Detecting Concealed Familiarity Using Eye Movements

The Effect of Leakage of Mock Crime Details to Innocents

Van der Cruyssen, I.; Ben-Shakhar, G.; Pertzov, Y.; Verschuere, B.

DOI
10.1037/mac0000140

Publication date
2024

Document Version
Final published version

Published in
Journal of Applied Research in Memory and Cognition

AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

2024, Vol. 13, No. 4, 516–525
https://doi.org/10.1037/mac0000140

EMPIRICAL ARTICLE

Detecting Concealed Familiarity Using Eye Movements: The Effect of Leakage of Mock Crime Details to Innocents

Ine Van der Cruyssen[1, 2], Gershon Ben-Shakhar[1], Yoni Pertzov[1], and Bruno Verschuere[2]
[1] Department of Psychology, Hebrew University of Jerusalem, Israel
[2] Department of Clinical Psychology, University of Amsterdam, The Netherlands

The present study examined the eye-tracking Concealed Information Test (CIT) in a mock crime scenario. Participants were instructed to either commit a mock crime on campus (guilty participants; $n = 42$), read an article about this mock crime (informed innocents; $n = 45$), or read an unrelated article (naïve innocent participants; $n = 46$). Afterward, all participants were presented with an eye-tracking CIT task. Based on preregistered analyses of participants' gaze behavior, we were able to distinguish the guilty participants from the naïve innocents (area under the curve [AUC] = .71, 95% CI [.60, .82]). Interestingly, we were also able to distinguish the guilty participants from the informed innocent ones (AUC = .65, 95% CI [.53, .77]). Although these results are promising, the observed detection efficiency was lower than both previous eye-tracking CIT studies that used highly familiar stimuli as well as mock crime CIT studies relying on physiological measures.

General Audience Summary

To examine whether a suspect of a crime has implicating crime knowledge, recognition of crime-related details can be detected using memory detection methods, such as the CIT. Traditionally, crime knowledge is detected with physiological measures, and studies exploring newer eye-tracking measures have produced a lower detection efficiency. Recently, an eye-tracking paradigm has been introduced that showed a high detection efficiency for highly familiar stimuli. The present study is the first to assess this promising eye-tracking CIT approach in a mock crime scenario. Participants were instructed to either commit a mock crime on campus (guilty participants; $n = 42$) or read an unrelated article (naïve innocent participants; $n = 46$). As leakage of information provides a great challenge to memory detection, we also added a group of informed innocents ($n = 45$) to whom critical crime knowledge was leaked through an article about the mock crime. All participants were presented with an eye-tracking memory detection task. Based on preregistered analyses of participants' gaze behavior, we were able to distinguish the guilty participants from the naïve innocents (AUC = .71, 95% CI [.60, .82]). Interestingly, we were also able to distinguish the guilty from the informed innocent ones (AUC = .65, 95% CI [.53, .77]). Although these results are promising, the observed detection efficiency was lower than both previous eye-tracking CIT studies that used highly familiar stimuli, as well as mock crime CIT studies relying on physiological measures.

Keywords: memory detection, eye movements, Concealed Information Test, leakage

This article was published Online First October 26, 2023.
Ine Van der Cruyssen https://orcid.org/0000-0002-9296-4055
Yoni Pertzov https://orcid.org/0000-0002-3395-0155
Bruno Verschuere https://orcid.org/0000-0002-6161-4415
This research was supported by the Barbara and Morton Mendel Foundation (awarded to Ine Van der Cruyssen). The funding source had no other role other than financial support.

The authors would like to thank Stijn Nuiten and Marcus Spaan for giving an introduction to eye tracking. They would also like to thank Luca Weiss, Michel Bonnet, Nesrine Sioua, Aimee Diepgrond, and Anika Schönberger for their help during the data collection.

Yoni Pertzov and Gershon Ben-Shakhar have a patent on concealed information testing using gaze dynamics (Patent No. US-11020034-B2, June 1, 2021).

This study was preregistered (https://osf.io/sa6yb). Materials, data, and analytic scripts have been made publicly available and can be accessed at https://osf.io/p562n/.

Ine Van der Cruyssen played a lead role in data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, and writing–original draft and an equal role in conceptualization and writing–review and editing. Gershon Ben-Shakhar played an equal role in conceptualization, supervision, and writing–review and editing. Yoni Pertzov played an equal role in conceptualization, supervision, and writing–review and editing. Bruno Verschuere played an equal role in conceptualization, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Ine Van der Cruyssen, Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 VZ Amsterdam, The Netherlands. Email: i.a.vandercruyssen@uva.nl

Ever since its introduction by David Lykken (see Lykken, 1959), the Concealed Information Test (CIT; originally called the Guilty Knowledge Test) has proven to be a reliable method for determining whether a suspect of a crime possesses crime-related knowledge (see Verschuere et al., 2011, for a review). The CIT relies upon the assumption that only the actual perpetrator possesses specific knowledge about the crime. By comparing the physiological responses of individuals to crime-related items only known by the perpetrator (e.g., the murder weapon) with equally plausible alternatives (e.g., gun, knife, rope), concealed knowledge can be detected. After decades of research, it is well established that the CIT provides a valid, theory-driven tool for the detection of "knowledgeable" individuals (Meijer et al., 2014) and is routinely used in criminal investigations in Japan (Osugi, 2011).

Traditionally, the possession of concealed knowledge is detected based on autonomic nervous system (ANS) activity (e.g., electrodermal, respiratory, and heart rate responses; Gamer, 2011). However, capturing physiological measures could be challenging in certain ecological settings, such as airports, border control, or police interviews, since they require a laboratory setup and direct contact with the examinee. Recently, the relative ease and rapidity of video-based eye tracking have raised the possibility of using this tool for the detection of concealed knowledge (Gamer & Pertzov, 2018; Lancry-Dayan et al., 2023).

