Using Pupillometry in a RSVP Task Can Detect Concealed-Information

Charlotte V. Bosch

S3589889

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

2223-1b-02

Supervisor: dr. Robbert van der Mijn

Second evaluator: Msc. Ana Vilotijevic

In collaboration with: Danie Hettema, Janette Possul, Tobias Kroner, Anouk Feijen and Teresa Nguyen

14 April 2023

Abstract

The Concealed Information Test presents examinee’s with a series of items that are relevant and irrelevant to a crime, to investigate whether this person is in possession of incriminating knowledge. This test relies on the body's physiological response to such items. However, which physiological response can be a reliable measure of concealed information has been questioned, because of the countermeasures that examinee’s might use. Rapid Serial Visual Presentation (RSVP) in combination with pupillometry has thus been explored as a paradigm to counteract those countermeasures. Previous studies have shown that pupil size can reveal concealed information using this paradigm. Since pupillometry shows to be a promising novel way, which is non-invasive and cost efficient, this study will try to replicate those findings and improve on its usability, with different measures of accuracy and different salient stimuli. In an experiment 57 university students were part of a mock crime. Participants had hidden a suitcase at a chosen place, which they were instructed to conceal. They were then given a name of a place to search for in an RSVP stream. Apart from these names control names also appeared in the stream while their pupil sizes were recorded. Our results show that mean pupil size does not reveal concealed information at group level. However, the pupil slope showed a more promising measure for concealed information at group and individual level. Further refinement of the experiment is necessary to create a more reliable test for concealed information.

Keywords: concealed information test, rapid serial visual presentation, pupillometry, locus coeruleus

Using Pupillometry in a RSVP Task Can Detect Concealed-Information

The body's physiological responses have long been used to detect deception and thereby reveal concealed information (Gamer, 2011). The information someone conceals can be anything, such as one's true identity or a hiding location. One of the settings where concealed information is of great interest is the forensic setting, in which it is essential to find out the truth about a certain crime. And thus, this is one of the most important settings in which concealed information testing can be applied (Matsuda et al., 2012).

One of such tests that has been created is the Guilty Knowledge Test, better known as Concealed-Information Test (CIT) (Lykken, 1959). The CIT is a test in which participants are presented with a series of items. The CIT is based on the idea that a participant will have a different physiological response to items presented that are relevant to them than to those that are irrelevant (Krapohl et al., 2009; Lykken, 1959). Originally, this test was done in combination with the galvanic skin response (GSR) to detect the differences in physiological responses to the relevant and irrelevant stimuli (Lykken, 1959).

Aside from the use of CIT in combination with a polygraph, researchers have been combining the CIT with other measurements such as electro-encephalography (EEG) (Harris et al., 2021; Rosenfeld et al., 1987), functional magnetic resonance imaging (fMRI) (Langleben et al., 2016), functional near-infrared spectroscopy (fNIRS) (Sai et al., 2014) and the pupillary response to measure difference in brain activity (Chen et al., 2022; Gamer, 2011). In an analysis carried out by Meijer et al. (2016), results from different studies were used to compare paradigm detection accuracy. Their analysis showed values ranging from *a* .72 to .98 (*a* = area under the receiver operating characteristic curve) using CIT in combination with skin conductance rate, reaction time, event related potentials (ERP) and fMRI. These results can be interpreted as to how well the paradigm can discriminate between innocent and guilty suspects (higher than .5 being above chance) (Meijer et al., 2016).

Although the previously mentioned ways of measuring physiological responses in the CIT paradigm have shown some success, the validity of CIT is questioned. This is because of the influence the examiner's behavior can have on the subject's response. For example behavior such as change in voice tone or stature when presenting the relevant stimuli (Davies & Beech, 2018). Second, the potential use of countermeasures is a predominant downside of this test: guilty suspects might for example inflict pain on themselves to keep their physiological responses high for all items (Davies & Beech, 2018; Peth et al., 2016). Because of the potential use of countermeasures and how it influences the reliability of the test, researchers have been trying to find a solution to this problem. For example, the use of subliminal stimuli (Maoz et al., 2012) and the Rapid Serial Visual Presentation (RSVP) technique (Bowman et al., 2014). RSVP, used frequently in research on memory and attention (Ariga & Yokosawa, 2008; Broadbent & Broadbent, 1987; Nieuwenstein & Potter, 2006; Potter, 1976), is a paradigm in which a series of visual information (e.g., pictures, words) are presented sequentially and rapidly at the same spatial location. By presenting these images in such a short time that only a small subset of them can be consciously perceived, it is thought that there are no ways to countermeasure the resulting physiological response (Bowman et al., 2014). The brain selectively attends to salient information, while disregarding irrelevant information. Once the subconscious brain detects these salient stimuli, they break through into awareness and cannot be hindered by volitional cognitive control and thus countermeasures cannot be applied (Bowman et al., 2014).

