Reading the criminal mind

Exploring novel methods of memory detection

Van der Cruyssen, I.A.

Publication date
2023

3

Memory detection based on eye movements

This chapter is under review at Journal of Applied Research in Memory and Cognition as:
Van der Cruyssen, I., Ben-Shakhar, G., Pertzov, Y., & Verschuere, B. (2022). *Detecting concealed familiarity using eye movements: The effect of leakage of mock crime details to innocents.*

Abstract

The current 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 (AUC = .71). Interestingly, we were also able to distinguish the guilty participants from the informed innocent ones (AUC = .65). 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.

The fact that physiological responses can be used to reveal hidden memories is of interest to crime investigations. Ever since David Lykken introduced the test in 1959 (Lykken, 1959), the CIT has proven to be a valid method for determining whether an examinee recognizes crime-relevant details (for a review, see Verschuere et al., 2011). The CIT relies upon the assumption that only the actual perpetrator has specific knowledge about a crime. By comparing the examinees' physiological responses to crime-relevant items known only to the actual perpetrator (e.g., the murder weapon), to 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 activity (i.e., electrodermal, respiratory, and heart rate responses; Gamer, 2011). However, capturing physiological measures could be challenging in certain ecological settings like 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 sparked an interest within the scientific community to use this tool for the detection of concealed knowledge (Gamer & Pertzov, 2018; Lancry-Dayan et al., 2022).

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. In a series of studies, Neil Cohen and colleagues showed that familiar faces are scanned with a smaller number of fixations and fewer regions sampled as compared to unfamiliar faces (e.g., Althoff & Cohen, 1999). This memory-guided gaze dynamic also occurs for recently learned faces (Heisz & Shore, 2008), as well as complex scenes (Ryan et al., 2000). 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 which are directed less toward the inner regions of the stimulus as compared to unfamiliar stimuli (Millen & Hancock, 2019; Millen et al., 2017, 2020; Peth et al., 2013, 2016). However, these studies typically resulted in a lower detection efficiency (measured as the area under the Receiver Operating Characteristic [ROC] 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 relatively low detection efficiency for eye movements may be accounted for by the sequential presentation of the familiar and unfamiliar items. A parallel presentation of both familiar and unfamiliar items can provide additional important information as participants can choose when and for how long they look at each stimulus, which could possibly unfold various patterns of visual attention related to recognition.

Lancry-Dayan et al. (2018) introduced a novel approach to memory detection based on eye tracking, which included both a parallel and a single stimulus display. 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 because a successful performance requires participants to focus more on the unfamiliar items and less on the familiar ones (Nahari et al., 2019). Lancry-Dayan et al. (2018) indeed found that when one of the four faces was personally familiar, participants' gaze was initially directed toward it, followed by a strong preference towards 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 has an influence on behavioral and neuronal responses (Buttle & Raymond, 2003; Leveroni et al., 2000). Specifically, it has been shown that newly learned stimuli -such as crime related items- can limit the scope of memory detection based on eye tracking (Millen et al., 2017, 2020). 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 eye movement measures in the paradigm proposed by Lancry-Dayan et al. (2018) in a mock crime context.

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 also show enhanced responses to the crime related items (Bradley et al., 2011). Several CIT studies based on autonomic nervous system 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 current 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 this crime. 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 current 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 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 non-leaked, 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 current 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 current 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 number 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 fMRI (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 3 participants because they had less than 70% valid trials (i.e., no pupil detected, due to closed eyes or large head movements), 8 participants because they had calibration/validation problems, 7 participants because they underperformed on the short term memory task (i.e., because they had less than 60% correct trials), 2 participants of the guilty condition who did not perform the mock crime properly (i.e., who did not return with the two stolen items), and 2 participants of the guilty condition who did not remember at least 5 out of 8 crime related items in the post-study 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. 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; 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). 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.