Several studies have already shown that eye movements are not only influenced by features of the stimuli but also by the observer's memory, which appears to have a substantial impact on visual exploratory behavior (see Hannula et al., 2010, for a review). For example, it has been shown that familiar faces are scanned with a smaller number of fixations, fewer regions sampled, fewer return fixations, smaller proportions of fixations to the inner regions of the face (e.g., Althoff & Cohen, 1999), and longer fixation durations (Schwedes & Wentura, 2012) as compared to unfamiliar faces. This memory-guided gaze dynamic extends to recently learned faces (Heisz & Shore, 2008) as well as complex scenes (Ryan et al., 2000). Moreover, the effects of memory on eye movements can occur rapidly, obligatorily, and even in the absence of conscious awareness (Hannula et al., 2010). This link between memory and eye movements is of immediate relevance for the detection of concealed information.

Indeed, initial CIT studies based on eye tracking revealed that familiar stimuli elicit fewer and longer fixations that are directed less toward the inner regions of the stimulus as compared to unfamiliar stimuli (Millen et al., 2017, 2020; Millen & Hancock, 2019; Peth et al., 2013, 2016). However, these studies typically resulted in a lower detection efficiency (measured as the area under the receiver operating characteristic curve [AUC]) than physiological measures. For example, Peth et al. (2016) directly compared CIT detection efficiency based on physiological to efficiency based on various eye movement measures and reported an AUC of 0.88 for physiological measures and AUCs ranging between 0.59 and 0.83 for eye movement measures. Only the number of fixations showed a relatively good detection efficiency with an AUC of 0.83. This lower detection efficiency for eye movements may be accounted for by the sequential presentation of the items. While a serial presentation only allows the participant to look at the currently presented stimulus, a parallel presentation of both familiar and unfamiliar objects allows participants to choose when and for how long to look at each stimulus. This could possibly unfold various patterns of visual attention associated with recognition. In addition, the type of task used is important, as it can impact and guide gaze behavior. For example, T. Nahari et al. (2019) showed that the deployment of gaze toward familiar items is substantially different when performing a short-term memory task than when performing a visual detection task.

Lancry-Dayan et al. (2018) introduced a novel approach to memory detection based on eye tracking, which included both parallel and single stimulus displays. Specifically, a display of four stimuli was presented, followed by a presentation of a single stimulus. Participants had to perform a short-term memory task and decide whether the single item appeared in the display of four items presented a few seconds earlier. The addition of this task is particularly important as familiar items require fewer resources for encoding into short-term memory (Jackson & Raymond, 2008), and therefore participants are expected to direct their gaze more toward the unfamiliar stimuli. This way, the task further contributes to the gaze-related measures that distinguish familiar from unfamiliar stimuli. Indeed, Lancry-Dayan et al. (2018) found that when one of the four faces was personally familiar, participants' gaze was initially directed toward it, followed by a strong preference toward the other, unfamiliar faces. This pattern was also found when participants were explicitly asked to conceal recognition of familiar faces or when they were instructed to look equally at all faces. Using the gaze deployment pattern, a signal detection analysis revealed high detection efficiency estimates (ranging from 0.89 to 0.97), which are even larger than estimates derived from physiological measures (see Meijer et al., 2014), suggesting the potential applicability of detecting "knowledgeable" individuals based on their eye movements.

However, Lancry-Dayan et al. (2018) used only highly familiar stimuli (faces of close friends), which can be used in restricted cases (e.g., identifying members of a terror group), but are not very common in realistic criminal investigations. Moreover, previous studies have found that the degree of familiarity influences behavioral and neuronal responses (Buttle & Raymond, 2003; Leveroni et al., 2000). Specifically, Millen et al. (2017, 2020) demonstrated that newly learned stimuli—such as crime-related items—can limit the scope of memory detection based on eye tracking. This is also the case for the paradigm developed by Lancry-Dayan et al. (2018). A recent eye-tracking CIT study that used this paradigm found lower detection efficiency estimates for newly learned compared to highly familiar items (Lancry-Dayan et al., 2021). Thus, the primary goal of the present study is to assess the detection efficiency of mock crime items based on eye movements in the paradigm proposed by Lancry-Dayan et al. (2018).

Furthermore, the notion that only guilty suspects have knowledge about a crime is, unfortunately, not always valid. Through newspapers, social media, or even during interrogation itself, innocent suspects may be exposed to critical details of the crime and could therefore show enhanced responses to the crime-related items. Several CIT studies based on ANS measures have already shown that such leakage can indeed result in false positive outcomes (Bradley et al., 2011). Various attempts have been made to overcome this problem (such as presenting the items on an exemplar level rather than a categorical level, which is less susceptible to leakage; Geven et al., 2019, 2022), but these attempts were only partially effective and so far no ideal solution was offered to the leakage problem. It is yet unclear how information leakage may affect the eye-tracking measures. Thus, the second goal of the present study is to examine whether and to what extent eye movement measures are affected by information leakage.

In the present study, we apply the novel eye-tracking paradigm (Lancry-Dayan et al., 2018) in a more ecologically valid "mock crime" situation. To that end, participants were randomly assigned to either the guilty condition (commit a mock crime), the informed innocent condition (read an article about the mock crime), or the naïve innocent condition (read an unrelated article). Next, participants were informed that they were suspected of committing a crime on campus. Our primary research question was whether guilty participants can be distinguished from naïve innocents based on their eye movements when performing a short-term memory task similar to the one used in Lancry-Dayan et al. (2018). Our secondary research question was whether informed innocents can be differentiated from guilty and naïve innocent participants.

Method

After running a preliminary study (preregistration: https://osf.io/eh9dv), which yielded no significant results, we modified the design and method, which resulted in the present study. There were two main problems with this preliminary study. First, we had some empty cells in our mixed design, preventing us from running all our planned analyses of variance (ANOVAs; i.e., we had a two [condition: guilty vs. innocent, between subjects] by two [item: crime-related vs. foil, within subjects] by two [leakage: leaked vs. nonleaked, within subjects] mixed design, but we did not leak foils, resulting in some empty cells). Second, our study had insufficient statistical power (there were only 14 guilty participants and 21 innocent participants and a substantial exclusion rate of 27%). The key changes applied in the present study were the between-subjects manipulation of leakage (instead of within subjects) and a strong increase in power. To be transparent, the data and materials of this preliminary study have been made publicly available and can be accessed at https://osf.io/se6nd/.