In the study done by Bowman et al. (2013), they showed that concealed identity can be detected using EEG and RSVP. Participants were asked to take on a fake name as their new identity. Different names - including their real name (probe), fake name (target), and irrelevant names (control) - were presented on-screen, while EEG was used to measure a P300 ERP, which is evoked by recognized items. They concluded from their experiment that the P300 potentials were provoked significantly earlier and with significantly larger amplitudes, when participants were presented with their real name, than when presented with their fake name (Bowman et al., 2013).

Another way to detect variation of arousal in response to stimuli could be to determine alteration in Locus Coeruleus (LC) activity. The LC is a large nucleus, which is involved in the neural pathway controlling autonomic functions and arousal. Activation of the LC activates sympathetic and inhibits parasympathetic output. One way to determine increased activity of the LC, and thus measuring variation of arousal, is an increase in pupil diameter (Murphy et al., 2014; Samuels & Szabadi, 2008).

Making use of the same RSVP and concealed identity test as Bowman et al. (2013), Chen et al. (2022) used pupillary response to detect deception and found that this paradigm can be used to detect concealed identity, thus, replicating the results of Bowman et al. (2013). However, the CIT and RSVP in combination with pupillometry, resulted only in significant detection of concealed identity on a group level but not on an individual level.

Although Chen et al. (2022) did not find an effect on individual level, further exploration of this study is of importance to find a cost-efficient way of implementing the CIT in forensic settings, for which pupillometry is a suitable technique since it is easily accessible and applicable with low cost of equipment and high time efficiency.

In this study we will replicate the research done by Chen et al. (2022) by combining pupillometry, RSVP and CIT to determine if concealed information can be detected. However, in contrast to Chen et al. (2022), who made use of autobiographical information (own name versus fake name), this study will try to expand on the type of information that is concealed, non-autobiographical, by imbedding a mock crime scenario. The use of a mock crime scenario is supported by Carmel et al. (2004) and Peth et al. (2012) who suggest that, although not representing real life, it can be effective if items directly related to the mock crime task are used, in this case a location name, in comparison to peripheral items (e.g. the weather at the day of the task). By generalizing the test not only to names, but also to locations, we aim to extend the usability in forensics, since a diversity of crimes can be committed and testing a variety of scenarios could prove its effectiveness in different types of criminal activities. Furthermore, as suggested by Chen et al. (2022) we will decrease the amount of trials based on their observation that fatigue might play a role in decreased pupil-size over time and that habituation to control stimuli might influence the results.

Using pupillometry we will measure involuntary responses to rapidly presented stimuli. The stimuli in this study will be names of locations where the participant could "hide" an object, in this case a hidden suitcase that is part of the mock crime scenario. Using this set-up, we will compare pupil characteristics between the real hiding location (probe), fake hiding location (target) and control locations (irrelevant-1 and irrelevant-2).

Because of the consequences the CIT outcome can have on individuals in the forensic setting, it is of great importance to build a reliable test that can determine on an individual level concealed information, and therefore it needs to have a high sensitivity and specificity. We aim to achieve this by extending on the statistical analysis that we replicate from Chen et al. (2022), by following the suggestion of Mathôt & Vilotijević (2022) to use linear mixed effect analysis, add a slope analysis and make use of classification.