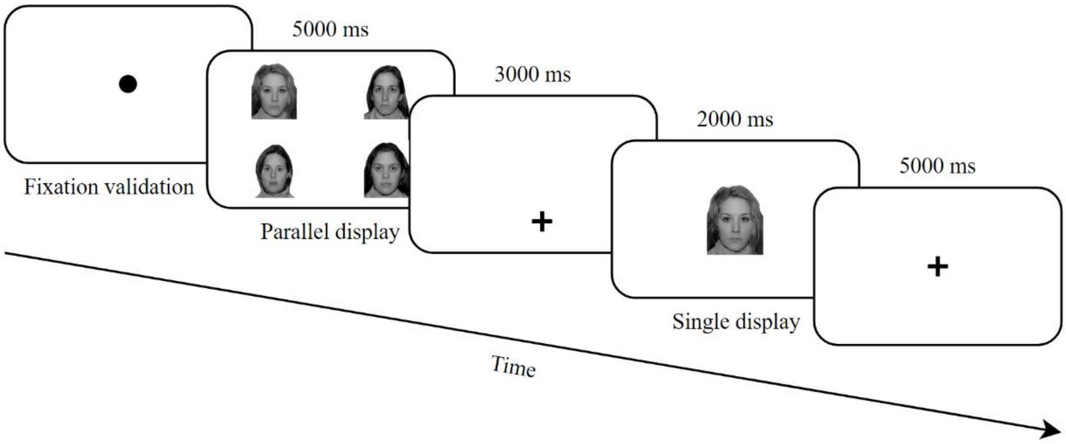
Short term memory task

The task was displayed on a 23 in. Syncmaster monitor, with a 120 Hz refresh rate, and a 1024 × 768 screen resolution. Monocular gaze position was tracked at 1000 Hz with an Eyelink 1000+ (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 nine-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 (5000 ms), followed by a fixation point (3000 ms), a single stimulus display (2000 ms), and a blank screen with a central fixation point (5000 ms; see Figure 1). 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.

Figure 1

Schematic overview of the course of a trial of the short term memory task



The short term memory task consisted of 64 trials in total. For each participant, there were 8 unique crime related items (see above) and 56 unique foils. Half of the trials contained one crime related item and 3 foils in the parallel display. The other half contained only foils. Each mock crime item was presented 4 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, report the items that were stolen during the mock crime (or guess, if they were innocent), and rate the significance level of all crime related items and 8 foils. This information was used to examine whether the effect was driven primarily by recognition and/or by item significance. Finally, 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 8000°/s² acceleration were defined as saccades, intervals in between saccades were defined as fixations).

Confirmatory primary analysis

AUC for guilty vs. naïve innocent participants

The main research question of the current 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 ROC curve was constructed. For this purpose, a combined index of all 6 relevant indices from Lancry-Dayan et al. (2018; i.e., number of visits, fixation count, dwell time during the first phase [1-1000 ms], dwell time during the second phase [1000-5000 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 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], $p = .001$.

Confirmatory secondary analyses

AUC for other conditions

The secondary research question of the current 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 vs. informed innocents and informed innocents vs. naïve innocents. These analyses showed that while informed innocents could be significantly distinguished from guilty participants (AUC = .65, 95% CI [.53; .77], $p = .028$), they could not be significantly distinguished from naïve innocents (AUC = .61, 95% CI [.49; .73], $p = .082$).

Analysis of variance (ANOVA)

Lastly, we ran two orthogonal contrasts on the combined index: 1. 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 2. 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 informed innocents, compared to naïve ones.

Exploratory 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 osf: https://osf.io/56qv7.

Table 1

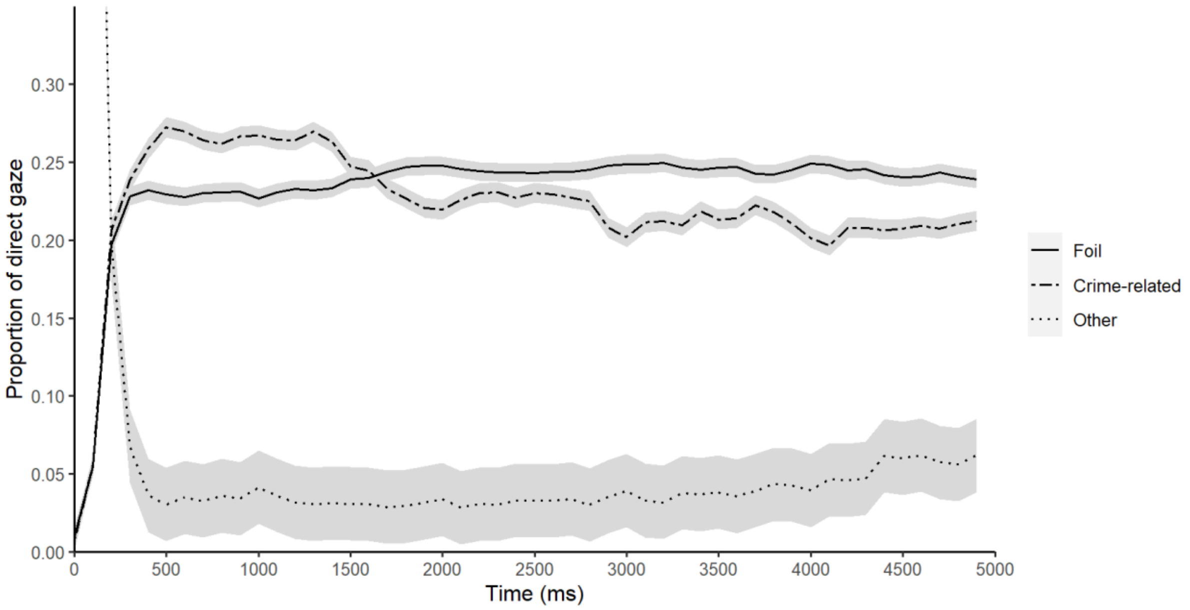
AUC, 95% CI, and p-values of the different measurement that were part of the combined index