The present study was preregistered (https://osf.io/sa6yb), and we made no deviations to the preregistration. All materials, data, and analytic scripts of the study can be accessed at https://osf.io/p562n/. The study was approved by the ethical committee of the Social and Behavioral Sciences faculty at the University of Amsterdam and registered as no. 2019-CP-11391. The protocol was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants

We determined our sample size based on a power analysis. Specifically, we aimed for 90% power to detect an AUC of at least .70. AUC is a standard measure used in CIT studies to assess how well guilty participants can be discriminated from naïve innocent participants. Meta-analyses of CIT studies revealed that the average AUCs range from .74 for heart rate to .94 for functional magnetic resonance imaging (Meijer et al., 2014, 2016). To provide a more conservative estimate that is practically meaningful, we aimed to be able to detect an AUC of at least .70. Using MedCalc as statistical software, this resulted in a required sample size of 41 participants per group, resulting in a total sample size of minimum $N = 123$. We continued testing until we had 41 inclusions in each group, which resulted in a total sample size of 155 participants.

Participants with normal or corrected-to-normal vision were recruited through the online recruitment system of the University of Amsterdam and received 1 credit or 10 euros for participation. Out of these 155 participants, following our preregistered criteria, we excluded three participants because they had less than 70% valid trials (i.e., no pupil detected, due to closed eyes or large head movements), eight participants because they had calibration/validation problems, seven participants because they underperformed on the short-term memory task (i.e., had less than 60% correct trials), two participants of the guilty condition who did not perform the mock crime properly (did not return with the two stolen items), and two participants of the guilty condition who did not remember at least five out of eight crime-related items in the poststudy memory test (see below).

The final sample size contained 133 participants, of which 42 were assigned to the guilty condition, 46 to the naïve innocent condition, and 45 to the informed innocent condition. In total, 77% of the sample was female, 22% was male, and 1% defined themselves as other. They had a mean age of 21 years ($SD = 4$ years) and were from the Netherlands (27%), Germany (11%), or one of 43 other countries.

Procedure

At the beginning of the experiment, participants gave written informed consent and were asked to shut down their phones.

Mock Crime

Participants were randomly assigned to either the guilty condition, the informed innocent condition, or the naïve innocent condition. Participants received instructions from the experimenter accordingly. Guilty participants were instructed to commit a mock crime on campus in which they had to steal a *wallet* and a *phone* from another *researcher*'s lab. The phone was hidden in a jacket, and the wallet was in a *safe* covered by a *scarf*. The key to the safe was hidden in a *purse*. The participants received *gloves* and a *backpack* from the experimenter to carry their loot (the eight crime-related items are italicized in the previous sentences; the exact instructions can be found at https://osf.io/5grs7; face stimuli were selected from the Nimstim database of facial expressions; Tottenham et al., 2009). Informed innocents were instructed to leave the lab and read an article about the crime in which pictures describing the theft were included. Naïve innocents were instructed to leave the lab to read an unrelated article.

Upon return, all participants were informed by the experimenter that there had been a crime on campus of which they were suspected. Then, participants did a short-term memory task similar to the one used in Lancry-Dayan et al. (2018). They were told that they should try to appear innocent and were promised an extra reward if they would be classified as innocent by the computer.

Short-Term Memory Task

The task was displayed on a 23 in. Syncmaster monitor, with a 120 Hz refresh rate, and a 1,024 × 768 screen resolution. Monocular gaze position was tracked at 1,000 Hz with an Eyelink 1,000+ (SR Research Ltd., Mississauga, Ontario, Canada). Participants' head was stabilized using a chinrest, situated 60 cm from the screen.

The experiment started with the standard 9-point calibration and validation procedure provided with the eye tracker. Afterward, all participants did five practice trials, with a set of stimuli not used during the actual test, to get familiar with the task. Participants had to complete at least three correct practice trials to be able to continue with the real experiment. Participants who failed in more than two out of the five trials did another session of five practice trials. Participants repeated this procedure until they reached the minimal threshold of three correct practice trials.

Each trial of the main task started with a fixation validation process. During this process, participants had to fixate on a fixation point in the middle of the screen and press the space bar. When the visual angle between the predicted gaze position and the center of the fixation point was less than one degree, participants were able to continue. Larger deviations were accompanied by an error beep and the opportunity to repeat the calibration process. After the fixation validation process, participants saw a parallel display of four stimuli (5,000 ms), followed by a fixation point (3,000 ms), a single stimulus display (2,000 ms), and a blank screen with a central fixation point (5,000 ms; see Figure 1). Critically, all stimuli were either related to the mock crime or foils. The fixation point before the single display was displayed below the stimulus to avoid a biased gaze position to any specific location on the stimulus (Arizpe et al., 2012; Peterson & Eckstein, 2013). During the single stimulus display, participants were required to press "p" if the single stimulus also appeared in the previous display with four stimuli, and "q" if it did not.

The short-term memory task consisted of 64 trials in total. Half of the trials contained one crime-related item and three foils in the parallel display. The other half contained only foils. For each participant, there were eight unique crime-related items (see above) and 56 unique foils. Each mock crime item was presented four times, once in each location (top left, top right, bottom left, and bottom right). In the single display, 25% of the displays contained a crime-related item, and the other 75% contained a foil. Half of the trials in the single display contained an item that also appeared in the parallel display and required the answer "p." The other half contained an item that did not appear in the parallel display and required the answer "q." Thus, a crime-related item could be displayed only in the parallel display, only in the single display, in both displays, or none.

After the short-term memory task, participants were informed that the deception detection study was over and were asked to honestly enter their demographics. Next, participants were asked to take a memory test. They were asked to answer eight multiple-choice recognition questions about the items that were stolen during the mock crime (i.e., there was one question about each stolen item). Each question had eight answer options (chance level: 12.5%). If the participants were innocent or did not know the answer, they had to guess. Finally, participants had to rate the significance level of all crime-related items and eight foils on a 9-point Likert scale (where 1 was *not significant at all* and 9 was *very significant*). This information was used to examine whether the effect was driven primarily by recognition and/or by item significance. In the end, guilty participants were asked to return the two stolen items.