We hypothesize a) that mean pupil size is significantly larger in trials where the real hiding location (probe) is shown compared to trials where irrelevant-2 (control) is shown during the time window of 600-900 ms after the critical item (T1), b) that slopes of pupil size are significantly steeper in trials where the probe is shown compared to trials where irrelevant-2 is shown during the time window of 600-900 ms after T1, and c) combining mean pupil size and pupil-slope measures will be a better predictor to classify participants as "guilty" or "innocent" than either of the factors alone. Additionally, we expect that there will be no differences in mean pupil size and slope between control conditions (irrelevant-1 and irrelevant-2).

Method

Participants

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

Apparatus & Stimuli

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

Design

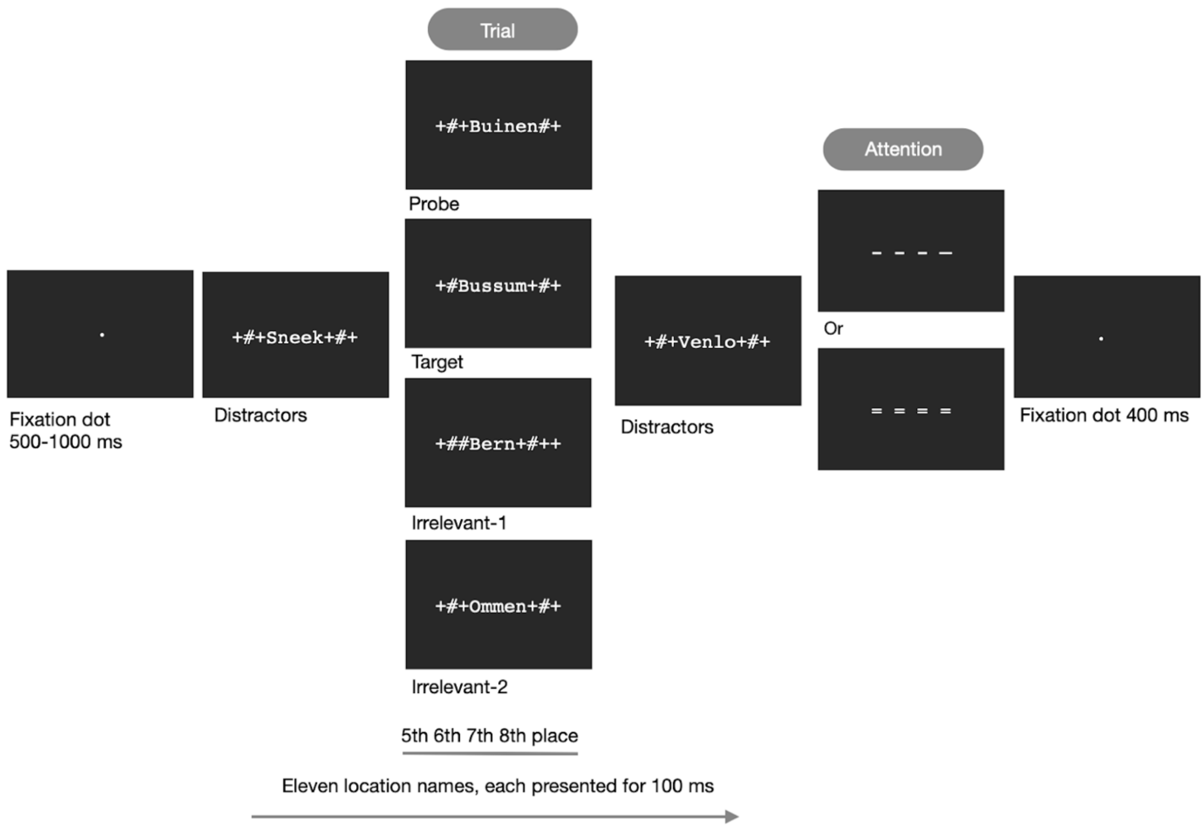
We used a within-participant experimental design. As part of a mock-crime scenario, an RSVP task was used, with a list of Dutch locations as stimuli (Klein, 2022; see OSF). From the list of 1105 Dutch locations, 15 were randomly chosen to be the potential real hiding location (probes) and removed from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

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

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

Figure 1

RSVP trial sequence



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

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

Procedure

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

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

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

Data processing

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

Analysis

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

Preprocessing

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

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

Group level analysis

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

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

Individual analysis

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

Classification

To investigate the discriminability of our testing method, we conducted a classification analysis. We used the mean pupil size and pupil derivative in the window of 600-900 ms after T1 onset as predictors. The participant was marked as guilty if the *t*-test between probe and irrelevant-2 exceeded the critical threshold $t(48) = 1.677$, $p < .05$. Second, if the *t*-statistic was below the critical threshold, the participant was marked as innocent. We also checked the significance of the difference between irrelevant-1 and irrelevant-2 for each participant. If a participant marked as guilty showed a significant difference between irrelevant conditions, we marked this participant as undetermined.

