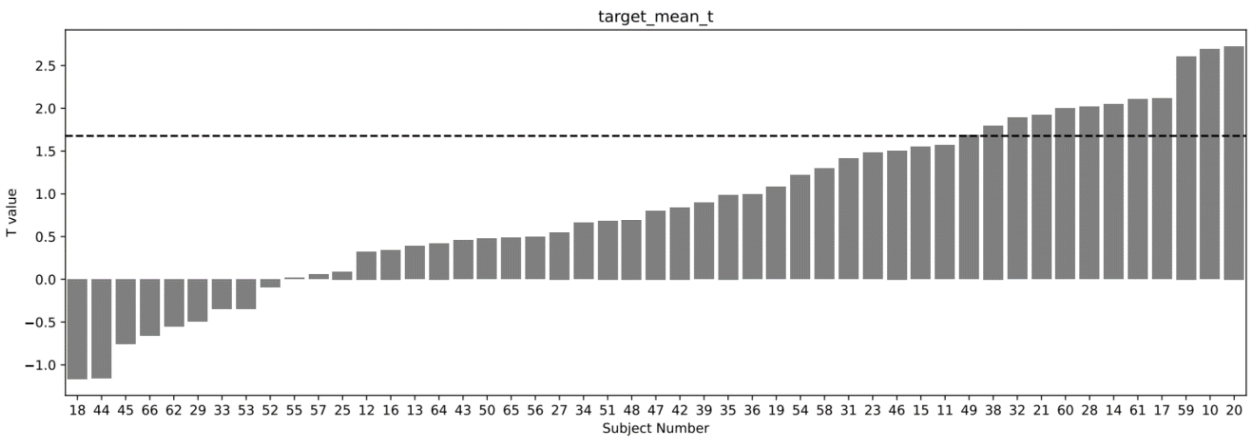
| Subject_nr | Probe derivative | Control derivative | Target derivative | Probe mean pupil size | Control mean pupil size | Target mean pupil size |
|---|---|---|---|---|---|---|
| 64 | -1.01 | -1.07 | -0.45 | -0.34 | -0.58 | 0.42 |
| 65 | 0.58 | 1.54 | 5.25** | -0.13 | 0.06 | 0.48 |
| 66 | 0.30 | 0.08 | 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

**Figure 7**

*Individual Mean Pupil Size Differences Between Target and Irrelevant-2 Conditions*




*Note. t*-tests were performed for each participant at $t = 1.677$ to compare mean pupil sizes between the target and irrelevant-2 conditions. Subject numbers begin at 10 as the first participant.

## Appendix D

**Table 4**

*Linear Mixed Effect Analysis for Derivative Post-Hoc*

| Condition | Coef (SE) | *z* | *p* |
|---|---|---|---|
| Intercept | 1.41 (0.15) | 9.59 | < .001 |
| Irrelevant-1 | 0.15 (0.11) | 1.32 | .188 |
| Probe | 0.36 (0.15) | 2.38 | .017 |
| Target | 1.32 (0.14) | 9.32 | < .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.