Results

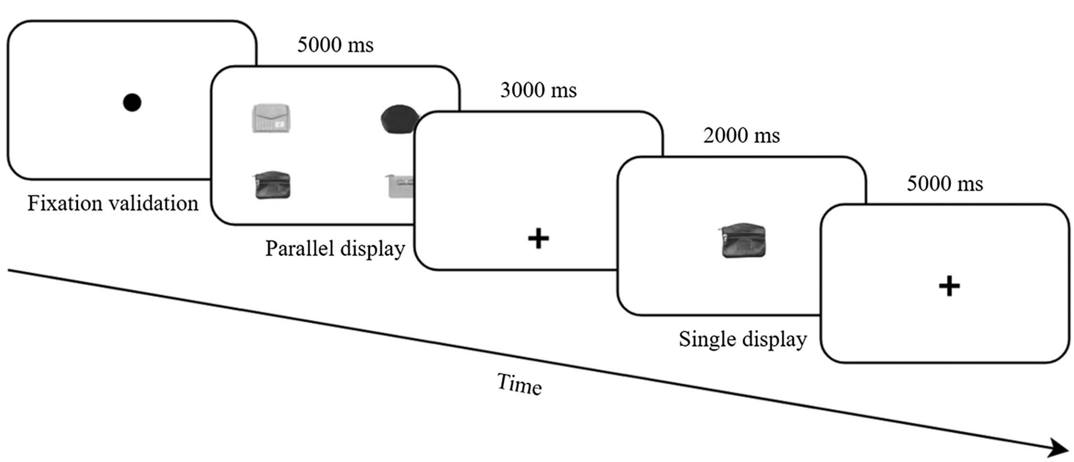
Eye movement data were parsed into saccades and fixations using Eyelink's standard parser configuration (i.e., samples of which the deviation of consecutive samples exceeded 30°/s velocity or 8,000°/s^2 acceleration were defined as saccades, intervals in between saccades were defined as fixations).

Confirmatory Primary Analysis

AUC for Guilty Versus Naïve Innocent Participants

The main research question of the present study was whether guilty participants can be distinguished from naïve innocent participants based on their gaze behavior. To examine the detection efficiency of the eye-tracking measures in differentiating between guilty and naïve innocent participants, a Receiver Operating Characteristic curve was constructed. For this purpose, a combined index of all six relevant indices from Lancry-Dayan et al. (2018; i.e., number of visits, fixation count, dwell time during the first phase [1–1,000 ms], dwell time during the second phase [1,000–5,000 ms] of the parallel display, mean fixation duration, and reaction time during the single display) was computed for crime items, for each participant. To this end, we

Figure 1
Schematic Overview of the Course of a Trial of the Short-Term Memory Task



first standardized all indices within each participant across all trials using a Z transformation. The direction of the Z score for each measure was defined according to the results reported by Lancry-Dayan et al. (2018; e.g., the number of visits was expected to be lower for guilty participants, so these z scores were multiplied by −1 such that higher indices reflected more recognition). Then we computed the detection score as the average of these 6 Z scores of the crime items, using equal weights. Finally, the detection-score distribution of the guilty participants was compared to the detection-score distribution of naïve innocent participants, and the AUC was computed.

This analysis showed that we were able to distinguish guilty participants from naïve innocent ones significantly above chance with an AUC of .71, 95% CI [.60, .82], p = .001. For the AUCs for the individual measures, see our nonpreregistered analyses. To define the optimal cutoff point, the Youden index was calculated. The maximum Youden index was obtained for cutoff point of 0.036, Jmax = 0.440. This cutoff point corresponded to a sensitivity of 0.810 (i.e., 81% hits, 19% misses) and a specificity of 0.630 (i.e., 37% false positives).

Confirmatory Secondary Analyses

AUC for Other Conditions

The secondary research question of the present study was whether guilty participants can be distinguished from informed innocents and whether informed innocent participants can be distinguished from naïve innocents. Similar to our primary analysis, we computed the AUCs for guilty versus informed innocents and informed innocents versus naïve innocents. These analyses showed that informed innocents could be significantly distinguished from guilty participants (AUC = .65, 95% CI [.53, .77], p = .028). The maximum Youden index was obtained for cutoff point of 0.045, Jmax = 0.251. This cutoff point corresponded to a sensitivity of 0.762 (i.e., 76% hits, 24% misses) and a specificity of 0.489 (i.e., 51% false positives). Informed innocents could not be significantly distinguished from naïve innocents (AUC = .61, 95% CI [.49, .73], p = .082). The maximum Youden index was obtained for cutoff point of 0.025, Jmax = 0.342. This cutoff point corresponded to a sensitivity of 0.733 (i.e., 73% hits, 27% misses) and a specificity of 0.609 (i.e., 39% false positives).

Analysis of Variance

Last, we ran two orthogonal contrasts on the combined index: (a) Comparing all knowledgeable participants (i.e., guilty + informed innocent participants) to unknowledgeable participants (i.e., naïve innocent participants), where the effect of interest was the interaction between stimulus type (crime-related item vs. foil) and group, and (b) comparing guilty to informed innocents, where once again the factor of interest was the interaction.

In the first ANOVA, the interaction was significant, $F(1, 131) = 8.75$, $p = .004$, $\eta^2 = 0.04$, indicating that the difference between crime-related items and foils was larger for knowledgeable participants than for unknowledgeable participants. In the second ANOVA, the interaction was also significant, $F(1, 85) = 7.80$, $p = .006$, $\eta^2 = 0.04$, indicating that the difference between crime-related items and foils was larger for guilty participants compared to informed innocent ones.

Nonpreregistered Analyses

Individual Measurements

To test the contribution of the individual measures to the combined index, we ran the analyses again with each separate measure, comparing the guilty participants to the naïve innocents. The results of these analyses can be found in Table 1. Results comparing guilty to informed innocents and comparing informed innocents to naïve innocents can be found on the OSF at https://osf.io/56qv7.

Although the differences between these measures are small, Table 1 suggests that the variables of the parallel display seemed to contribute more than the variables of the single display. As compared to naïve innocent participants, guilty participants visited the crime-related items less, dwelled more on them during the first phase of the parallel display, and less during the second phase of the parallel display. This pattern of an initial preference followed by an avoidance of crime-related items is similar to the pattern observed by Lancry-Dayan et al. (2018). This particular pattern in guilty and informed innocent participants is visualized in Figure 2.