Results

Task performance

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

Group level analysis

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

Figure 2a

Traces of Mean Pupil Size

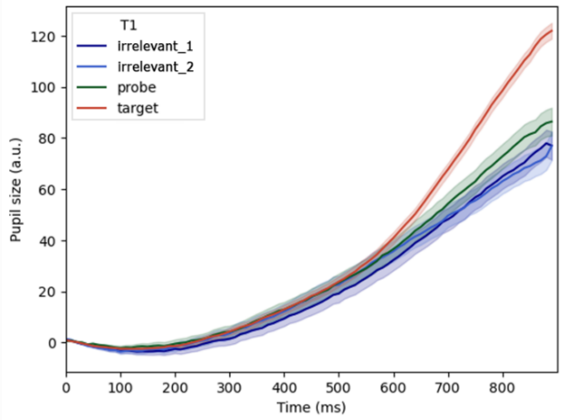
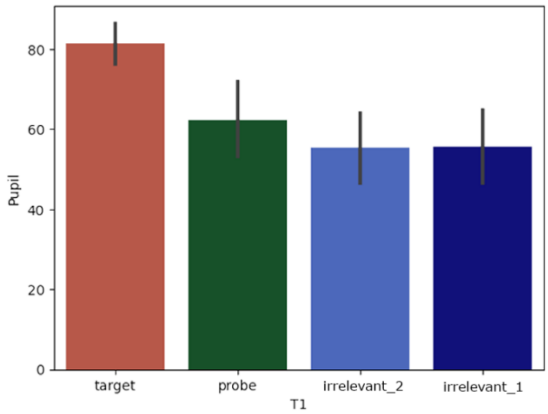


Figure 2b

Mean Pupil Size for Each Condition



Note. Time window from 600-900 ms.

Figure 3a

Traces of Mean Pupil Derivative

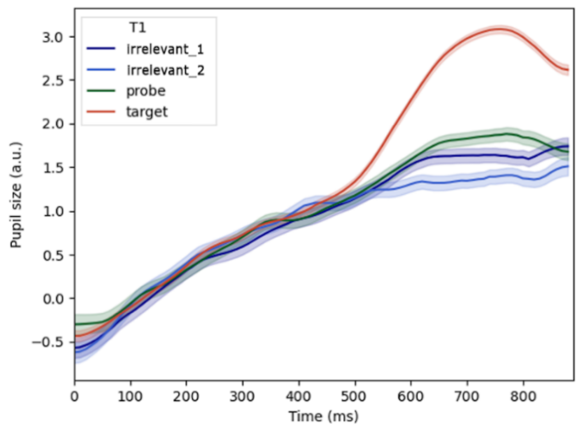
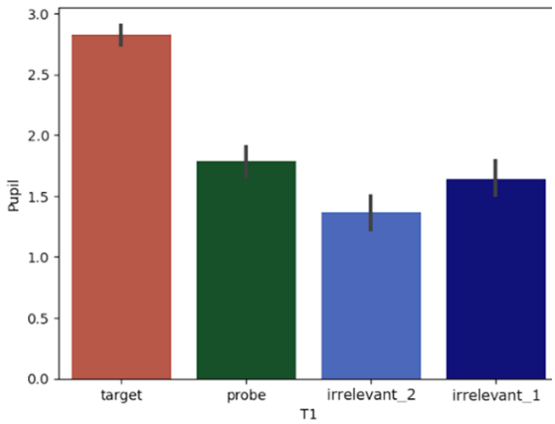


Figure 3b

Mean Pupil Derivative for Each Condition



Note. Time window from 600-900 ms.

