PRE-PROOF AUTHOR COPY

Caught virtually lying – crime scenes in virtual reality help to expose suspects' concealed recognition

Danielle G. Norman[1], Kimberley A. Wade[1], Mark A. Williams[2] and Derrick G. Watson[1]

[1]Department of Psychology, University of Warwick

[2]WMG, University of Warwick

Cite as: Norman, D.G., et al Caught Virtually Lying - Caught virtually lying – crime scenes in virtual reality help to expose suspects' concealed recognition. Journal of Applied Memory and Cognition (2020), https://doi.org/10.1016/j.jarmac.2019.12.008

Abstract

We explore how virtual reality could be used in police investigations to take a suspect 'back in time' and demonstrate that they recognize a crime scene despite claiming not to. In this study, participants committed a mock crime before being incentivized to conceal recognition of crime related details (e.g., the stolen item or crime scene). The crime scenes and objects were laser scanned, converted to photo-realistic models, and presented to suspects either in Virtual Reality (VR) or as 2D images on a computer screen. While concealing recognition of crime information, participants' heart rate and skin conductance were measured using a Concealed Information Test (CIT) to assess recognition. Detection of concealed recognition increased by over 25% when participants viewed crime items in VR compared to 2D images. Our findings suggest that revisiting crime scenes or objects in VR may enhance stimulus recognition and salience resulting in increased CIT diagnosticity.

Keywords: Concealed Information Test (CIT), Laser Scanning, Virtual Reality (VR), Deception, Memory Recognition, Skin Conductance

Caught virtually lying – crime scenes in virtual reality help to expose suspects' concealed recognition

Imagine if, during a police interrogation, you could take a suspect 'back in time', using Virtual Reality (VR) and demonstrate that the suspect recognizes the crime scene despite them claiming not to. Recent advancements make this possible with technologies available to digitally capture, document, and visualize scenes and objects with exceptional precision (Puente, González-Jorge, Martínez-Sánchez, & Arias, 2013). In domains such as medicine, engineering and emergency response, VR has been shown to be effective for training professionals (Häfner, Häfner, & Ovtcharova, 2013). Yet, other studies assessing memory for information acquired in VR, but tested in the real world, suggests there is either no advantage of VR as a learning tool, compared to traditional methods (Voinescu & David, 2019), or that learning suffers (Lanen & Lamers, 2018). Although research into the benefits of acquiring knowledge via VR is mixed, there are good reasons why VR may enhance memory strength, or the salience of information learned in the real world; something not yet tested.

Decades of memory research show that remembering is improved when individuals recall knowledge in the same context in which the knowledge was learned, known as *Context Reinstatement* (Bailenson et al., 2008; Godden & Baddeley, 1975; Smith & Vela, 1992; Tulving & Thompson, 1973). Relatedly, memory is enhanced when the cognitive operations carried out at encoding are reinstated and used at retrieval (Dewhurst & Brandt, 2007). Presumably, the reactivation of neural patterns established during encoding enhances one's ability to bring that information back to mind (Staudigl & Hanslmayr, 2018). Such "transfer appropriate processing" may explain the *Pictorial Superiority Effect* in which memory retrieval is typically superior when the retrieval modality is a picture, image or photograph (Hockley, 2008). Given the current realism of VR environments, we might expect higher levels of context reinstatement and stronger modality congruence when memories are retrieved in VR rather than 2D images or other modalities. That is, we might observe a 'VR superiority effect'.

A powerful memory-detection paradigm with important legal applications is the Concealed Information Test (CIT) (Granhag, Vrij & Verschuere, 2015; Verschuere, Ben-Shakhar, & Meijer, 2011; Osugi, 2011). The CIT aims to determine whether a suspect recognizes information that only the culprit would know (Verschuere, Ben-Shakhar, & Meijer, 2011). In CIT studies, participants are typically instructed to carry out a simulated crime, such as stealing a specific item from a location (Verschuere, Ben-Shakhar, & Meijer, 2011) or from within a virtual environment (Hahm et al., 2009). The CIT presents 'suspects' with various stimuli and determines their recognition of crime details by measuring their physiological responses (e.g., skin conductance, heart rate, P300) to crime-related items (termed probe items, e.g., a tablet computer stolen from a handbag) and to control items (irrelevant items, e.g., other portable electronic items or bags). A person without knowledge of the crime would be unable to discriminate the crime from control items. Whereas a person with knowledge of the crime would show a larger physiological response to crime items than to control items (Verschuere, Ben-Shakhar, & Meijer, 2011). This difference in recognition response to crime items versus control items is known as the CIT effect.

The CIT typically identifies the guilty suspect approximately 8 out of 10 times in lab-based experiments while nearly always correctly rejecting innocent suspects (Ben-Shakhar & Elaad, 2003). While a great deal of CIT research has been conducted over the past fifty years, little has examined the benefits of increasing modality congruence in the CIT. One recent study showed a strong effect of modality congruence using P300 (an event related potential brain wave indicating recognition) CIT when comparing pictures against verbal presentation (Rosenfeld, Ward, Frigo, Drapekin, & Labkovsky, 2015). Another demonstrated both a picture superiority effect in the P300 CIT (Zheng et al., 2019) and a modality effect (Deng, Rosenfeld, Ward, & Labkovsky, 2016) when suspects attempt to conceal recognition of mock crime details. These findings suggest that physiologically measured recognition should be larger in VR compared to viewing 2D images due to VR-driven increased modality congruence.

Feature matching theory conceptualizes the above memory models while offering a specific framework for understanding physiological orienting and its relationship to recognition intensity. The physiological response caused by orienting is monotonically related to the similarity between the encoded and test stimuli (Ben-Shakhar & Gati, 1987). The magnitude of the physiological CIT effect increases as the number of overlapping features between the encoded and test image being recognized, increases. This has important implications for the CIT because the CIT effect should increase as the similarity between the presented crime item and the actual crime memory increases (Marchand, Inglis-Assaff, & Lefebvre, 2013). Hence, feature matching theory predicts that, compared to 2D image stimuli, VR presentation of real-world digital recreations will increase the physiological CIT effect by increasing responses to crime items. This is due to the increased number of features available, that is, scale and 3D depth, for the suspect to correctly match the crime item to a memory of the real-world. Thus, both memory recognition research and CIT theory suggest that VR-based retrieval should increase the measurable recognition of crime details encoded in the real-world in comparison to 2D images.

We examined whether VR enhances physiologically measured recognition strength and thus improves the diagnosticity of the CIT. Participants committed a mock crime before being incentivized to conceal recognition of details relating to that crime. The crime scenes and objects were laser scanned, converted to photo-realistic models, and revisited by suspects either in VR or as 2D images on a computer screen (Figure 1). The VR condition provided a rich and realistic environment that was tracked to the participant's head position. Neither the VR environment nor the 2D images contained independent object motion. We then assessed participants' recognition by measuring heart rate and skin conductance in a CIT. We expected that participants taking the VR CIT would show a greater skin conductance response and heart rate deceleration to crime items, and thereby show a stronger CIT effect, than those taking the 2D image CIT.