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 fixation	.60 [.48; .72]	.102
	Reaction time	.57 [.44; .69]	.301

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 participants is visualized in Figure 2.

Figure 2

Time course of gaze position of guilty participants towards crime related items, foils, or other areas.



Memory test

At the end of the experiment, participants were asked to try to recognize (or guess) the crime-related items in a multiple choice question (chance level: 12.5%). The results for each crime related item for the guilty, informed innocent, and naïve innocent participants can be found in Table 2. 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%), $t(85) = 11.34$, $p < .001$, $d = 2.43$, 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.

Table 2

Memory test results for each condition

	Guilty	Informed innocent	Naïve innocent
Phone	71%	53%	24%
Wallet	79%	62%	26%
Victim	98%	96%	26%
Safe	93%	40%	24%
Scarf	93%	38%	22%
Purse	98%	4%	20%
Gloves	95%	42%	9%
Backpack	24%	33%	22%
Average	**81%**	**46%**	**21%**

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 condition can be found in Table 3. 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, $r = 0.12$, $p = .177$.

Table 3

Significance ratings of the crime items for each condition

	Guilty	Informed innocent	Naïve innocent
Phone	4.55	4.42	4.07
Wallet	4.07	3.91	3.46
Victim	5.81	5.11	3.72
Safe	4.64	3.51	3.13
Scarf	4.55	3.36	3.13
Purse	5.45	3.27	3.67
Gloves	5.00	3.16	3.20
Backpack	2.86	3.20	3.72
Average	**4.62**	**3.74**	**3.51**

Discussion

Can eye movements reveal crime enactment? As expected, we found that it is possible to distinguish guilty participants from naïve innocents based on their gaze behavior. Additionally, we found that it is possible to distinguish guilty participants from innocent participants who have crime knowledge because they read about the crime.

Mock crime

This was the first study that tested the short term memory CIT task in a mock crime design. Although we found that it is possible to distinguish between guilty and innocent in a crime-like scenario, the detection efficiency observed in the current study (AUC = .71) was considerably 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 as compared to highly familiar 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 recognition test at the end of the study indeed showed suboptimal recognition (81%). One crime related item was not even recognized by a majority of the guilty group and had an accuracy rate of only 24%. The fact that poor recognition can lower the accuracy of the CIT is highly relevant for forensic applications and highlights the importance of using only very salient crime-related items (e.g., Gamer et al., 2010; 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 current study are also lower than what is usually observed with physiological measures that are typically used to detect concealed information. The mean detection efficiency of autonomic nervous system activity is 0.81 (Lancry-Dayan et al., 2022). While awaiting a direct comparison, these estimates seem substantially better than the AUC found in the current study. On the other hand, eye tracking does come with some important advantages as compared to physiological measures. For example, no lab, expensive equipment, or direct contact with the examinee is needed. Detection can be achieved fast and possibly even without the suspect’s knowledge. Furthermore, the current method might be better resistant to leakage and countermeasures (Lancry-Dayan et al., 2022). Because of these advantages, future studies should examine the added value of the eye tracking CIT by testing it in combination with physiological measures.