Mean pupil size

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

Table 1

Linear Mixed Effects Analysis for Mean Pupil Size

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

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

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	$p < .001$

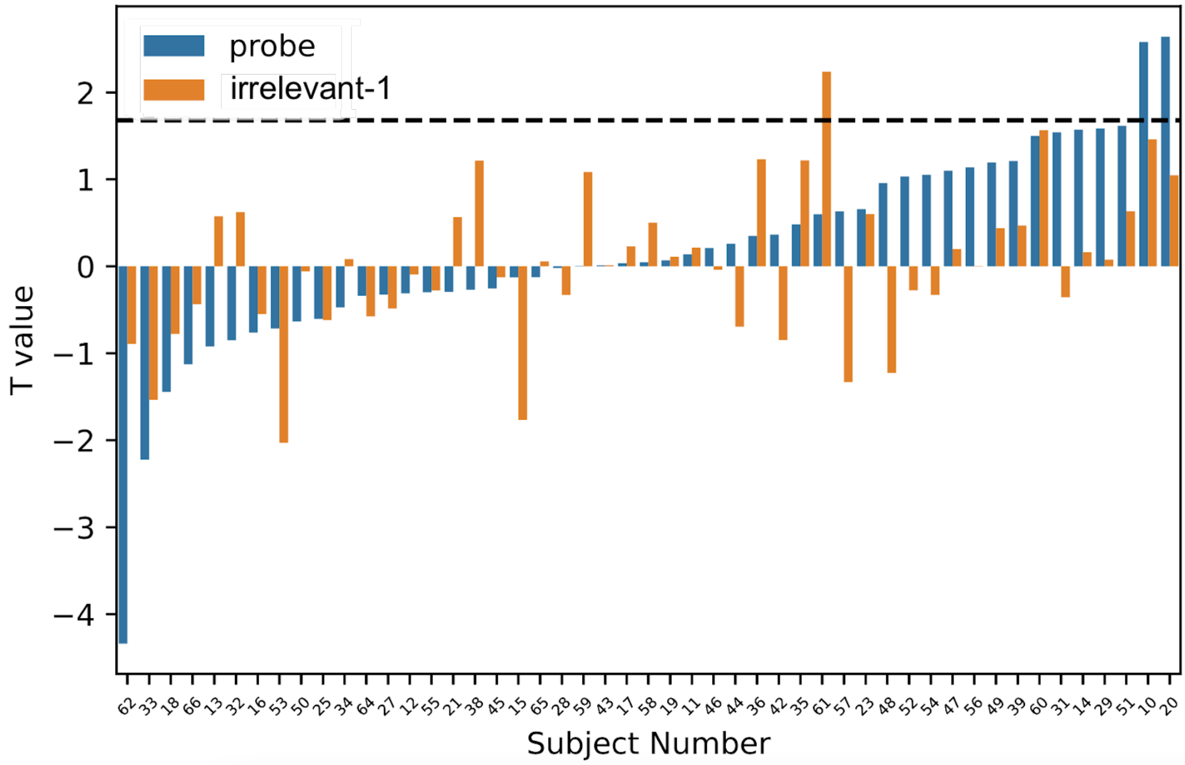
Note. $N = 6876$ (trials). $\alpha < .05$.

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 condition. Nine participants showed a significant difference between the irrelevant-1 and irrelevant-2 conditions for pupil derivative. Four out of these nine participants also had a significant difference between the probe and irrelevant-2 condition (Figure 4b). Appendix B summarizes all individual results.