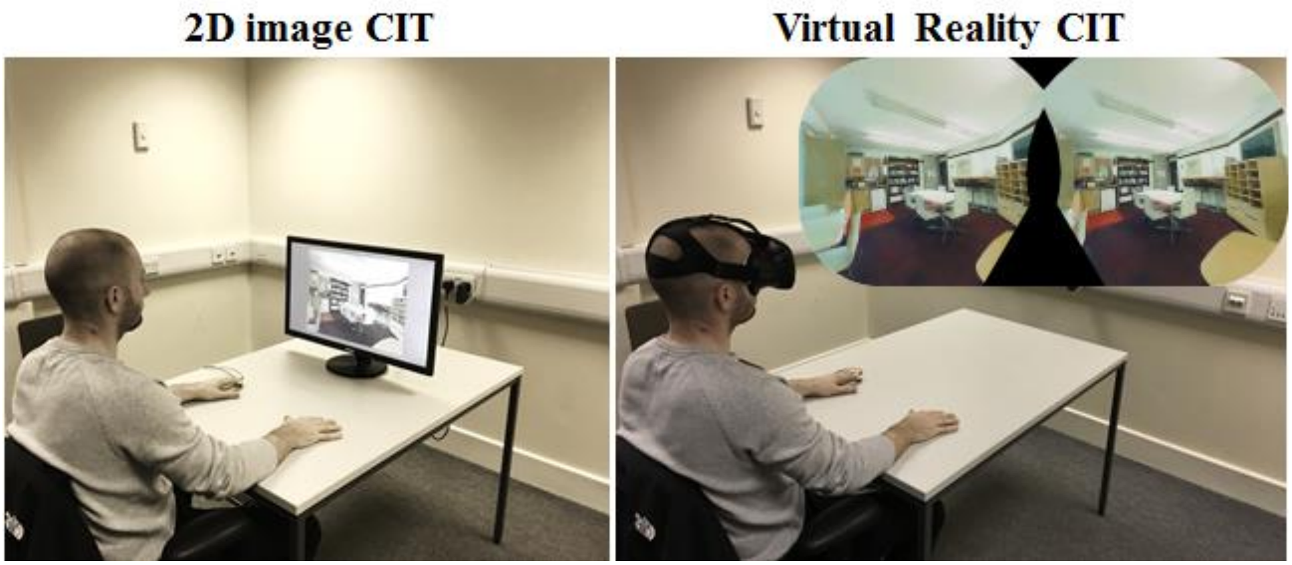

Figure 1. A participant viewing the crime scene; as 2D images on a monitor (Left), or within a 3D, 1:1 scale, head-tracked virtual reality environment (Right).

Method

Participants

According to a meta-analytic review, the CIT effect between innocent and guilty is typically large (d = 1.55), indicating that 8 people per group is sufficient for finding a main CIT effect (Ben-Shakhar & Elaad, 2003). Estimating the effect size of testing modality and feature matching was more problematic given no previous studies had compared physiological recognition for VR against 2D photographs. One study revealed a large effect for modality congruence between picture and verbal stimuli using the P300 CIT (Rosenfeld et al., 2015). In providing evidence for their feature matching theory, Ben-Shakhar and Gati (1987) found large modality effects with groups of 30 participants. Based on these findings it was estimated that there would be a large modality effect size (Cohen, 1988). A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), assuming a large effect size of $d = 0.8$, and $\alpha = 0.05$ for a single group, suggested a minimum sample size of 23 participants would be sufficient for a power of 0.95. Because up to 25% of participants could be skin conductance non-responders (Venables & Mitchell, 1996) the sample was increased to 32 per each of the four groups for a total sample size of 128.

One-hundred and twenty-eight adults (59% women, 18 - 46 years, *mean* = 21, *SD* = 4.1) were recruited via a university online participant panel at the authors' institution. Participants received £4 payment for participating in the 30-minute testing session and the opportunity to receive their 'lie detection score'. Participants were incentivized with the chance to win a £25 Amazon voucher if they obtained the lowest score. Participants were equally split and randomly allocated to one of four experimental conditions: Guilty with a VR-CIT, innocent with a VR-CIT, guilty with a 2D image CIT and innocent with a 2D image CIT.

Materials

The CIT. Participants in the guilty condition carried out a mock crime in which they entered an office, identified an unattended handbag and stole a tablet computer before handing it over to an ‘accomplice’ in the department common room. The four key crime details used in the CIT were the office, handbag, tablet and common room. The CIT therefore consisted of four questions/blocks each containing one crime item and three control items (Figure 2). CIT blocks were presented in a random order and each began with a question presented for 10s followed by a 1s blank. The four items were then presented sequentially for 5s followed by a 10s blank (Figure 3). Three seconds prior to each item (excluding the first item), a sub-section of the question was presented as a reminder (e.g., “Was this the bag?”). The first item presented in each CIT question was a buffer; a control item used to absorb the initial orienting to that item group. The four CIT blocks were then randomized again and repeated resulting in participants seeing eight CIT blocks in total. Participants were instructed to respond verbally with ‘no’ or ‘don’t know’ in response to each item.

Figure 2. The four CITs with 2D images of the VR models used in this study with crime items on the right.