AUC Including Other Exploratory Measures

Previous research has proposed several other measures to signal recognition of familiar items. These proposed measures include first fixation duration (Millen & Hancock, 2019), second fixation duration (Schwedes & Wentura, 2012), and visit duration (T. Nahari et al., 2019). We tested whether these measures could increase the AUC value by adding them to the combined index. Including these measures, the AUC increased to .77, 95% CI [.67, .87], $p < .001$.

Memory Test

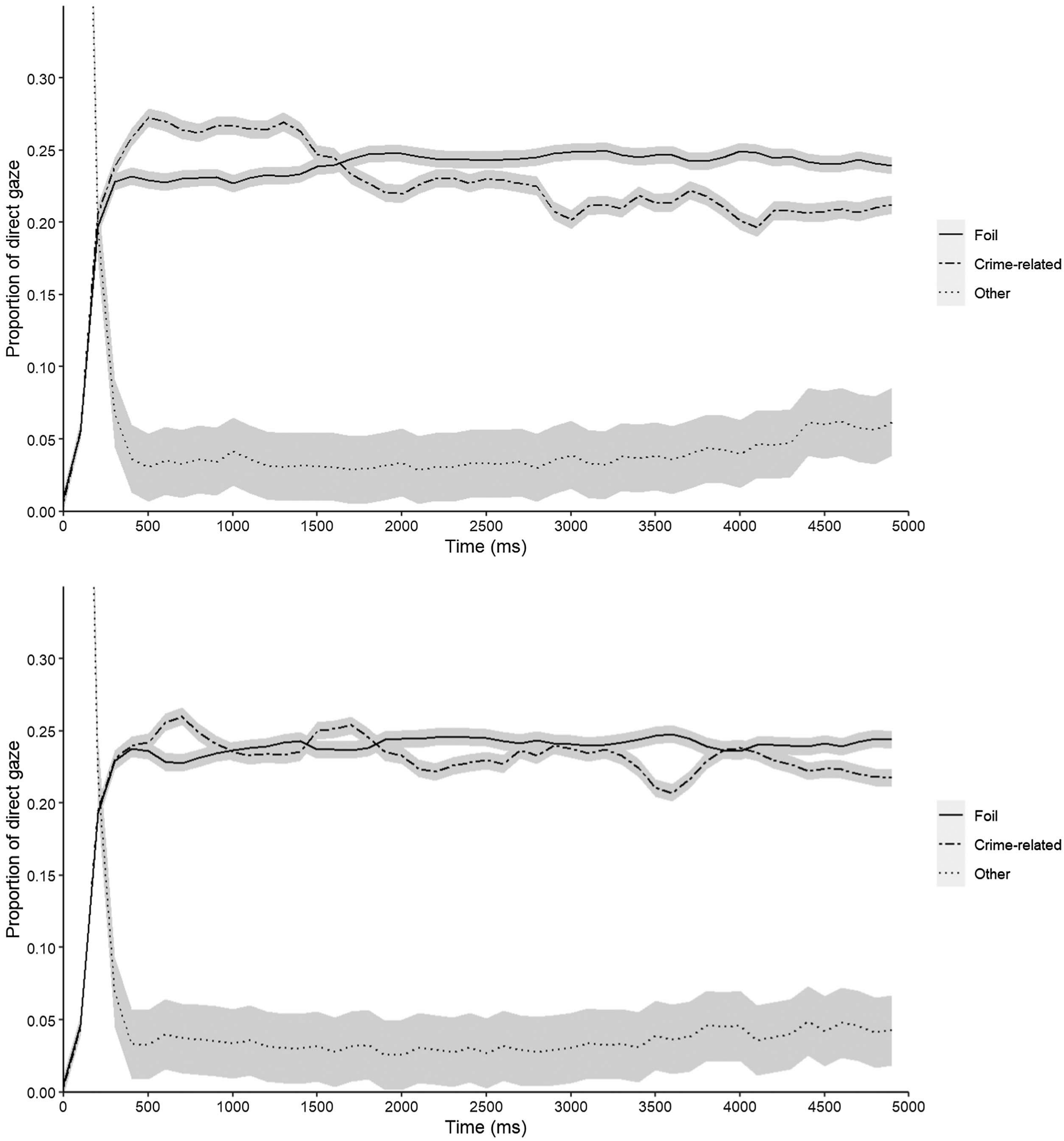
At the end of the experiment, participants were asked to recall (or guess) the crime-related items in a multiple-choice test (chance level: 12.5%). The results for each crime-related item for the guilty, informed innocent, and naïve innocent participants can be found in Figure 3. A one-way ANOVA showed that there was a significant difference between the three conditions, $F(2, 130) = 189.25$, $p < .001$, $\eta^2 = 0.74$. Post hoc comparisons with Tukey's correction revealed that guilty participants had a higher accuracy (81%) than informed innocent participants (46%), $t(85) = 11.34$, $p < .001$,

Table 1
AUC, 95% CI, and p Values of the Different Measurement That Were Part of the Combined Index

Display	Measure	AUC [95% CI]	p
Parallel display	Dwell time first phase	.67 [.56, .79]	.005
	Dwell time second phase	.63 [.51, .75]	.041
	Number of visits	.70 [.59, .82]	.001
	Fixation count	.60 [.48, .72]	.114
Single display	Mean duration of fixations	.60 [.48, .72]	.102
	Reaction time	.57 [.44, .69]	.301

Note. AUC = area under the curve; CI = confidence interval.

Figure 2
Time Course of Gaze Position of Guilty Participants (Top) and Informed Innocent Participants (Bottom) Toward Crime-Related Items, Foils, or Other Areas

