Do your eyes tell the truth? Detecting concealed information using RSVP and pupillometry

Danie Hettema

S4348516

Department of psychology, University of Groningen

PSB3N-BT15: Bachelor Thesis (Nederlands) (1b)

Group number: 2223_1b_02

Supervisor: Dr. Robbert van der Mijn

Second evaluator: MSc. Ana Vilotijevic

In collaboration with: Charlotte Bosch, Anouk Feijen,

Tobias Kroner, Teresa Nguyen, Janette Possul

14th of April 2023

Summary

This study uses Rapid Serial Visual Presentation (RSVP) combined with a Concealed Information Test (CIT) to detect concealed information using pupillometry. In this experiment, participants (N = 49) chose a hiding location (probe). They received another location and pretended it was theirs (target). This study wants to determine if mean pupil size and derivative can be used as classifiers of concealed information. The results show no significant difference between the probe and control conditions for mean pupil size in a time window of 600-900 ms after the onset of the critical stimulus. The target is, as expected, significantly larger than the control conditions. For pupil derivatives both the probe and the target conditions are significantly larger than the control conditions. Both control conditions, however, also differ significantly, making it difficult to interpret these results. On the individual level, mean pupil size and pupil derivative correctly classified respectively 4% and 16% of the participants. A post hoc analysis (N = 46) on a wider time window of 600-1000 ms shows that for pupil derivatives both the target and probe are significantly larger than the control conditions, but the control conditions no longer differ. These results lead us to conclude that mean pupil size can not be used as a classifier of concealed information. Derivative pupil size might be a good classifier on a wider time window. Future research must show whether the wider time window is useful and should assess derivative pupil size more often as a possible classifier.

Keywords: Concealed Information Test (CIT), Classification, Mean pupil size, Pupil derivative, Rapid Serial Visual Presentation (RSVP)

Do your eyes tell the truth? Detecting concealed information using RSVP and pupillometry

There are many situations in which people might want to conceal the truth. You can think about crime situations in which the culprit wants to hide incriminating knowledge. Another example, which is a little less serious, is when you are planning to surprise a friend. They can obviously not know about it, so sometimes you have to conceal your true intentions. The first example is from a societal perspective more severe and unwanted than the second, socially acceptable one (Saxe, 1991). It raises the question: how to combat lying in severe contexts, such as criminal situations?

That is where lie detection comes in. Over the years there have been different forms of lie detection, such as the polygraph (Lykken, 1959). The Concealed Information Test (CIT) (Lykken, 1959) is another instrument to determine whether a person has knowledge of a certain word or image. The CIT works by measuring physiological responses to stimuli. Some of these stimuli might be related to the information a subject tries to hide (Ben-Shakhar & Elaad, 2003). If the participant has any hidden knowledge about a stimulus, a physiological response will be detectable and measurable (Verschuere et al., 2011). A few examples of these physiological responses are change in relative blood pressure, skin conductance, and event-related potentials (ERP’s) (Ben-Shakhar & Elaad, 2003). These physiological responses are caused by the autonomic nervous system. The emotional effect of lying and concealing information is reflected in these physiological responses (Cutrow et al., 1972), but we will purely focus on the recognition of salient stimuli.

When looking at the brain, it is known that there is an increase in activity at the moment someone is concealing information about a presented salient stimulus. One measure of brain activity is the P300-response. This is a large ERP typically measurable 300-400 ms after a significantly salient stimulus was shown (Nieuwenhuis et al., 2011). Recent studies have shown that these P300-responses are related to responses of the locus coeruleus-norepinephrine system, which are also related to pupil dilations (Gilzenrat et al., 2010). Examples of salient stimuli are the name or a picture of someone you know. Alsufyani et al. (2021), for example, used names of famous people for this.

By using EEG to compare P300-responses, you can try to determine whether a subject is concealing information (Rosenfeld et al., 2013). When someone is presented with a salient stimulus, a larger P300-response is elicited than when a non-salient stimulus is presented. This response therefore reflects the recognition of the stimulus which might be information that the subject is trying to hide.