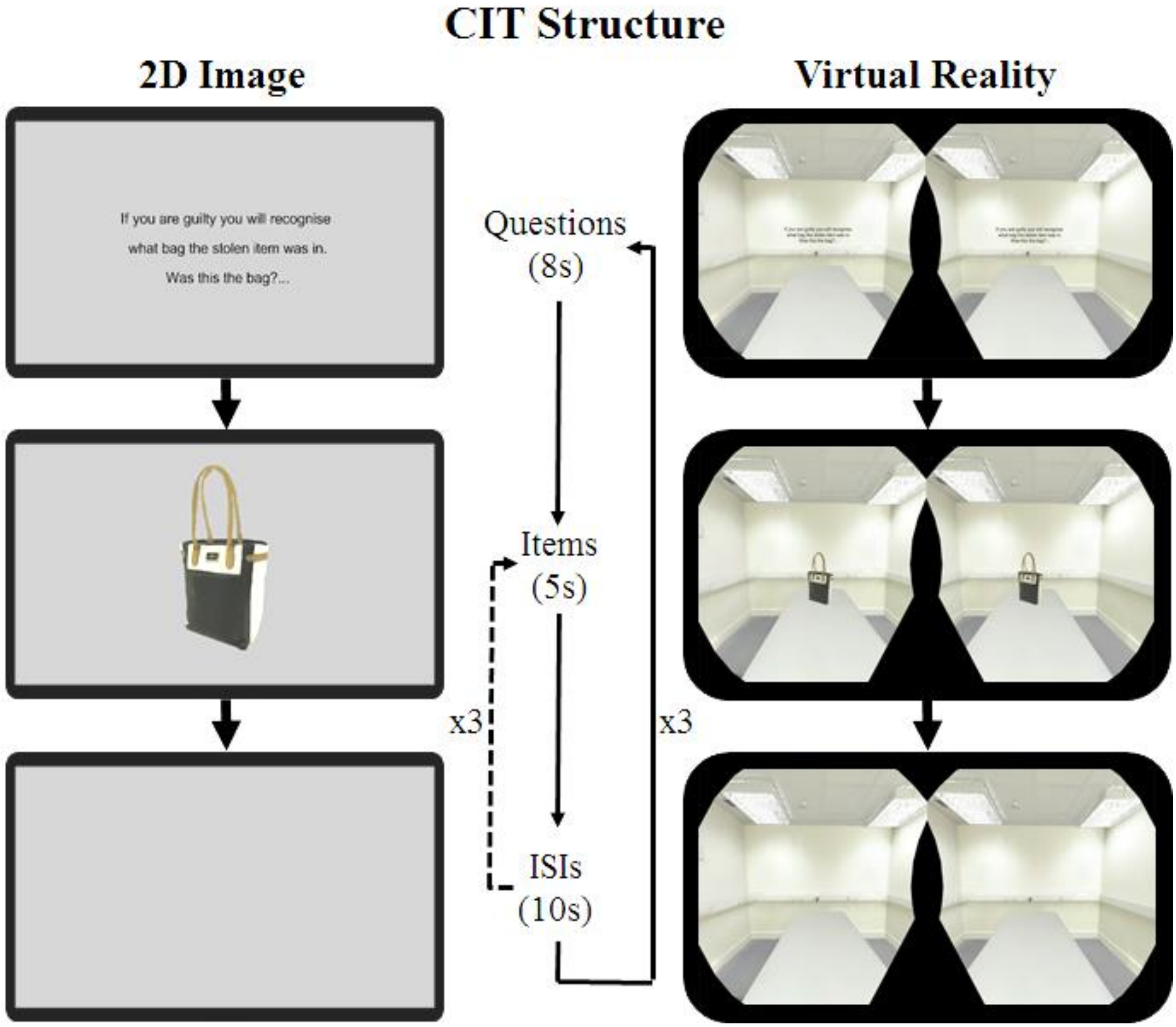

Figure 3. *The CIT structure (centre) for both the* 2D image *(left) and Virtual Reality (right) CITs. From the top, the CIT begins with the question (8s) followed by presentation of one of the three control items (5s), followed by removal of that item (10s). Another item is then presented, and this section is repeated until all four items, three control and one crime, have been presented. The next CIT question is presented, and the process is repeated until all four CITs have been presented. This is then repeated once to complete the main testing phase.*

Virtual Reality Stimuli. A FARO Focus 3D X330 Laser Scanner was used to capture multiple colored point cloud models of 11 scenes. The scanning parameters were: Resolution = 8192 pt/360^o with point cloud size = 7984 × 3414 (i.e. 27 million points) and Quality = 4x resulting in 9-minute scans. The scenes were cleared of clutter with window blinds drawn and available indoor lighting switched on. Objects were captured using a FARO Freestyle 3D handheld scanner which is designed to scan objects with a resolution <1.5mm. The resulting models were photorealistic 1:1 scale, 360 degree and 3D detailed copies of real scenes and objects (Figure 4). The VR condition contained motion from changes in the participant's viewpoint as a result of tracked head movements, however, none of the stimuli contained independent object motion. For maximal control between the VR and 2D image condition, 1920 x 1080 screenshots of the VR models were taken from the view of the participant in VR to act as the stimuli for the 2D image CIT condition. The virtual reality models used in this research are available on request from the authors.

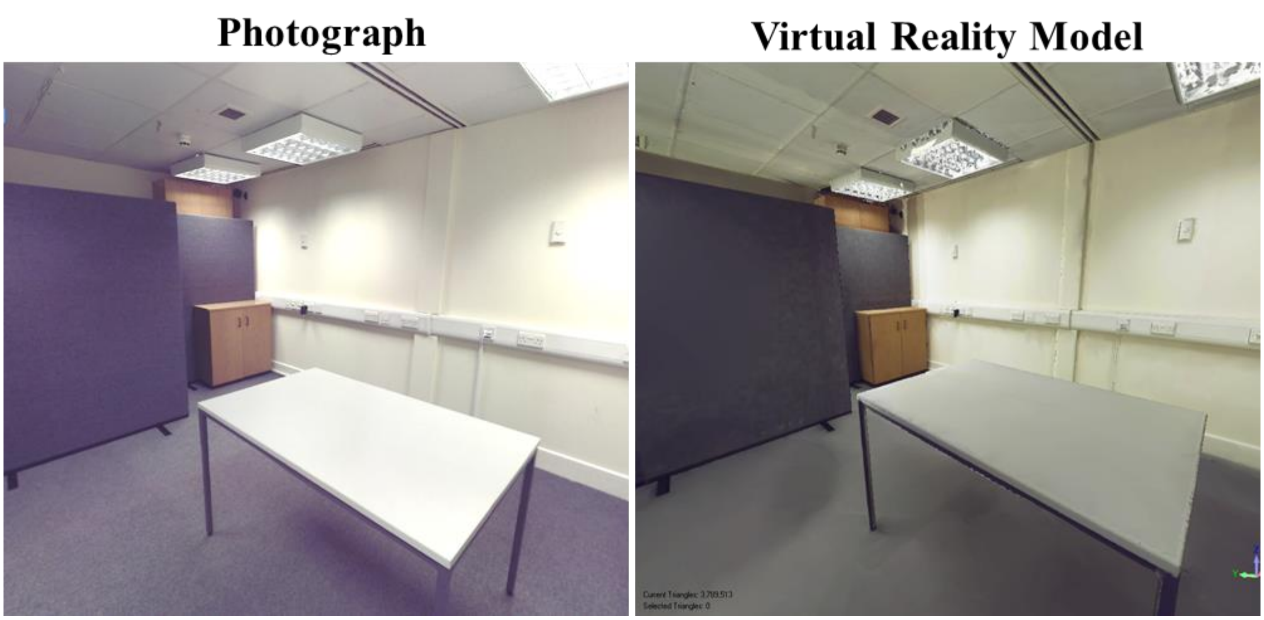

Figure 4. *Photograph of the lab where participants underwent the CIT (Left). Photorealistic virtual model of the lab used as a base for participants in the VR condition (Right).*

Physiological Data

Electrodermal activity (EDA) and heart rate were recorded using a MP36R data acquisition unit (*Biopac Systems Inc*) with pre-gelled disposable Ag/AgCL electrodes (EL507 and EL501 for EDA and heart rate respectively). EDA electrodes were attached to the distal phalanges of the first and middle finger of the non-dominant hand with EDA signals sampled at 1000Hz at ×2000 gain and filtered using a 66.5Hz low pass filter. For heart rate, Electrocardiogram (ECG) electrodes were placed in a standard Einthoven Lead I Configuration: one placed on the ventral side of the dominant wrist, another on the non-dominant lateral aspect of the distal fibula, and the third electrode utilising the EDA ground electrode placed on a non-dominant distal phalange. ECG signals were sampled at 1000Hz at ×1000 gain, with a 66.5Hz low pass filter and a 0.5Hz high pass filter. Electrodes were attached for approximately 5 minutes before data collection. A webcam recorded participants from a side view to allow noise removal if participants made substantial movements.