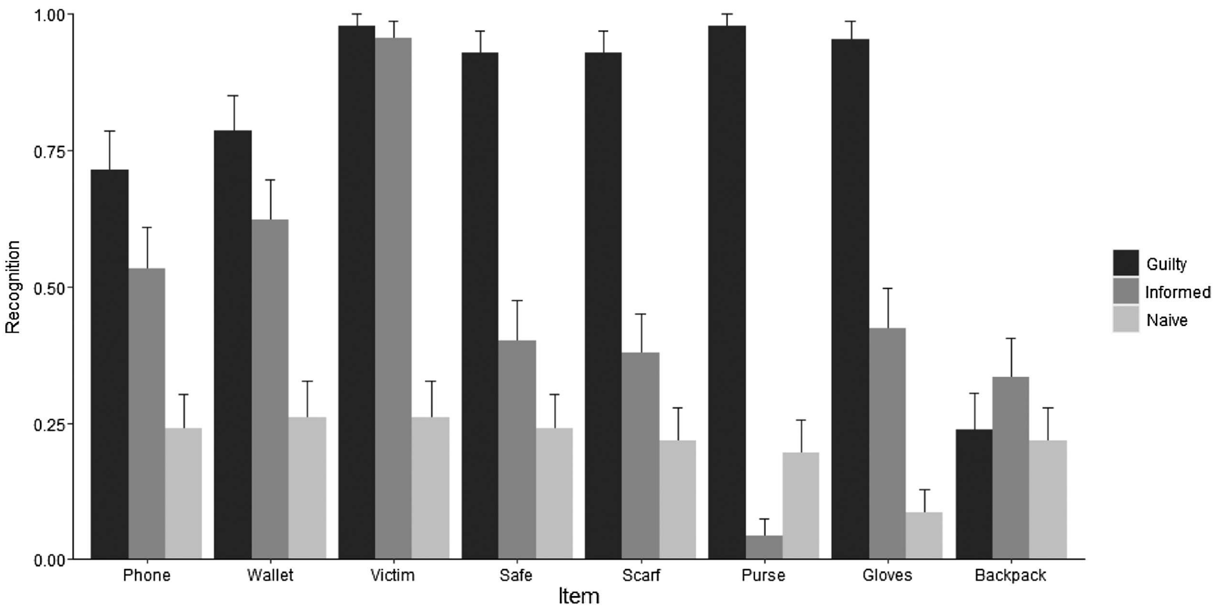


$d = 2.43$ and than naïve innocent participants (21%), $t(86) = 19.40$, $p < .001$, $d = 4.14$. Informed innocent participants also scored significantly higher than naïve innocent participants, $t(89) = 8.14$, $p < .001$, $d = 1.71$. Across conditions, there was a significant positive correlation between participants' memory test results and their combined index, $r = 0.26$, $p = .002$. More accurate results on the memory test were related to higher combined indexes. When we recalculated the AUC comparing guilty participants to naïve innocent ones using only the three items that had the largest memory differences (victim, purse, and gloves), the AUC increased to .77, 95% CI [.67, .87], $p > .001$.

Significance Ratings

At the end of the experiment, participants also had to indicate how significant they found each crime-related item on a 9-point Likert scale (where 1 was *not significant at all* and 9 was *very significant*). The results for each crime-related item in each

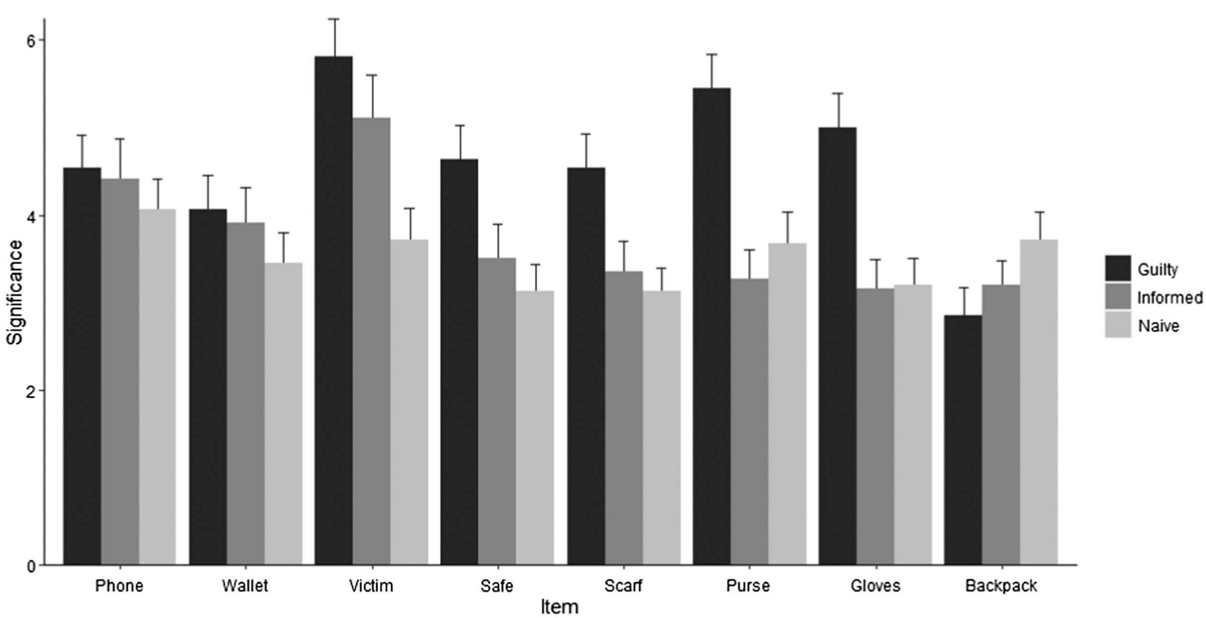
Figure 3
Memory Test Accuracies for Each Condition



Note. The error bars represent standard errors.

condition can be found in Figure 4. A one-way ANOVA showed that there was a significant difference between the three groups, $F(2, 130) = 4.55$, $p = .012$, $\eta^2 = 0.07$. Post hoc comparisons with Tukey's correction showed that significance ratings were higher for guilty participants (4.62) compared to naïve innocents (3.51), $t(86) = 2.88$, $p = .013$, $d = 0.61$, but not in comparison to informed innocents (3.74), $t(85) = 2.27$, $p = .064$, $d = 0.49$. Informed innocent participants did not rate the items significantly higher than naïve innocent participants, $t(89) = 0.61$, $p = .814$, $d = 0.13$. Across conditions, the correlation between participants' significance ratings and their combined index was not significant, $r = 0.12$, $p = .177$.

Figure 4
Significance Ratings for Each Crime-Related Item for Each Condition



Note. The error bars represent standard errors.

We ran a linear mixed model analysis with the combined index as the dependent variable and recognition, significance, and their interaction as fixed effects. Condition (guilty vs. informed innocent vs. naïve innocent) was included as a random effect. The main effect of recognition was not significant ($\beta = -0.03$, $t = -0.46$, $p = .649$), the main effect of significance was not significant ($\beta = -0.01$, $t = -1.28$, $p = .204$), and the recognition by significance interaction was also not significant ($\beta = 0.03$, $t = 1.85$, $p = .066$).