Since the use of lie detection and the CIT, there has been a search for ways to improve the results of these tests and make them more reliable. The CIT is, after all, quite susceptible to countermeasures (Rosenfeld et al., 2004; Peth et al., 2016). Countermeasures, such as inflicting pain on yourself, change or activate a physiological response and therefore change the outcome of the test. Rapid Serial Visual Presentation (RSVP) might counter that problem. RSVP is a method in which stimuli are presented to subjects on the fringe of awareness (Potter, 1976). Each stimulus is presented for approximately 100 ms (Broadbent & Broadbent, 1987; Chen et al., 2022) and is directly followed by the next stimulus. By doing so, the participant has not enough time to use top-down cognitive control to actively process what kind of stimuli they perceive. Therefore, they do not have enough time to decide whether to use a countermeasure to hide their response. This makes it impossible to use countermeasures, other than easily detectable ones (Broadbent & Broadbent, 1987; Bowman et al., 2013).

Bowman et al. (2013) used RSVP in combination with the CIT. They studied the brain's P300-response to an RSVP task containing names. Participants had to pick a “fake” name: a different name than the participants' own name. This became the target: a stimulus the participant is actively searching for. The participants' own and thus real name is the probe:

a stimulus the participant is not searching for, but which is relevant to the participant. In the experiment of Bowman et al. the participants were instructed that their real name could show up in an RSVP stream, but that they should hide a reaction to it. Targets and probes were presented in a stream of distractors. These distractors should hold no salient value for the participants. Some of these distractor words were used as a control condition to compare to the target and probe. The probe, target and control words, are called critical items, or T1-stimuli. Bowman et al. (2013) found that probes elicited a significantly higher P300-response than control names. This effect was found, even whilst the participants were told to conceal a reaction to the probe. This effect was expected since the probe is salient for the participants.

This research paper focuses on CIT and the recognition of salient stimuli. Whilst recognizing familiar information, an increase in pupil dilation is expected, due to the increase in recognition memory processes (Maw & Pomplun, 2004, as cited in Otero et al., 2011). Therefore, an even greater dilation can be expected when lying about or actively hiding information in comparison to only processing unknown stimuli, since the lying also increases one's arousal and thus psychosensory response (Mathôt, 2018).

The use of pupil responses to classify whether or not people are lying is not a new subject. Lubow and Fein (1996) found in their two experiments that they could correctly detect approximately 50-70% of the participants who were hiding information and a 100% of the participants who were not hiding anything.

A more recent study of Chen et al. (2022) focused on pupil dilation as a possible indication for concealed information. Their results show that it is possible to use RSVP and pupillometry to classify whether someone is concealing information at a group level. Chen et al. found that the pupil size of their probe condition was significantly larger than the control condition. This effect was, however, not significant for most participants individually. Chen et al (2022) also found that a time window of 600-1200 ms after the onset of the critical item seems to be the most optimal time window to assess differences in pupil size. A bachelor thesis study by Göl et al. (2022) replicating Chen et al. (2022) found a narrower time window of 640-920 ms to be optimal.

The aim of this study is to conceptually replicate the study of Chen et al. But aside from only using mean pupil size, we will also use pupil-slope. This is because Seymour et al. (2012) showed that pupil-slope also produces an efficient way to classify someone as concealing information.

Research question

Combining the information from previous studies, this study analyzes if concealed information in a mock crime scenario can be detected using RSVP and pupillometry. Since the use of RSVP in combination with pupillometry is more practical than EEG, it is important to determine if it is a suitable alternative for it. The previous studies from Lubow and Fein (1996) and from Chen et al. (2022) show promising results about using RSVP and pupillometry. By fine tuning it, it might one day be used to detect concealed information in police or other forensic situations.