Skin conductance responses were defined as the difference in absolute magnitude of tonic skin conductance peaks and their respective peak onsets. Skin conductance peaks were identified using an Acq*Knowledge* v4.2 propriety algorithm (Kim, Bang & Kim, 2004), with parameters ensuring peak onsets were within a 0.5-5s window following stimulus presentation and maximum peaks within 10s (Gamer, 2011) - this output was manually checked for errors. For heart rate, an *AcqKnowledge's* propriety Heart Rate algorithm was used on the ECG signal to detect R peaks, classify the time interval between them, and automatically filter artefacts. The R-R interval was then converted to instantaneous heart rate (beats per minute) before baseline-correction via subtraction of the 3s mean heart rate prior to stimulus onset. The average baseline-corrected heart rate was calculated between stimulus onset and 15s after, resulting in the mean heart rate change measure. This measure has been shown to outperform other measures of heart rate change when analysing physiological data from the CIT (Gamer, 2011). Due to individual differences in physiological responsiveness, within-subjects standardised scores (z-scores) were calculated for each individual measure (Ben-Shakhar, 1985). Responses to mean normalised physiological responses to crime items were used to indicate the CIT effect (Meijer, Selle, Elber & Ben-Shakhar, 2014).

We removed data from a trial if there was excessive movement within a 0-2s window prior to individual stimulus onsets (Klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2016). Signals were removed if the sensors became dislocated or dislodged during the experiment. Participants with a standard deviation of raw SCR responses below 0.01μS were considered skin conductance non-responders and the EDA data were removed from analysis (Klein Selle et al., 2016). Finally, the first trial in each CIT block, always a control item, was removed prior to analysis as its sole role is to absorb the initial orienting to that CIT item group. In total five participants met our criteria for SCR non-responders and their data were excluded from the SCR analysis. Six participants' heart rate data were excluded from the analysis due to dislodged sensors; one guilty participant in the 2D image and three in the VR condition, and two innocent participants in the 2D image condition. Out of 4096 trials, 21 (0.5%) were removed from analysis due to large movement artifacts. All raw physiological data can be found in the supplementary materials.

Procedure

Participants were provided with an overview of the study procedure (but not told about the different conditions) including their right to withdraw, given the opportunity to ask any questions, and invited to provide consent and demographic information (age and gender).

Mock Crime. Participants assigned "guilty" carried out a mock crime. The experimenter[1] made it clear that it was not a real crime, but participants should try to imagine it was and therefore not get caught. Participants were asked to imagine that they were partners in crime with the experimenter, and that the experimenter had identified an unattended bag in an office that had a tablet computer poking out the front pocket. Participants were asked to discretely steal the computer before returning it to their accomplice in the common room who could then sell it for money. While viewing a floor plan of the building with photographs of the key details, participants were given the following instructions: i) Head left from the start point [highlighted on the floor plan] and find the open office H122 [highlighted on the floor plan]; ii) Knock on the door and tell the person inside that: "Danni has asked that I wait for her in here"; iii) Sit at the desk in the corner with the handbag next to it for about a minute; iv) Steal the tablet from the handbag and leave; and v) Wait for me in the psychology common room - I will be a few minutes) [highlighted on the floor plan]. Participants were not told that the people in the office knew about the experiment (which they did but were asked to ignore the participants' activities). However, if participants did directly ask whether people in the office knew about the experiment, the experimenter confirmed that they did (this happened twice). Participants typically completed the mock crime in approximately 3 minutes. Participants assigned "innocent" moved immediately to the CIT phase.

Instructions. Following the mock crime—or following the consent phase for those in the innocent group—participants were asked to imagine that they had been contacted by the

---

[1] The procedure prevented the experimenter from being blind to both the suspect and modality conditions. This was because the mock crime condition required additional instructions and management by the experimenter and the requirement to apply the head-tracked VR headset in the VR condition.

authorities informing them that they were now a suspect in a recent crime and therefore would undertake a lie detection test. They were reminded to try to appear as innocent as possible and therefore to deny any knowledge of the crime. The EDA and ECG electrodes were then applied as described in the Physiological Data section.

VR Setup. Participants in the VR condition put on and adjusted the VR headset for a comfortable fit. The default pupil distance was set as 62mm, but participants could change it in the preview if needed (no participants did). Participants were advised that they could remove the headset at any point if they felt any form of motion sickness or fatigue or for any other reason (none did). Participants were told that one potential method that could be used to fool the test would be to simply close their eyes. The experimenter instructed participants not to do this because in the real-world eye trackers could be installed in the headset to detect when people were intentionally closing their eyes. Participants were asked to remain seated during the CIT and to keep their head relatively still and forward facing. A similar instruction was given to participants in the 2D image condition.

Stimulus Preview. Previewing all items in the CIT prior to testing is recommended to reduce the novelty for each stimulus preventing a confounding orienting signal (Verschuere & Crombez, 2008). It also allowed the experimenter to visually explain the CIT procedure as well as familiarize participants with the task. In the preview, participants saw each CIT question followed by the four stimuli in a random order, either all in VR or all as 2D images depending on the condition, that accompanied that question. The stimuli lasted for 5s with no inter-stimulus interval. Following the preview, participants could ask any questions before the main CIT commenced. This phase was particularly useful for participants who had not previously used a VR headset before or experienced a virtual environment.

The CIT. Participants were reminded to appear as innocent as possible and deny all knowledge of the crime. They were also reminded that they were being filmed and that they should try to remain as still as possible. During the CIT the experimenter sat quietly behind a screen out view of the participant. The participants then underwent the CIT as described above.

Post-CIT Questionnaire. Participants were given a paper-based questionnaire consisting of multiple-choice questions to check that they had remembered the crime items. Participants were also asked to rate their motivation to beat the CIT on a 6-point scale (1 = *no motivation,* 6 = *highly motivated)*, their stress during the mock crime and CIT on a 6-point scale (1= *no stress* to 6 = *highly stressed),* how immersive they found the mock crime scenario and how well they believed they appeared innocent on a 6-point scale (1= *not immersive* 6 = *highly immersive)*. Participants were also asked to provide an open answer to the question: *"Did you do anything to try and fool the polygraph test? If you did or didn't please bullet point below – either case is fine."* Finally, participants were debriefed.