Discussion

Our findings demonstrate that CIT based on eye tracking can distinguish guilty from naïve innocent participants. This suggests that eye tracking could be a valuable tool in scenarios where physiological measures may not be feasible. Second, our study revealed that eye tracking could differentiate guilty from innocent participants who have crime knowledge because they read about the crime. This finding highlights the potential of eye tracking to protect individuals who possess knowledge about a crime but were not actively involved in it.

In our study, guilty participants dwelled more on crime-related items during the first phase of the parallel display. This is consistent with the orienting response theory, according to which attention is attracted by significant stimuli (Sokolov, 1990). This behavior is essential for adapting to a constantly changing environment, as it enables us to quickly make decisions and navigate through the world (Hannula et al., 2010). Participants dwelled less on the crime-related items during the second phase of the parallel display and also visited the items less. This finding could be explained by inhibition theory. According to this theory, the CIT effect is (partly) driven by attempts to inhibit recognition-induced arousal (klein Selle et al., 2016). When suspects recognize a crime-related item, they try to suppress the response in order to appear innocent. A recent study by klein Selle et al. (2023) showed that fixations and blinks seem to reflect arousal inhibition in the CIT. However, the shorter dwell time and fewer visits could also be explained by the nature of the task. For a successful performance of the task, participants needed to focus more on the unfamiliar items of the parallel display, because familiar items require fewer resources for encoding into short-term memory (Jackson & Raymond, 2008). This finding is in line with previous studies using a similar short-term memory task (Lancry-Dayan et al., 2018, 2021).

Mock Crime

Although we found that it is possible to distinguish between guilty and innocent in a crimelike scenario, the detection efficiency observed in the present study (AUC = .71) was lower compared to the results reported by Lancry-Dayan et al. (2018; AUC = .89). One explanation for this difference could be that the original study used highly familiar faces, while our study used newly learned stimuli. Indeed, previous research showed that highly familiar items are detected more efficiently than newly learned items (Lancry-Dayan et al., 2021; Millen et al., 2020).

The lower detection efficiency for newly learned items could be explained by a poorer memory for those items (i.e., not all guilty participants may have seen and remembered all relevant information). The recall test at the end of the study indeed showed suboptimal recall (81%). The fact that poor recall can lower the accuracy of the CIT highlights the importance of using only salient crime-related items (e.g., Gamer et al., 2010; G. Nahari & Ben-Shakhar, 2011). This is particularly important when the CIT is administered a long time after the crime.

The detection efficiency estimates in the present study are also lower than what is usually observed with physiological measures typically used to detect concealed information. The mean detection efficiency of ANS measures is 0.81 (Lancry-Dayan et al., 2023), which is substantially better than the AUC found in the present study. To compare the different measures in a more controlled way, direct comparative studies should examine eye movements alongside physiological measures within the same experimental design; this comparison would provide insights into the unique contributions and limitations of each measure (klein Selle et al., 2016, 2023). Moreover, such a study would reveal whether a combination of eye movement and physiological measures can further improve detection efficiency.

Nevertheless, eye tracking does come with some important advantages as compared to physiological measures. One advantage is the nonintrusive nature of eye tracking, as it does not require direct contact with the examinee. Unlike physiological measures (which involve attaching electrodes or sensors to the body), eye tracking can be conducted using remote, noninvasive devices. Another advantage of eye tracking is the portability and affordability of the equipment, enabling researchers to conduct studies even outside the lab. Importantly, eye-tracking measures might be better resistant to countermeasures (Lancry-Dayan et al., 2023), which is a major limitation of CIT based on ANS measures.

Leakage

Leakage is one of the biggest challenges for the application of the CIT (Lykken, 1998). The CIT is based on the premise that people with crime knowledge respond to crime-related information, regardless of how this information was acquired. Although leakage is a major concern, several studies have demonstrated that in some cases it is possible to distinguish guilty participants from informed innocents (e.g., Geven et al., 2019, 2022), implying that knowledge is not the only factor that determines the CIT effect. Factors such as motivation to avoid detection and the significance of the CIT items also affect the responses elicited by the critical items (e.g., Ben-Shakhar & Elaad, 2003). Informed innocents may be less engaged in processing or encoding the leaked items, resulting in poorer recognition and lower detection scores (McDermott & Roediger, 2018; Seymour & Fraynt, 2009). This is particularly evident when the test is delayed (e.g., G. Nahari & Ben-Shakhar, 2011). In the present study, better memory contributed to higher detection scores. Surprisingly, significance did not correlate with detection scores, suggesting that the CIT effect is driven primarily by memory of the critical items. However, a linear mixed models analysis did not confirm this idea.

The present study used a form of leakage that seems unlikely in real life, where leakage often happens only verbally (e.g., by hearing or reading about the crime). However, in the present study, leakage was introduced both verbally and pictorially. This procedure enhances the effect of leakage because leakage of verbal items is only a partial description of the item (e.g., if a participant was told that a phone has been stolen, it would be impossible for him to show an effect if four different phones are presented in the CIT).

Although the strict form of leakage examined in the present study might be less likely in real life, it does show the effects in its most drastic form. This could mean that in real life, it might be easier to distinguish between guilty and informed innocent suspects. Moreover, research has shown that a pictorial test could lead to better detection as it is more likely to result in a better match between encoding and testing (Van der Cruyssen et al., 2021). In sum, the present study might underestimate the potential of the eye-tracking CIT because pictorial leakage is rare in real life and because a pictorial test might be more accurate than the verbal test, which is still primarily used in practice today (Osugi, 2011).

However, the current method is not fully protected against leakage. The present study showed that leakage did reduce detection efficiency (AUCs decreased from .71 to .65) and increased the number of false positives from 37% to 51%.

Limitations

In the computation of the combined index, we assigned equal weights to the six components to avoid the risk of capitalization by chance. Equal weights are recommended when no independent validation of optimal weights based on a large sample is available (Dawes, 1979; Wainer, 1976). Clearly, once a proper estimation of optimal weights is available, detection efficiency based on the eye movement indices can be increased.