Hypotheses

We hypothesize that on both group and individual level: A) mean pupil size will be significantly larger in probe trials than in control trials during a time-window of 600-900 ms after the onset of the T1-stimulus. B) The change in pupil size will be significantly steeper in probe trials than in control trials during a time-window of 600-900 ms after the onset of the T1-stimulus. C) Combining the measures “mean pupil size” and “pupil size derivative” will be a better predictor to classify participants as “guilty” or “innocent” than both measures apart. D) We expect no differences in the two control conditions for both pupil measures.

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/] 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

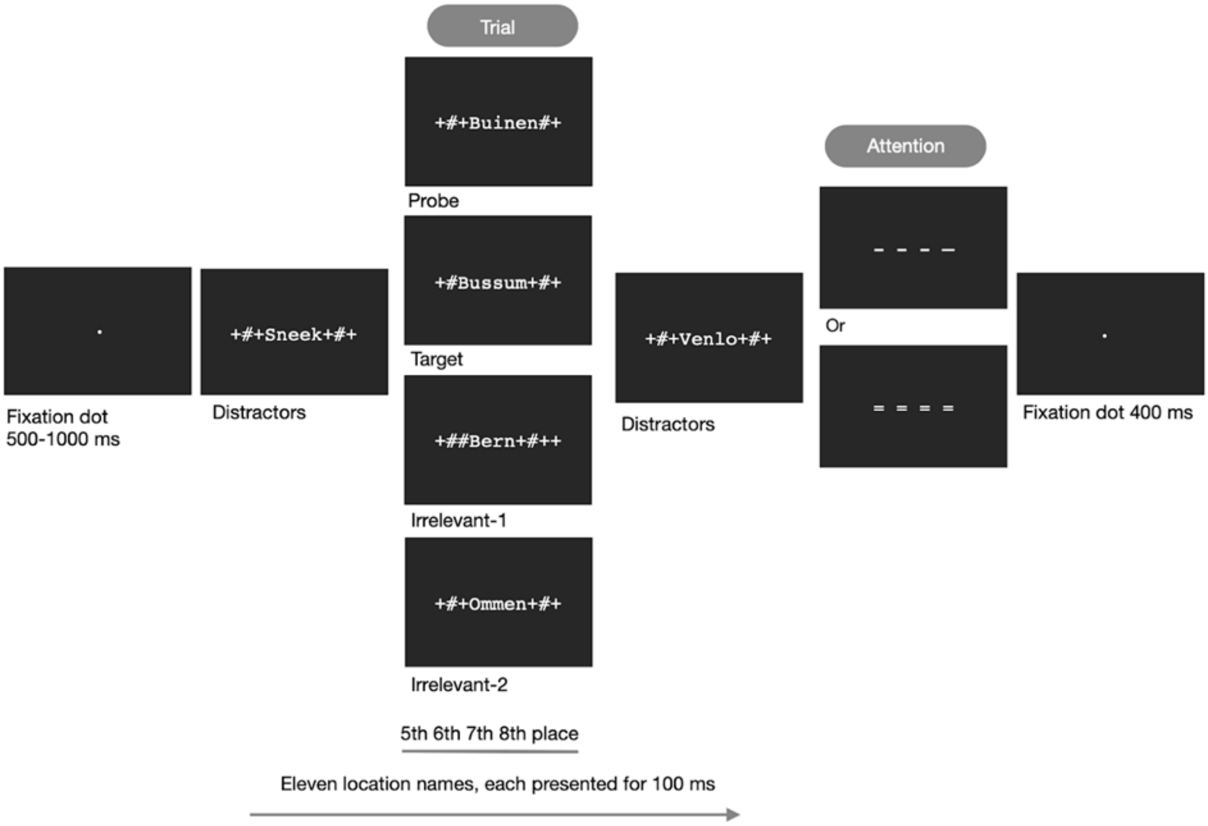
We used a within-participant experimental design. As part of a mock-crime scenario, an RSVP task was used, with a list of Dutch locations as stimuli (Klein, 2022; see OSF). From the list of 1105 Dutch locations, 15 were randomly chosen to be the potential real hiding location (probes) and removed from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