Leakage

This is the first study to explore the effect of leakage in the short term memory CIT task. Leakage is one of the biggest challenges for the application of the CIT. The idea behind the CIT is that people with crime knowledge respond to crime related information, regardless of how this crime knowledge was acquired. This is also why David Lykken, one of the pioneers in CIT research, suggested that if there was any concern that information about the crime had been leaked, the test should be avoided (1998). 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). The observation that guilty participants show a larger CIT effect than informed innocents suggests 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 the leaked items, resulting in poorer recognition and lower detection scores. This is particularly evident when the test is delayed. For example, Nahari and Ben-Shakhar (2011) demonstrated that guilty and informed innocents show similar recognition and detection efficiency when the CIT is administered immediately after the mock crime. However, when introducing delays, both measures attenuate much more in informed innocents. In the current study, better memory contributed to higher detection scores, but significance did not correlate with detection scores. This suggests that the CIT effect is driven primarily by memory of the critical items.

The current study used a form of leakage that may be unlikely in real life. In real life, leakage often happens only verbally (e.g., by reading or hearing about the crime), while in the current study, leakage was introduced both verbally and pictorially (the article which described the crime verbally, also contained pictures of all crime details). 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 this strict form of leakage 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, and it could be resistible to verbal leakage (see Chapter 2; Van der Cruyssen et al., 2021). In sum, the current 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 a verbal test which is still primarily used in practice today (Osugi, 2011).

However, it should be noted that even though the comparison between guilty and informed innocent participants indicates that the two groups can be differentiated, this does not imply that the current method is fully protected against leakage. The current study showed that leakage did reduce detection efficiency (AUCs decreased from .71 to .65). Although a carefully chosen threshold could protect most informed innocents, false-positive outcomes could still occur.

Limitations

In the control condition of naïve innocents, participants read an unrelated article, rather than executing an irrelevant activity. However, the critical factor in the attempt to simulate innocent suspects who are unaware of the crime scene, is to insure that the control participants will not be exposed to the critical items. From this respect, reading an unrelated article or executing an unrelated activity are similarly valid procedures. In the computation of the combined index, we used a conservative approach and assigned equal weights to the 6 components. This was done to avoid the risk of capitalization on 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. From this respect, the present detection efficiency estimate should be viewed as an underestimate.

Conclusion

In conclusion, the current study found that it is possible to distinguish guilty participants 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 innocent participants who have crime knowledge because they read about the crime. Although the observed detection efficiency in the current study is lower than what was found in previous eye tracking CIT studies and than CIT based on autonomic nervous system 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., 2022). Moreover, future CIT studies should use both eye movement and physiological measures as a combination of the 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. doi: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), e31106. doi: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. doi:10.1037/0021-9010.88.1.131

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. *Memory detection: Theory and application of the Concealed Information Test,* 187 –199. doi: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. doi: 10.3758/bf03194853

Dawes, R. M. (1979). The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist, 34*(7), 571-582. doi:10.1037/0003-066x.34.7.571

Gamer, M. (2011). Detecting concealed information using autonomic measures. *Memory detection: Theory and application of the Concealed Information Test,* 27–45. doi: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. doi: 10.1016/j.biopsycho.2009.11.005

Gamer, M., & Pertzov, Y. (2018). Detecting Concealed Knowledge From Ocular Responses. *Detecting Concealed Information and Deception,* 169–186. doi: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), 1-11. doi: 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), e13957. doi: 10.1111/psyp.13957

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision, 7*(9), 24-24. doi:10.1167/8.1.9

Lancry-Dayan, O. C., Ben-Shakhar, G., & Pertzov, Y. (2022). The promise of eye-tracking in the detection of concealed memories. *Trends in Cognitive Sciences.* doi: 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. doi:10.1016/j.jarmac.2018.01.011

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.* doi: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. *Journal of Neuroscience, 20*(2), 878-886. doi: 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. doi:10.1037/h0046060

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

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. doi: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. doi:10.1111/psyp.12609

Millen, A. E., & Hancock, P. J. (2019). Eye see through you! Eye tracking unmasks concealed face recognition despite countermeasures. *Cognitive research: principles and implications, 4*(1), 1-14. doi: 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), 1-18. doi: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. doi: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. doi: 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), 1-16. doi: 10.1186/s41235-019-0162-7

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. *Memory detection: Theory and application of the Concealed Information Test,* 253. doi: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. doi: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. doi: 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. doi: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. doi:10.1111/1467-9280.00288

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. doi: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. doi:10.1017/cbo9780511975196

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