Conclusion

In conclusion, the present study found that it is possible to distinguish guilty from naïve innocent participants based on their eye movements in a mock crime scenario. We found that it is also possible to distinguish guilty participants from informed innocents. Although the observed detection efficiency in the present study is lower than what was found in previous eye-tracking CIT studies and CIT based on ANS measures, the advantages of eye-tracking imply that this approach may still be relevant for forensic applications. Future research focusing on eye-tracking-based CIT should examine whether and how detection efficiency with this method could be improved (Lancry-Dayan et al., 2023). Moreover, future CIT studies should use both eye movement and physiological measures, as a combination of these measures is likely to yield improved detection efficiency.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PLOS ONE*, *7*(2), Article e31106. https://doi.org/10.1371/journal.pone.0031106

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

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 187–199). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.011

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & Psychophysics*, *65*(8), 1296–1306. https://doi.org/10.3758/BF03194853

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 27–45). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.003

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biological Psychology*, *83*(2), 101–107. https://doi.org/10.1016/j.biopsycho.2009.11.005

Gamer, M., & Pertzov, Y. (2018). Detecting concealed knowledge from ocular responses. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 169–186). Academic Press. https://doi.org/10.1016/B978-0-12-812729-2.00008-2

Geven, L. M., Ben-Shakhar, G., Kindt, M., & Verschuere, B. (2019). It's a match!? Appropriate item selection in the Concealed Information Test. *Cognitive Research: Principles and Implications*, *4*(1), Article 11. https://doi.org/10.1186/s41235-019-0161-8

Geven, L. M., Verschuere, B., Kindt, M., Vaknine, S., & Ben-Shakhar, G. (2022). Countering information leakage in the Concealed Information Test: The effects of item detailedness. *Psychophysiology*, *59*(2), Article e13957. https://doi.org/10.1111/psyp.13957

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*, Article 166. https://doi.org/10.3389/fnhum.2010.00166

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8*(1), Article 9. https://doi.org/10.1167/8.1.9

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 556–568. https://doi.org/10.1037/0096-1523.34.3.556

klein Selle, N., Suchotzki, K., Pertzov, Y., & Gamer, M. (2023). Orienting versus inhibition: The theory behind the ocular-based Concealed Information Test. *Psychophysiology*, *60*(3), Article e14186. https://doi.org/10.1111/psyp.14186

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

Lancry-Dayan, O. C., Ben-Shakhar, G., & Pertzov, Y. (2023). The promise of eye-tracking in the detection of concealed memories. *Trends in Cognitive Sciences*, *21*(1), 13–16. https://doi.org/10.1016/j.tics.2022.08.019

Lancry-Dayan, O. C., Nahari, T., Ben-Shakhar, G., & Pertzov, Y. (2018). Do you know him? Gaze dynamics toward familiar faces on a concealed information test. *Journal of Applied Research in Memory and Cognition*, *7*(2), 291–302. https://doi.org/10.1037/h0101821

Lancry-Dayan, O. C., Nahari, T., Ben-Shakhar, G., & Pertzov, Y. (2021). Keep an eye on your belongings: Gaze dynamics toward familiar and unfamiliar objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(11), 1888–1901. https://doi.org/10.1037/xlm0001086

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience*, *20*(2), 878–886. https://doi.org/10.1523/JNEUROSCI.20-02-00878.2000

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

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. Plenum Press.

McDermott, K. B., & Roediger, H. L. (2018). Memory (encoding, storage, retrieval). In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology* (pp. 117–153). DEF Publishers.

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

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

Millen, A. E., & Hancock, P. J. B. (2019). Eye see through you! Eye tracking unmasks concealed face recognition despite countermeasures. *Cognitive Research: Principles and Implications*, *4*(1), Article 23. https://doi.org/10.1186/s41235-019-0169-0

Millen, A. E., Hope, L., & Hillstrom, A. P. (2020). Eye spy a liar: Assessing the utility of eye fixations and confidence judgments for detecting concealed recognition of faces, scenes and objects. *Cognitive Research: Principles and Implications*, *5*(1), Article 38. https://doi.org/10.1186/s41235-020-00227-4

Millen, A. E., Hope, L., Hillstrom, A. P., & Vrij, A. (2017). Tracking the truth: The effect of face familiarity on eye fixations during deception. *Quarterly Journal of Experimental Psychology*, *70*(5), 930–943. https://doi.org/10.1080/17470218.2016.1172093

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*(6), 733–744. https://doi.org/10.1111/j.1469-8986.2010.01148.x

Nahari, T., Lancry-Dayan, O., Ben-Shakhar, G., & Pertzov, Y. (2019). Detecting concealed familiarity using eye movements: The role of task demands. *Cognitive Research: Principles and Implications*, *4*(1), Article 10. https://doi.org/10.1186/s41235-019-0162-7

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 253–275). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.015

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684

Peth, J., Kim, J. S., & Gamer, M. (2013). Fixations and eye-blinks allow for detecting concealed crime related memories. *International Journal of Psychophysiology*, *88*(1), 96–103. https://doi.org/10.1016/j.ijpsycho.2013.03.003

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

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science*, *11*(6), 454–461. https://doi.org/10.1111/1467-9280.00288

Schwedes, C., & Wentura, D. (2012). The revealing glance: Eye gaze behavior to concealed information. *Memory & Cognition*, *40*(4), 642–651. https://doi.org/10.3758/s13421-011-0173-1

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback*, *34*(3), 177–187. https://doi.org/10.1007/s10484-009-9092-3

Sokolov, E. N. (1990). The orienting response, and future directions of its development. *The Pavlovian Journal of Biological Science*, *25*(3), 142–150. https://doi.org/10.1007/BF02974268

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Van der Cruyssen, I., Regnath, F., Ben-Shakhar, G., Pertzov, Y., & Verschuere, B. (2021). Is a picture worth a thousand words? Congruency between encoding and testing improves detection of concealed memories. *Journal of Applied Research in Memory and Cognition*, *10*(4), 667–676. https://doi.org/10.1016/j.jarmac.2021.03.001

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975196

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*(2), 213–217. https://doi.org/10.1037/0033-2909.83.2.213

Received December 6, 2022
Revision received September 3, 2023
Accepted September 6, 2023 ■