Results

Skin Conductance Responses

The key finding was that SCRs to crime items were larger for participants undertaking the VR-CIT compared to the 2D image equivalent, but only for guilty participants and not innocent participants. Mean normalized crime item SCRs were analyzed using a 2 (Modality: VR vs. 2D) x 2 (Suspect: Guilty vs. Innocent) ANOVA (Figure 5). This revealed significant main effects of Suspect, $F(1, 119) = 49.3$, $p < .001$, $MSE = 3.22$, $\eta_p^2 = .293$, SCRs were larger for guilty participants than for innocent suspects. There was a marginal difference for Modality, $F(1, 119) = 3.69$, $MSE = .24$, $p = .057$, $\eta_p^2 = .030$, SCRs to crime items were marginally larger in the VR condition than in the 2D condition. There were significant two-way interactions between:

Modality and Suspect, $F(1, 119) = 7.2$, $MSE = .471$, $p = .008$, $\eta_p^2 = .057$, SCRs to crime items were larger for participants in the VR-CIT compared to the 2D image condition but only for guilty participants. A follow-up t-test revealed that for Guilty participants, SCRs to crime items were larger in VR compared to the 2D image condition suggesting that VR enhanced guilty suspects' recognition strength of the crime related items, $t(62) = 3.26$, $p = .002$, $d = .813$ ($MD = .213$). There was no effect of Modality for innocent participants, $t(57) = .544$, $p = .589$, $d = .130$ ($MD = .035$). Finally, crime item SCRs were larger for Guilty participant compared to Innocent in both the VR, $t(61) = 6.98$, $p < .001$, $d = 1.75$ ($MD = .448$), and 2D image condition, $t(58) = 3.02$, $p = .004$, $d = .798$ ($MD = .200$).

Heart Rate Change

The ΔHR CIT effect was larger for guilty participants, but this was not affected by the modality of the CIT. Mean normalized crime item ΔHR values were analyzed using a 2 (Modality: VR vs. 2D) x 2 (Suspect: Guilty vs. Innocent) ANOVA (Figure 6). This revealed significant main effect of Suspect, $F(1, 118) = 9.9$, $MSE = .994$, $p = .002$, $\eta_p^2 = .077$, heart rate decelerated more for guilty participants than innocent participants. However, there was no Modality effect, $F(1, 118) = .001$, $MSE = 0$, $p = .980$, $\eta_p^2 = 0$, or interaction between Suspect and Modality, $F(1, 118) = .079$, $MSE = .008$, $p = .780$, $\eta_p^2 = .001$.

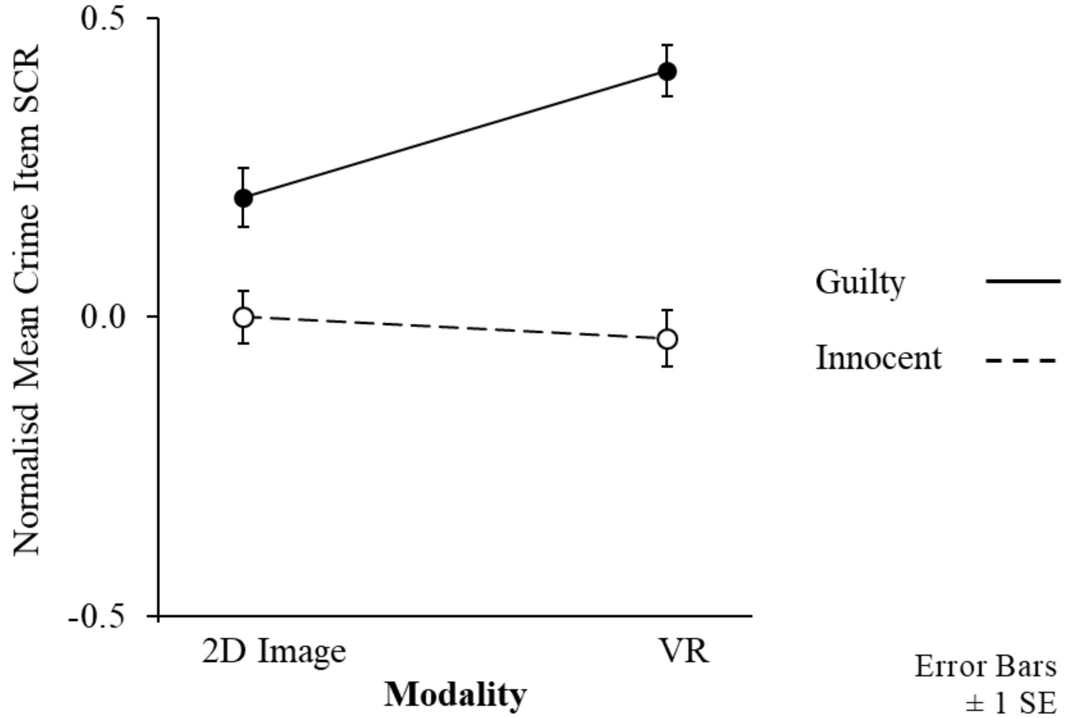

Figure 5. *Mean normalized SCR as a function of Modality and Suspect.*

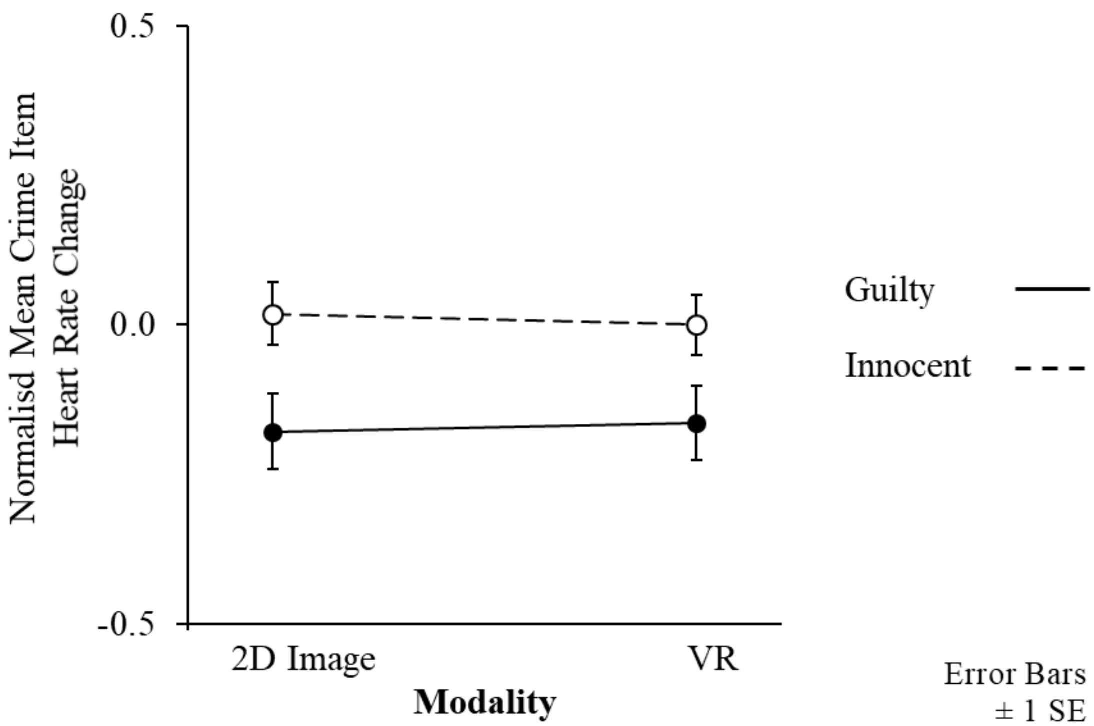

Figure 6. *Mean normalized heart rate deceleration as a function of Modality and Suspect.*