Figure 4a

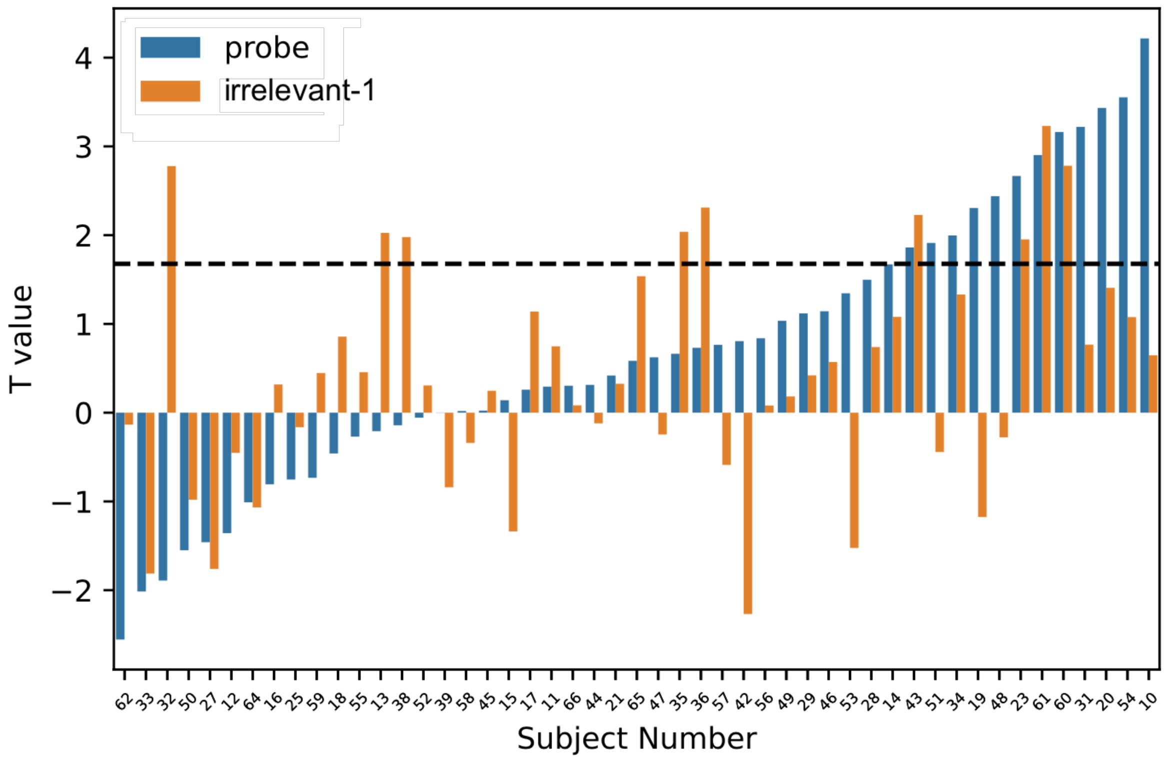
Individual Mean Pupil Size Differences (Probe and Irrelevant-1 vs. Irrelevant-2)



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

Figure 4b

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



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

Classification

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

Discussion

In this study we used pupillometry combined with Rapid Serial Visual Presentation (RSVP) and the Concealed Information Test (CIT) (Lykken, 1959) to determine if detection of concealed information in forensic settings can be achieved using this paradigm.

Our primary hypothesis was that mean pupil size is significantly larger in trials where the real hiding location (probe) is shown compared to trials where irrelevant-2 (control) was shown during the time window of 600-900 ms after the critical item (T1). The mean pupil size was not significantly different for the probe in comparison to the irrelevant-2 at group level. However, on an individual level, two participants out of 49 did show a significant difference. The results of this study indicate that mean pupil size could not reveal concealed information in an RSVP task at group level and only predict concealed information for some participants at individual level.

Second, we hypothesized that the slopes of pupil size (derivative) were significantly steeper in trials where the probe was shown compared to trials where irrelevant-2 was shown during the time window of 600-900 ms after T1. Our results showed a significant difference in pupil derivative when presented with the probe at group analysis. On an individual basis, only twelve out of 49 participants showed a significant difference.

Our third hypothesis was that mean pupil size and pupil derivative combined would be a better predictor to classify participants as “guilty” or “innocent” than either of these factors alone. However, due to insignificance of the mean pupil size, we restricted our analysis to mean pupil size and pupil derivative as separate classifiers. Mean pupil size had a low accuracy with a hit rate of 4%. Second, classification accuracy based on pupil derivative showed a hit rate of 16% and four participants were marked as undetermined.