Figure 2a

Traces of Mean Pupil Size

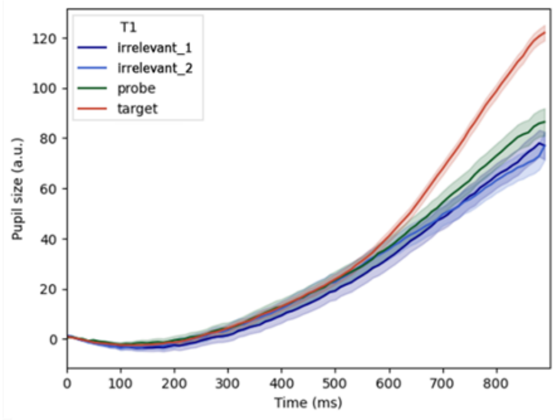
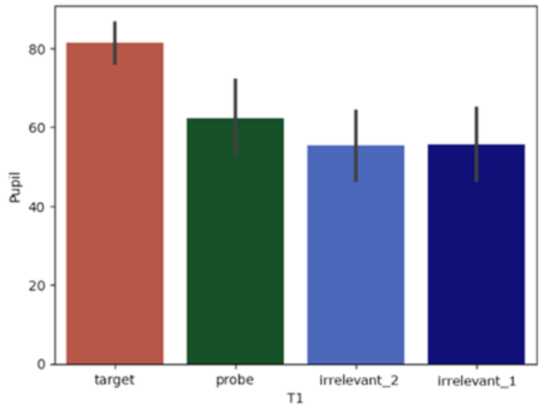


Figure 2b

Mean Pupil Size for Each Condition



Note. Time window from 600-900 ms.

Figure 3a

Traces of Mean Pupil Derivative

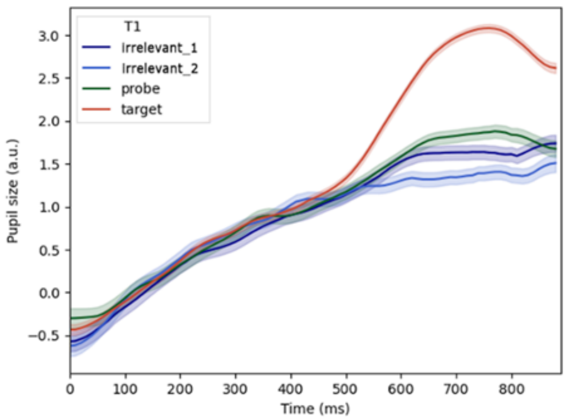
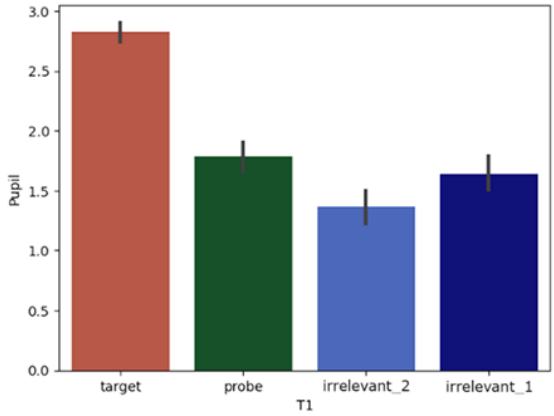


Figure 3b

Mean Pupil Derivative for Each Condition



Note. Time window from 600-900 ms.