Signal Detection Analysis

To assess the efficiency of deception detection for both modalities used in this study, a signal detection analysis was conducted to determine the degree of separation between the guilty and innocent participants. A Receiver Operator Curve (ROC) was generated using the SCR data for both the guilty and innocent groups in the VR and 2D image conditions. As shown in Figure 7, the curves are closer to the upper left-hand corner of the ROC which indicates high overall accuracy (Zweig & Campbell, 1993). The area under this curve (*AUC*) allows an objective measure of diagnosticity; the accuracy trade-off between the test sensitivity and specificity. An *AUC* of 0.5 suggests no discrimination, 0.7-0.8 is considered fair, 0.8-0.9 is excellent, and 0.9+ is outstanding (Hosmer Jr, Lemeshow, & Sturdivant, 2013). In the VR condition the SCR CIT effect's *AUC* was .901 (0.822 - 0.981, $CI^{95\%}$), indicating an excellent diagnostic test with a large guilty-innocent effect size, $d = 1.75$. In the 2D image condition the SCR CIT effect's *AUC* was .709 (0.577 - 0.840, $CI^{95\%}$) indicating a fair diagnostic test and the effect size was large ($d = .798$). The detection rate in VR was significantly better that the 2D image condition, *AUC diff* = .192, $SE = .077$, $z = 2.47$, $p = .007$. Note, no effect of Modality was revealed for ΔHR therefore a combined *AUC* = .664 (0.567 - 0.761, $CI^{95\%}$), and ROC (Figure 7) were computed which indicated a limited diagnosticity with a medium guilty-innocent effect size $d = .723$.

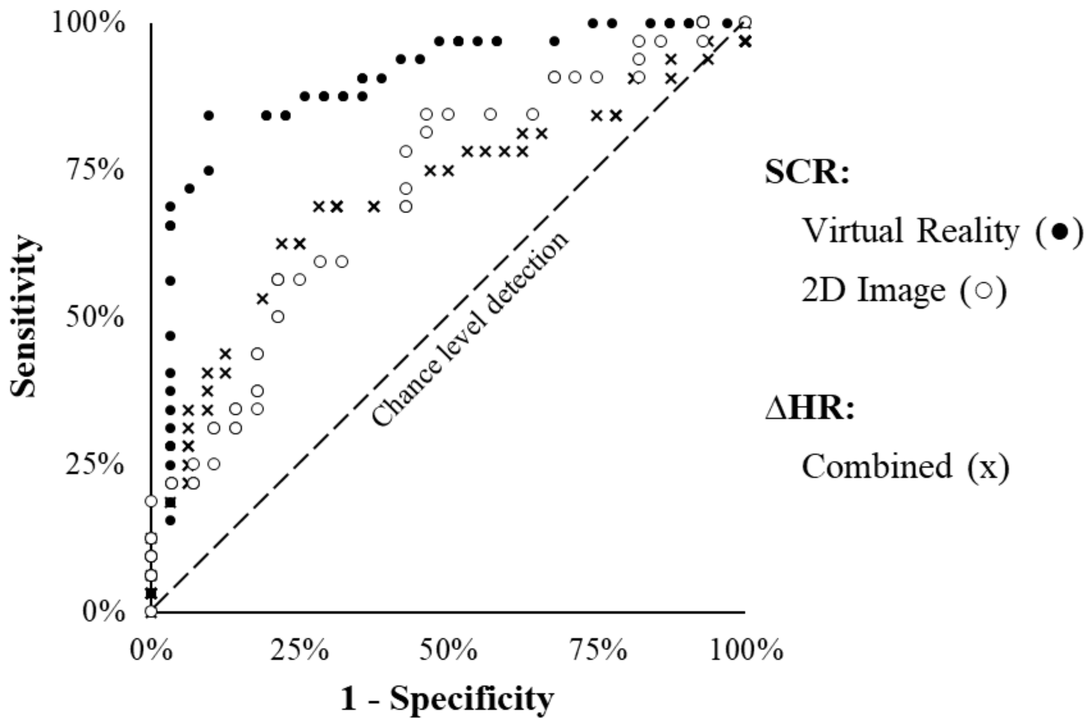

Figure 7. *Signal detection curve (ROC) showing the detection sensitivity and specificity between guilty and innocent participants who either took the CIT in VR or on a 2D computer monitor.*

Post-CIT Questionnaire

No additional factors measured in the Post-CIT Questionnaire significantly differed across between groups, suggesting that it was unlikely that they influenced the above findings. All guilty participants correctly recalled all crime items whereas innocent participants performed at chance level in the memory test, $t(63) = 1.1$, $p = .28$, $MD = .125$. Overall participants' self-reported motivation was moderate, $M = 4.9$, $SD = 0.9$ (Scale = 1_{low} to 6_{high}). Mean self-reported motivation was analysed using a 2 (Modality: VR vs. 2D) × 2 (Suspect: Guilty vs. Innocent) ANOVA. This revealed a significant main effect of Suspect, $F(1, 127) = 14.1$, $MSE = 9.6$, $p < .001$, $\eta_p^2 = .974$, with guilty participants self-reporting as more motivated ($M = 5.2$, $SD = 0.7$) than innocent participants ($M = 4.7$, $SD = 0.7$). However, there was no significant interaction between Modality and Suspect, $F(1, 127) = .56$, $MSE = .38$, $p = .455$, $\eta_p^2 \approx .0$, and no main effect of Modality, $F(1, 127) = .01$, $MSE = 0$, $p = .915$, $\eta_p^2 \approx 0$.

Overall, participants' self-reported stress during the mock crime was neutral, $M = 3.2$, SD = 1.6 (range = $1_{\text{no stress}}$ to $6_{\text{highly stressed}}$), with no significant difference between the 2D image or VR condition, $t(62) = .54$, $p = .588$, $MD = 2.19$. Overall, participants' self-reported stress during the CIT was low, $M = 2.5$, $SD = 1.3$ (range = $1_{\text{no stress}}$ to $6_{\text{highly stressed}}$). Mean self-reported stress during the CIT was analysed using a 2 (Modality: VR vs. 2D) × 2 (Suspect: Guilty vs. Innocent) ANOVA. This revealed a significant main effect of Suspect, $F(1, 127) = 17.2$, $MSE = 27.2$, $p < .001$, $\eta_p^2 = .122$, with guilty participants reporting more stress ($M = 3.0$, $SD = 1.4$) than innocent participants ($M = 2.0$, $SD = 1.1$). There was no significant interaction between Modality and Suspect, $F(1, 127) = 1.1$, $MSE = 1.76$, $p = .293$, $\eta_p^2 = .009$ and no main effect of Modality, $F(1, 127) = .84$, $MSE = 1.32$, $p = .362$, $\eta_p^2 = 007$.