In contrast to Chen et al. (2022) we were not able to detect the same significant difference for differentiating between probe and fake hiding locations (target) using mean pupil size difference. However, our experiment did support previous findings that activity of the Locus Coeruleus (LC) is involved in the pupillary response to salient stimuli, represented by the significant difference in pupil derivative seen when participants were presented with the target compared to irrelevant-2 (Murphy et al., 2014; Samuels & Szabadi, 2008). This leaves the question why in contrast to Chen et al. (2022), we have not found a significant difference for mean pupil size as a predictor. A possible explanation might lie in the salience of our probe, which unlike in Chen et al. (2022) experiment, was not autobiographical and thus left the participant unaroused. A second possibility might be that the time window of 600-900 ms is not sufficient for full pupil dilation, as Chen et al. (2022) saw their significant difference between a time window of 600-1200 ms. A third possibility is that participants did not recall the probe, after they were asked to focus on the target. Although we did make use of an assignment to make the participant familiar with the probe and thus try to make it salient, we did not determine whether the participant actually remembered the probe.

One of the unexpected findings of this work is that the starting letter of probe, target and irrelevant-2 may have influenced the pupil dilation of the participant. This is supported by further data which shows a significant pupil derivative difference between the irrelevant-1 and irrelevant-2 conditions at group level analysis. On an individual level analysis, one participant showed significant differences for the mean pupil size, and nine participants showed significance for the pupil derivative analysis. This is in contradiction with our expectation that no differences in mean pupil size and derivative between irrelevant-1 and irrelevant-2 would be seen. Meaning that the participants possibly made use of a search mechanism (Harris et al., 2021), which in this case was to search for the starting letter of the target, as participants were instructed to look for this location name, however the starting letter was equal for probe and the irrelevant-1, but not irrelevant-2 condition. Which means that a possible explanation for pupil dilation is, the thought of encountering the target as a consequence of a search strategy that participants might employ. It is thus recommended for future studies to make use of different starting letters such that these search mechanisms don't interfere with the task.

Upon conducting individual pupil derivative analysis, it was found that the ten participants who did not exhibit significant difference on the target trials also did not demonstrate significant difference on the probe trials. A plausible explanation could be that these participants simply failed to perceive either location. However, participants were excluded based on their accuracy scores in identifying the target during the task. In such cases, it may be more appropriate to consider the possibility that the LC was not activated by the salient stimuli and thus the resulting lack of change in pupil size as a potential explanation.

It would be interesting to look into causes of potential differences in pupil reaction between individuals, of which some may be task related. For example, it might be that the dyslexic brain is not able to detect words at a high speed or needs more time to process the information which in turn delays reaction time to salient stimuli (Jasińska & Landi, 2019). Further research is needed to establish validity of this research paradigm in different subpopulations.

Another suggestion is to take the use or abuse of psychotropic medication or recreational drugs into account. Our data do support activation of the LC. Thus, certain types of drugs that alter state of arousal (sedatives and stimulants), may influence LC activity directly or indirectly via inputs from the sleep/arousal network (Samuels & Szabadi, 2008; Szabadi, 2018), and additionally can influence the pupillary muscle (Kardon, 2011). Resulting in no or unreliable change in pupil size. Seen as half of the detainees at the Dutch Custodial Institutions Agency have an addiction to substances (Nationale Drug Monitor, 2023), this would be a crucially important factor to consider, if this method is to be further developed and researched for use in forensics.

Lastly, the use of our mock-crime might not have been sufficient to enhance salience, since salience of items can be higher if emotional load is ascribed to it. This could be due to the fact that the fear of negative consequences, in contrast to the real life forensic setting, is missing in this experimental setting and that the mock-crime was not distressing, leaving the participant unaroused (Peth et al., 2012).

Our findings highlight the necessity for a reliable and well-established test to accurately detect concealed information in forensic settings. One promising approach is the use of the RSVP paradigm, as demonstrated by Bowman et al. (2013), which has shown potential to counteract countermeasures taken by participants. However, to better comprehend the limitations of our results, further investigation into the use of pupillometry in combination with the CIT may provide a more cost and time-effective solution.

In conclusion, our results show that mean pupil size in our time window did not prove to be a reliable predictor of concealed information, however pupil derivative shows to be a more promising measure for detection of concealed information a pupillometry, RSVP and CIT paradigm. Future studies should take factors as previously suggested into account that could influence the responses to determine individual differences.