Mean pupil size

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

Table 1

Linear Mixed Effects Analysis for Mean Pupil Size

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

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

Derivative

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

Table 2

Linear Mixed Effects Analysis for Derivative

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

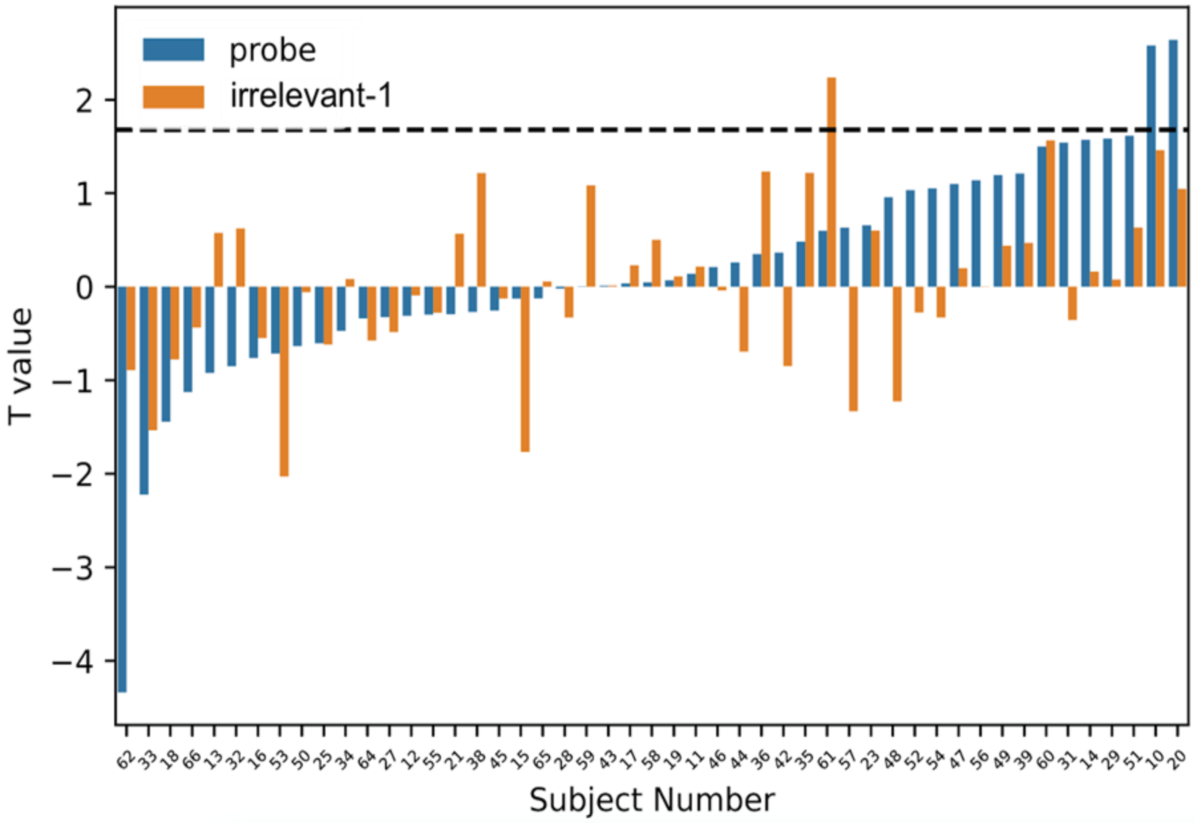
Note. $N = 6876$ (trials). $\alpha < .05$.