In the guilty group, 50% of participants indicated that they had used a VR headset at least once before, versus 38% of the innocent, this was not significant, $\chi^2(1, N = 64) = 1.02$, $p = .313$. Twenty-six participants (21%), all from the guilty condition, reported using some form of countermeasure to fool the test: Nine reported trying to imagine a different image when the crime item appeared; sixteen reported trying to control or relax their breathing; and one tried to answer verbally "no" in the same way. The difference between the number of guilty participants attempting countermeasures in either of the Modality conditions, 16%, was non-significant, $\chi^2(1, N = 64) = 1.04$, $p = .309$.

Discussion

Our findings show that skin conductance responses, taken to indicate recognition, are enhanced when crime scenes and objects, initially encoded in the real-world, are presented as VR models as opposed to 2D images. This is important to researchers and practitioners because VR may improve the diagnosticity of the CIT as a forensic memory test. To our knowledge, our study is the first to show that the SCR CIT effect is larger for guilty suspects, but not innocent suspect, who undertake a VR-CIT compared to a 2D image equivalent. We propose that our findings indicate recognition intensity increases for scenes and objects viewed in VR compared to 2D images.

Our findings fit with well-established theories of memory. Modality congruence predicts that memories retrieved in the same or similar modality as they were encoded are stronger than if the modalities mismatch (Dewhurst & Knott, 2010). The VR stimuli likely produced a closer match to the real-world in which the memory was encoded, thereby increasing the recognition signal. Transfer-appropriate processing theory suggests that recognition increases when cognitive processes used during encoding are reinstated at retrieval (Dewhurst & Brandt, 2007; Lanen & Lamers, 2018). Our VR models closely mirrored the real-world, which likely activated systems used to process and encode the real-world equivalents. Finally, feature matching theory predicts that as the number of matching features between the test stimulus (VR crime scene) and the encoded memory (actual crime scene) increase, so too does the physiological orienting magnitude (Ben-Shakhar & Gati, 1987). We observed an increase in SCR to crime items which we propose is due to an increased feature overlap between the real-world crime details and the VR models resulting in enhanced recognition strength for the crime items.

Our proposal for our findings is that the larger SCRs to crime items in the VR condition, represents greater recognition strength. This is due to the increased feature match (size, depth etc) between the stimuli presented in VR and the participant's memory of the stimulus (Ben-Shakhar & Gati, 1987; Stelmack, Plouffe, & Winogron, 1983; Marchand, Inglis-Assaff, & Lefebvre, 2013). However, of note, there was no reliable difference between recognition in the VR versus 2D conditions in the explicit memory test presented after the CIT. Indeed, all guilty participants explicitly recognized all crime related items when given a recognition test following the CIT. It is possible that the explicit test was simply not sensitive enough to detect differences that were nonetheless detectable by the physiological SCR measure.

An alternative explanation for our findings, however, is that the crime stimuli were more salient when presented within VR than as 2D images. Previous work (e.g., Kleinberg and Verschuere, 2015, see also klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2017; Meijer, Verschuere & Ben-Shakhar, 2011; Jokinen, Santtila, Ravaja & Puttonen, 2006) has shown that items with higher personal salience (e.g., country of origin or birthday) produce a larger reaction time CIT effect than less personally salient stimuli (e.g., favorite color or animal). It is possible that the VR presentation differentially increased the salience of the crime items based on their personal relevance (e.g., related to the memory of a crime) whilst having little impact on the less salient irrelevant items. By this account a larger CIT effect (physiological orienting) would occur without a change in explicit memory – as was found in our study. Although the present data do not allow us to differentiate between these alternatives' explanations, this would be a useful goal for future research.

In contrast to the SCR findings, there was no effect of modality (VR/2D-image) on heart rate indicating that only one of our hypotheses was confirmed. Response fractionation theory attempts to explain why SCR and heart rate (along with other parasympathetic measures) do not always correlate (Verschuere, Meijer, & De Clercq, 2011). This theory postulates that while SCRs are related to recognition via orienting processes, heart rate change reflects arousal inhibition experienced when actively concealing and suppressing recognition (Klein Selle et al., 2016; Klein Selle et al., 2017; Klein Selle, Verschuere, & Ben-Shakhar, 2018). One might expect that if VR facilitates a stronger recognition of the criminal activity and associated emotional arousal, then a greater amount of inhibition would be required by the guilty suspect – this was not found. One possibility for this is that inhibition might already have been at ceiling in the 2D image condition leaving no additional inhibition to be measured within the VR condition. Further exploration of this fractionation finding would be beneficial.

Motivation to beat the CIT (Ben-Shakhar & Elaad, 2003), stress during encoding and retrieval (Verschuere, Ben-Shakhar, & Meijer, 2011) and attention during retrieval can influence the CIT. However, none of these factors appear to account for our findings. Specifically, there were no significant differences in ratings of motivation or stress between modality conditions. There was also no difference in previous VR experience between our modality conditions. Indeed, we attempted to mitigate against both the novelty of the VR and the novelty of each scene and object stimulus by having participants preview all questions and items before the CIT (Verschuere & Crombez, 2008). A noteworthy issue with our study was that, compared to the innocent group, guilty participants took part in a longer (approximately 3 minutes) and more complex procedure. Although this is unlikely to have impacted our VR effect, it is of course possible that it may have interfered with the suspect effect and interaction. Further work could test this by having innocent participants carry out a similarly complex and timely task that does not relate to the mock crime.

Although not part of the planned analysis, we explored whether there was any impact on whether the CIT stimuli were Scenes (Office and Common Room) or Objects (Handbag and Tablet) on physiological responses. This could have modulated the results as viewing scenes in VR compared with 2D images is different both quantitatively (size and scale) and qualitatively (being inside the scene, the level of immersion). In contrast the difference between VR and 2D image presentation is smaller for individual objects. Additionally, scene and object images are processed differently within the brain (Oliva & Torralba, 2006). Nevertheless, including this factor in our analysis revealed no significant interactions or main effects of stimulus type (all ps > .05) thus providing no evidence that recognition strength was modulated by whether the stimulus was an object or scene.

The typical diagnosticity in mock crime paradigms with SCR is approximately AUC = 0.84, (0.83 - 0.87, $CI^{95\%}$) (Meijer et al., 2014). Although our VR-CIT AUC is relatively high (AUC = 0.901), the diagnosticity for our 2D image condition is relatively low (AUC = .710). This could be due to the smaller number of control items (three instead of the four typically used) and the smaller number of CIT questions (four instead of five) used in our study. Notably, the SCR diagnosticity for our 2D-image condition was within the 95% confidence range for mock crime studies with only four CIT questions ($AUC_{4\ CITs}$ = 0.81, 0.71 - 0.88 $CI^{95\%}$, Meijer et al., 2014).

The benefits of using VR to increase ecological validity while maintaining experimental control is well documented (Krokos, Plaisant, & Varshney, 2019; Parsons, 2015; Reggente et al., 2018) and other work demonstrates possible clinical applications (Negut, Matu, Sava, & David, 2016). Creating photo-realistic VR models from real-world scenes however, requires specialist technology, time and expertise, and presents complications such as how to deal with a sky in outdoor scenes or how-to laser scan reflective surfaces. Nevertheless, the use of digitally captured objects and scenes is increasing across a range of industries, including forensic crime scene documentation, and as the technology validation cycle continues, the more accessible and user friendly these technologies will become.