Reference

Ariga, A., & Yokosawa, K. (2008). Attentional awakening: Gradual modulation of temporal attention in rapid serial visual presentation. *Psychological Research*, *72*(2), 192–202. https://doi.org/10.1007/s00426-006-0100-4

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

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

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

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2004). Estimating the Validity of the Guilty Knowledge Test From Simulated Experiments: The External Validity of Mock Crime Studies. *Journal of Experimental Psychology. Applied*, *9*, 261–269. https://doi.org/10.1037/1076-898X.9.4.261

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*, *60*(1). 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

Davies, G., & Beech, A. R. (2018). Detecting Deception. In *Forensic psychology: Crime, justice, law interventions* (3rd ed.). The British Psychological Society.

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory Detection* (1st ed., pp. 27–45). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.003

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

Jasińska, K., & Landi, N. (2019). Dyslexia and Its Neurobiological Basis. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 625–646). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190672027.013.25

Kardon, R. (2011). Regulation of Light through the Pupil. In *Adler's Physiology of the Eye* (pp. 502–525). Elsevier. https://doi.org/10.1016/B978-0-323-05714-1.00025-X

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

Krapohl, D. J., McCloughan, J. B., & Senter, S. M. (2009). *How to Use the Concealed Information Test*.

Langleben, D. D., Hakun, J. G., Seelig, D., Wang, A.-L., Ruparel, K., Bilker, W. B., & Gur, R. C. (2016). Polygraphy and Functional Magnetic Resonance Imaging in Lie Detection: A Controlled Blind Comparison Using the Concealed Information Test. *The Journal of Clinical Psychiatry*, *77*(10), 1372–1380. https://doi.org/10.4088/JCP.15m09785

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

Maoz, K., Breska, A., & Ben-Shakhar, G. (2012). Orienting response elicitation by personally significant information under subliminal stimulus presentation: Demonstration using the Concealed Information Test: OR elicitation without awareness in a CIT paradigm. *Psychophysiology*, *49*(12), 1610–1617. https://doi.org/10.1111/j.1469-8986.2012.01470.x

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

Matsuda, I., Nittono, H., & Allen, J. J. B. (2012). The Current and Future Status of the Concealed Information Test for Field Use. *Frontiers in Psychology*, *3*, 532. https://doi.org/10.3389/fpsyg.2012.00532

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

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Nationale Drug Monitor, editie 2023. Criminaliteit en overlast 17.0 Laatste feiten en trends - Nationale Drug Monitor. https://www.nationaledrugmonitor.nl/criminaliteit-en-overlast-laatste-feiten-en-trends/. Geraadpleegd op: 11 april 2023. Trimbos-instituut, Utrecht & WODC, Den Haag.

Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal Limits of Selection and Memory Encoding: A Comparison of Whole versus Partial Report in Rapid Serial Visual Presentation. *Psychological Science*, *17*(6), 471–475.

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

Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*(3), 381–390. https://doi.org/10.1111/j.1469-8986.2011.01313.x

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

Sai, L., Zhou, X., Ding, X. P., Fu, G., & Sang, B. (2014). Detecting Concealed Information Using Functional Near-Infrared Spectroscopy. *Brain Topography*, *27*(5), 652–662. https://doi.org/10.1007/s10548-014-0352-z

Samuels, E. R., & Szabadi, E. (2008). Functional Neuroanatomy of the Noradrenergic Locus Coeruleus: Its Roles in the Regulation of Arousal and Autonomic Function Part II: Physiological and Pharmacological Manipulations and Pathological Alterations of Locus Coeruleus Activity in Humans. *Current Neuropharmacology*, *6*(3), 254–285. https://doi.org/10.2174/157015908785777193

Szabadi, E. (2018). Functional Organization of the Sympathetic Pathways Controlling the Pupil: Light-Inhibited and Light-Stimulated Pathways. *Frontiers in Neurology*, *9*, 1069. https://doi.org/10.3389/fneur.2018.01069

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 Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. (2013). *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*
23	2.67**	1.95*	4.44**	0.66	0.60	1.48

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

Subject_nr	Probe derivative	Control derivative	Target derivative	Probe mean pupil size	Control mean pupil size	Target mean pupil size
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
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. *$p < .05$. **$p < .01$.