Individual level analysis

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

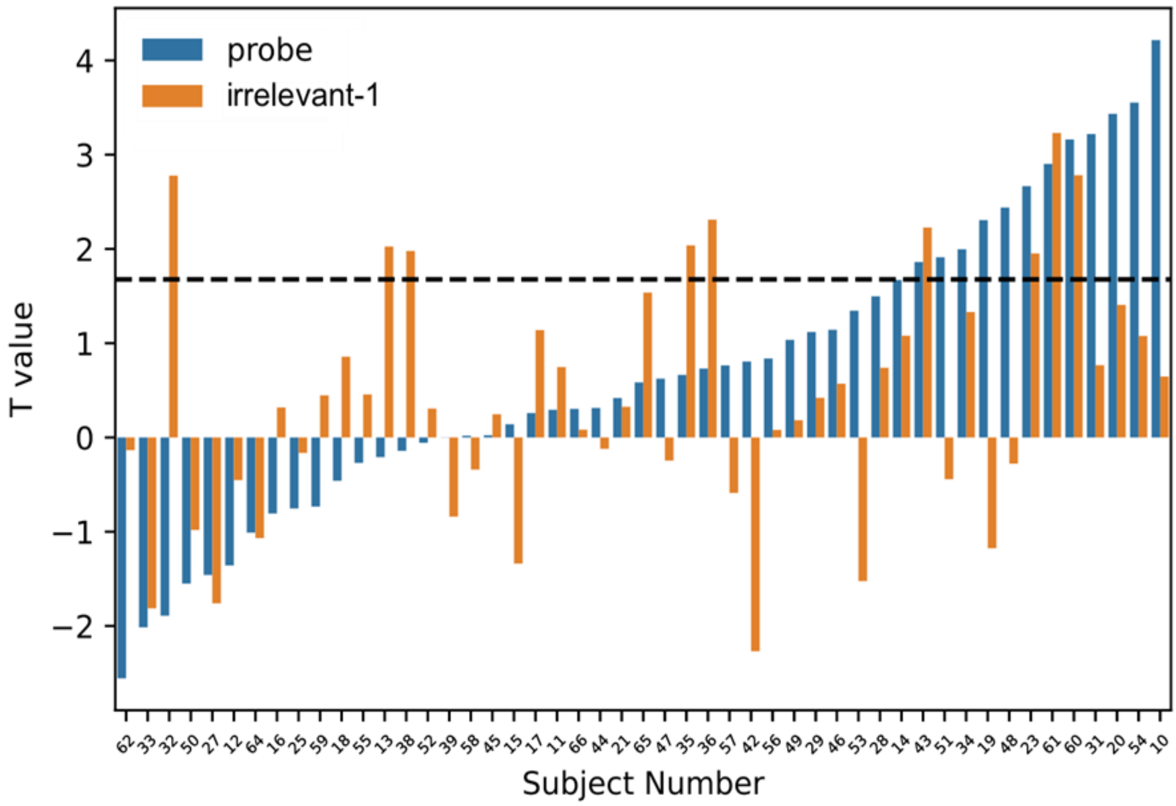
Individual Mean Pupil Size Differences (Probe and Irrelevant-1 vs. Irrelevant-2)



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

The goal of this study was to determine if pupillometry can be used to detect concealed information in a CIT paradigm where the stimuli are presented on the fringe of awareness. Based on the results, we conclude that the mean pupil size is not significantly larger in probe trials than in irrelevant-2 trials ($z = 1.09$, $p = .275$). Therefore it is questionable if mean pupil size is usable as a good predictor for the detection of concealed information in this setting. As expected, there was no significant difference between the irrelevant-2 and irrelevant-1 conditions for this measure. These findings lead us to reject hypothesis A: an expected significant difference between the probe and irrelevant-2 for mean pupil size.

For our second hypothesis, we found that the derivative of the pupil size was significantly larger in probe trials than in irrelevant-2 trials ($z = 2.92$, $p = .004$). For this measure there was, however, also a significant difference between irrelevant-2 and irrelevant-1 trials. This makes it questionable what causes the significant difference between probe and irrelevant-2 conditions, since it might also be due to extreme sensitivity of this measure. These findings lead us to accept hypothesis B: an expected significant difference between the probe and irrelevant-2 for pupil derivative. However, the findings also lead us to reject hypothesis D: no expected differences between the two control conditions.

Based on the results, we reject our third hypothesis: that mean pupil size and pupil derivative together would be a better classifier of concealed information than both measures apart. The mean pupil size had a 4% hit rate and the derivative pupil size had a 16% hit rate. Every participant that was classified with the mean pupil size was, however, also classified though the derivative pupil size. Therefore, we conclude that we should not add both measures together, since mean pupil size does not add any new information.