Currently in lab-based and applied CITs, images or words are presented to suspects to elicit physiological recognition responses. Models of memory suggest that returning suspects to the actual real-world crime scene would elicit the greatest recognition response. However, this would not be possible in a CIT as many stimuli are required to be presented sequentially in a tightly controlled and timed manner; additionally, crime scenes typically change over time which would weaken their match with the culprit's memory. Laser scanning the crime scenes and/or objects and presenting them to suspects in VR is the next best option. This approach means that the suspect can be visually 'taken back' to the crime scene without physically leaving the interview room. As more and more crime scenes and objects are digitally scanned, these virtual scenes and objects could potentially form a database of CIT stimuli ready for use as control items within a VR-CIT – much like the database of digital faces that are drawn on as foils in police line-ups. Clearly the adoption of a VR-CIT procedure would be a radical change. Thus, further systematic study of the VR-CIT as a memory detection test and the efficacy and reliability of a potential 'VR superiority effect' will be essential. Nonetheless our findings provide a promising start.

References

Bailenson, J. N., Davies, A., Blascovich, J., Beall, A. C., McCall, C., & Guadagno, R. E. (2008). The effects of witness viewpoint distance, angle, and choice on eyewitness accuracy in police lineups conducted in immersive virtual environments. Presence: Teleoperators and virtual environments, 17(3), 242-255.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. Psychophysiology, 22(3), 292-299.

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.

Ben-Shakhar, G., & Gati, I. (1987). Common and distinctive features of verbal and pictorial stimuli as determinants of psychophysiological responsivity. Journal of Experimental Psychology: General, 116(2), 91.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic.

Deng, X., Rosenfeld, J. P., Ward, A., & Labkovsky, E. (2016). Superiority of visual (verbal) vs. auditory test presentation modality in a P300-based CIT: The Complex Trial Protocol for concealed autobiographical memory detection. International Journal of Psychophysiology, 105, 26-34.

Dewhurst, S. A., & Brandt, K. R. (2007). Reinstating effortful encoding operations at test enhances episodic remembering. The Quarterly Journal of Experimental Psychology, 60(4), 543-550.

Dewhurst, S. A., & Knott, L. M. (2010). Investigating the encoding-retrieval match in recognition memory: Effects of experimental design, specificity, and retention interval. Memory & cognition, 38(8), 1101-1109.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191.

Gamer, M. (2011). Detecting concealed information using autonomic measures. Memory detection: Theory and application of the Concealed Information Test, 27-45.

Godden, D. R., & Baddeley, A. D. (1975). Context‐dependent memory in two natural environments: On land and underwater. British Journal of psychology, 66(3), 325-331.

Granhag, P. A., Vrij, A., & Verschuere, B. (Eds.). (2015). Detecting deception: Current challenges and cognitive approaches. John Wiley & Sons.

Häfner, P., Häfner, V., & Ovtcharova, J. (2013). Teaching methodology for virtual reality practical course in engineering education. Procedia Computer Science, 25, 251-260.

Hahm, J., Ji, H. K., Jeong, J. Y., Oh, D. H., Kim, S. H., Sim, K. B., & Lee, J. H. (2009). Detection of concealed information: combining a virtual mock crime with a P300-based Guilty Knowledge Test. Cyberpsychology & behavior, 12(3), 269-275.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. Memory & Cognition, 36(7), 1351-1359.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

Jokinen, A., Santtila, P., Ravaja, N., & Puttonen, S. (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. International Journal of Psychophysiology, 62(1), 175-184.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. Medical and biological engineering and computing, 42(3), 419-427.

Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test. PloS one, 10(4), e0118715.

Klein Selle, N., Verschuere, B., & Ben-Shakhar, G. (2018). Concealed Information Test: Theoretical Background. In Detecting Concealed Information and Deception (pp. 35-57). Academic Press.

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.

Klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. Psychophysiology, 54(4), 628-639.

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. Virtual Reality, 23(1), 1-15.

Lanen, M., & Lamers, M. H. (2018, October). Context-Dependent Memory in Real and Virtual Reality. In International Conference on Virtual Reality and Augmented Reality (pp. 177-189). Springer, Cham.

Marchand, Y., Inglis-Assaff, P. C., & Lefebvre, C. D. (2013). Impact of stimulus similarity between the probe and the irrelevant items during a card-playing deception detection task: The "irrelevants" are not irrelevant. Journal of clinical and experimental neuropsychology, 35(7), 686-701.

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed I nformation T est: A meta analysis of skin conductance, respiration, heart rate, and P300 data. Psychophysiology, 51(9), 879-904.

Negut, A., Matu, S. A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. The Clinical Neuropsychologist, 30(2), 165-184.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. Progress in brain research, 155, 23-36.

Osugi, A. (2011). 14 Daily application of the Concealed Information Test: Japan. Memory detection: Theory and application of the Concealed Information Test, 253.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Frontiers in human neuroscience, 9, 660.

Puente, I., González-Jorge, H., Martínez-Sánchez, J., & Arias, P. (2013). Review of mobile mapping and surveying technologies. Measurement, 46(7), 2127-2145.

Reggente, N., Essoe, J. K. Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the ecological validity of fMRI memory research using virtual reality. Frontiers in neuroscience, 12.

Rosenfeld, J. P., Ward, A., Frigo, V., Drapekin, J., & Labkovsky, E. (2015). Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed autobiographical memory detection. International Journal of Psychophysiology, 96(1), 16-22.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. Applied Cognitive Psychology, 6(2), 125-139.

Staudigl, T., & Hanslmayr, S. (2018). Reactivation of neural patterns during memory reinstatement supports encoding specificity. bioRxiv, 255166.

Stelmack, R. M., Plouffe, L. M., & Winogron, H. W. (1983). Recognition memory and the orienting response: An analysis of the encoding of pictures and words. Biological Psychology, 16(1-2), 49-63.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373

Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. Biological psychology, 43(2), 87-101.

Verschuere, B., & Crombez, G. (2008). Déjà vu! The effect of previewing test items on the validity of the Concealed Information polygraph test. Psychology, Crime & Law, 14(4), 287-297.

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

Verschuere, B., Meijer, E., & De Clercq, A. (2011). Concealed information under stress: A test of the orienting theory in real-life police interrogations. Legal and criminological psychology, 16(2), 348-356.

Voinescu, A., & David, D. (2019). The Effect of Learning in a Virtual Environment on Explicit and Implicit Memory by Applying a Process Dissociation Procedure. International Journal of Human- Computer Interaction, 35(1), 27-37.

Zheng, H., Rosenfeld, J. P., Deng, X., Lu, Y., Xue, C., Wang, Y., ... & Ouyang, D. (2019). Visual presentation modality's superiority in the detection of concealed information: A comparison of the efficiencies of the P300-based Complex Trial Protocol in visual versus auditory modalities. International Journal of Psychophysiology, 137, 32-40.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), 561-577.