Lastly, responses to target trials show a clear significant difference with irrelevant-2 trials for both mean pupil size ($z = 5.08$, $p < .001$) and derivative pupil size ($z = 10.48$, $p < .001$). This was as expected due to the nature of the task: actively searching for the target. This confirms earlier studies pertaining to the locus coeruleus-norepinephrine system and its correlation with pupil dilation (Gilzenrat et al., 2010; Nieuwenhuis et al., 2011).

Post Hoc analysis

As stated in the method section, we could not analyze pupil data after 900 ms after the onset of the T1-stimulus. To check whether this had an effect on the results, we performed a post-hoc analysis in which all the trials where T1 was presented at position eight were excluded. By doing so, we could analyze up to 1000 ms after the onset of T1. This is in between the recommendations of Chen et al. (2022) and Göl et al. (2022).

For the post hoc analysis on group level, linear mixed effects analyses were carried out for the derivative of pupil size and the mean pupil size. The results are presented in Appendix C. The derivative of pupil size was significantly larger in the probe condition than in the irrelevant-2 condition. Furthermore, irrelevant-1 and irrelevant-2 did not significantly differ anymore. This combined shows us that we could use the derivative of the pupil size at a time window from 600 to 1000 ms to possibly detect concealed information on a group level. This furthermore acknowledges the results of Chen et al. (2022) that the most optimal time window is longer than 600 to 900 ms.

For the analysis on individual level, t-tests were computed for all participants' mean pupil sizes and derivatives. For both measures we compared the probe to irrelevant-2, and irrelevant-1 to irrelevant-2. There was, however, not much of a difference there.

Based on these results, we can conclude that within the time window of 600 to 1000 ms, the derivative shows a significant effect for the probe on the group level ($z = 2.80$, $p =$

0.005). This makes this time window better suited to analyze the derivative on the group level. Not much changed on the individual level and that implies that we should still refrain from using mean pupil size as a classifier, since it still holds too little information and since pupil derivative is a more sensitive measure.

Recommendations

To improve on this research paper, we recommend trying to gather a broader sample. Our sample consisted of only native Dutch students. This is problematic for the generalizability of these results. It is, however, expected that the results would be similar for everyone, since our pupil measures are biological processes, which are assumed to work mostly similarly in all humans as explained by Nieuwenhuis et al. (2011).

It would also be recommended to analyze pupil traces for longer than 900 ms after the onset of the T1-stimulus. Due to our last fixation dot of only 400 ms, we could only analyze a time window with a maximum of 900 ms after the onset of the T1-stimulus. This is problematic since Chen et al. (2022) recommended a time window of 600 until 1200 ms after the T1-stimulus and since Göl et al. (2022) recommended a time window of 640 until 920 ms after the onset of the T1-stimulus. Since our last fixation dot was not presented long enough, we have not collected pupil data after this 900 ms limit for trials in which the T1-stimulus was positioned at the 8th position. This might have caused non-significant differences between the probe and control conditions for our pupil measures at the time window of 600 to 900 ms. The post hoc analysis showed that it is likely that the duration of our time window was indeed problematic, since the wider time window of 600 to 1000 ms shows a significant difference for probe and target compared to irrelevant-2 and not for irrelevant-1 compared to irrelevant-2.

Another recommendation would be to assess the salience level of the probe. We made use of a story and practice task to make the probe as salient as possible. We did, however, not assess how salient the probe became. Therefore, the non-significance on the individual level might be because of participants who forgot and therefore not recognized the probe. Rosenfeld et al. (2006) also found that classification accuracy of pupil measures was greater when stimuli with a greater salience were used, such as autobiographical items (own names, own living place, etc.). Instead of using an autobiographical item as a probe, we used an incidentally acquired probe (Rosenfeld et al., 2006). Such an incidentally acquired probe is more realistic for a forensics setting, since not all criminal information is autobiographical. This might, however, have led to a lower salience level and therefore these hard to interpret results.

Future experiments could try prompting participants with a series of questions at the end of the experiment regarding their chosen probe. As of now, we cannot conclude that everyone actually saw the probe or even remembered it during the experiment. To combat this, one can ask questions like: “What was your chosen hiding location?” Another question could be: “Did you see [probe]?” This question can be used to assess if people were on a conscious level able to perceive the probe. Or that they might have reacted solely to, in our case, the same starting letter as the target.

We also recommend carefully selecting the T1-stimuli and to make sure the stimuli are not too similar. We have used the same starting letter for our target, probe and irrelevant-1 to combat search strategies. This was an improvement on earlier papers. This similarity, however, might have caused the significant difference between irrelevant-1 and irrelevant-2. Participants possibly did not see and recognize the probe or target, but might have thought they saw it, because it had the same starting letter. This is, however, not supported by the accuracy scores on the question “did you see the target?” We removed participants with an accuracy lower than 50% for the target condition. After removing those participants, none of the remaining participants had an accuracy lower than 50% on the combined non-target condition. This makes it unlikely that participants got confused because of the same starting letter.

Lastly, we recommend creating an experiment in which participants can be assigned to both a “guilty” and "innocent" group. Our experiment had only “guilty” participants. Therefore, we could not assess a false positive or true negative rate for our pupil measures.

Despite the possibilities for improvement, our study still grants valuable information. The derivative of pupil size is something to further assess, as this shows promising results, but is not investigated often yet. As of now we can conclude that using RSVP in a CIT paradigm might be useful as a substitute for EEG, but further research is needed to fine tune the RSVP CIT paradigm and to assess what measures create the most optimal classification result.

References

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

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

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

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

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

Cutrow, R. J., Parks, A., Lucas, N., & Thomas, K. (1972). The Objective Use of Multiple Physiological Indices in the Detection of Deception. *Psychophysiology*, *9*(6), 578–588. https://doi.org/10.1111/j.1469-8986.1972.tb00767.x

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

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

Göl, S., Jansen, C., & Rasztar, F. (2022). *Measuring Pupil Dilation in a Concealed Information Test Using Rapid Serial Visual Presentation* [Unpublished undergraduate thesis]. University of Groningen.

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

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

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

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

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

Mathot, S. (2023). *python-eyelinkparser: An extensible parser for EyeLink data files (EDF)* (0.17.3) [Python]. https://github.com/smathot/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

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

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

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

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

Rosenfeld, J. P., Ben-Shakhar, G., & Ganis, G. (2013). Detection of Concealed Stored Memories with Psychophysiological and Neuroimaging Methods. *Memory and Law*, 263–303. https://doi.org/10.1093/acprof:oso/9780199920754.003.0011

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

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

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

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

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

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

Appendix A

Figure 5

Map of Potential Probes

Appendix B

Table 3

Individual t*-values for Each Condition*

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

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

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

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

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

Appendix C

Table 4

Derivative Pupil Size Analyzed for Time Window 600-1000 ms

	Coef. (SE.)	z	p
Intercept	1.41 (0.15)	9.59	< 0.001
Irrelevant_1	0.15 (0.11)	1.32	0.188
Probe	0.36 (0.15)	2.38	0.017
Target	1.32 (0.14)	9.32	< 0.001

Note. This table contains data of a Post-Hoc analysis. Irrelevant_2 is the intercept. In this table the data no longer contains the trials where the T1-position was the 8th place. N = 46. 11 participants were excluded: 6 for missing data, 5 for accuracy.

Table 5

Mean Pupil Size Analyzed for Time Window 600-1000 ms

	Coef. (SE.)	z	p
Intercept	60.41 (11.02)	5.48	< 0.001
Irrelevant_1	-6.28 (7.33)	-0.86	0.391
Probe	4.44 (7.52)	0.59	0.555
Target	29.04 (6.17)	4.71	< 0.001

Note. This table contains data of a Post-Hoc analysis. Irrelevant_2 is the intercept. In this table the data no longer contains the trials where the T1-position was the 8th place. N = 46. 11 participants were excluded: 6 for missing data, 5 for accuracy.