Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

http://wrap.warwick.ac.uk/138186

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Seen this scene?

Scene recognition in the reaction-time concealed information test

Danielle G. Norman, Daniel A. Gunnell, Aleksandra J. Mrowiec &

Derrick G. Watson

University of Warwick

Abstract

Detecting a suspect's recognition of a crime scene (e.g. a burgled room or a location visited for criminal activity) can be of great value during criminal investigations. Although it is established that the reaction-time Concealed Information Test (RT-CIT) can determine whether a suspect recognizes crime related *objects*, no research has tested whether this capability extends to the recognition of *scenes*. In Experiment 1, participants were given an autobiographic scene-based RT-CIT. In Experiment 2, participants watched a mock crime video before completing an RT-CIT which included both scenes and objects. In Experiment 3, participants completed an autobiographic scene-based RT-CIT, with half instructed to perform a physical countermeasure. Overall, the findings showed that an equivalent RT-CIT effect can be found with both scene and object stimuli and that RT-CITs may not be susceptible to physical countermeasure strategies thereby increasing its real-world applicability.

Keywords: Deception Detection; Reaction Time Concealed Information Test; RT-CIT; Scenes; Recognition Memory

Introduction

A key objective in forensic science is to link a culprit(s) to the crime scene (Fisher, 2004). DNA matching, tread mark or fingerprint analysis can be used to establish a connection between the culprit and the crime, however, such physical evidence, is not always available or adequately preserved (Peterson, Sommers, Baskin, & Johnson, 2010). In these situations, evidence of a connection between the culprit and crime often remain solely within the culprit's memory which the suspect will try to conceal. Information that a suspect may conceal recognition of include: i) Indoor or outdoor scenes of the crime, e.g., a room burgled; ii) Scenes which the culprit frequently visits to conduct criminal activity e.g., the transfer of illegal contraband; iii) Scenes where the culprit has hidden something or someone of interest e.g., a murder victim; iv) Scenes where the culprit has conducted reconnaissance for the purpose of planning criminal, military or terrorist activity; v) Protected facilities, e.g., government or military bases, accessed without authorization and; vi) Autobiographic scenes relating to locations that a person of interest denies recognizing e.g., schools, homes, workplaces. These are just some examples where detection of a culprit's concealed recognition of relevant scenes would be beneficial.

The Concealed Information Test (CIT) is a cognitive test designed to determine whether a suspect is concealing knowledge of hidden crime information that only the culprit would recognize (Ben-Shakhar & Elaad, 2003). Typically, the CIT determines a suspect's hidden recognition of crime details via analysis of their physiological response (typically skin conductance) to *crime* items (*probes*) compared with their responses to non-crime *control* stimuli, (*irrelevants*). Compared with control items, crime items elicit larger skin conductance responses, indicating

an increased level of orienting, taken to indicate recognition, to those stimuli (for a review see Verschuere, Ben-Shakhar & Meijer, 2011). This physiology-based CIT is well established and frequently applied in real-world cases, namely within the Japanese criminal justice system (Osugi, 2011).

An alternative, recently developed memory detection paradigm is the reaction time CIT (RT-CIT) which measures reaction times (RTs) instead of physiological responses. The RT-CIT relies on the idea that concealing knowledge of a crime requires inhibition of the truth leading to an overall slowing of responses made to crime-related items (see Verschuere, Suchotzki, & Debey, 2014). Compared to the physiological-CIT, the RT-CIT is less expensive, faster, easier to administer and analyze and does not require specialist training or equipment. The RT-CIT is an oddball task in which participants respond, using a keyboard, to a sequence of briefly presented stimuli. Each trial consists of the presentation of one of three types of stimulus: a *crime* (aka probe), a *control* (aka irrelevant), or a *target* item (Figure 1). Crime items are details that guilty a participant recognizes and control items are unrelated to the crime but are matched to crime items on relevant characteristics. For example, if the crime item was a set of bolt-cutters used to break a lock, control items would be equivalent tools such as a hacksaw or hammer. Participants are instructed to respond 'No' to indicate that they do not recognize either the crime or the control items. Target items were the third stimulus type that the participants were shown before the test and were instructed to respond 'Yes' to and do not related to the crime information in question. Without these items, participants could simply respond 'No' on every trial without processing the stimuli.
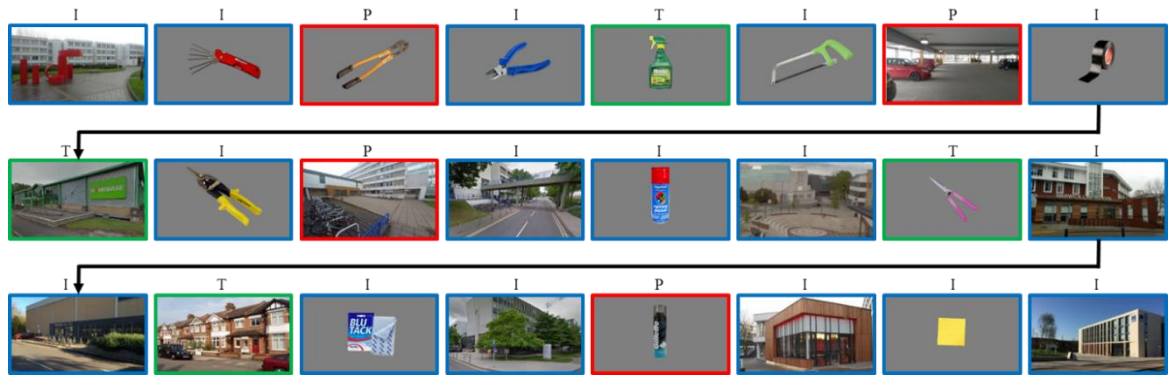
Figure 1. Example of one block of 24 images in an RT-CIT containing four crime items (red and 'P') each with four controls (blue and 'I') and one target (green and 'T').

Typically, guilty suspects are slower to respond 'No' (untruthfully), that they do not recognize a crime item, than they are to respond 'No' (truthfully), that they do not recognize a control item. It is proposed that this slowing reflects response inhibition experienced by participants whilst they resolve the conflict between recognizing an object yet reporting that they do not (Debey, Ridderinkhof, De Houwer, De Schryver, & Verschuere, 2015). Furthermore, this conflict sometimes results in an increased number of errors, for example, pressing 'Yes' to the crime items or 'No' to the target items (Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017). The target items serve no diagnostic role and are presented to ensure that participants process the stimuli and engage with the task. The diagnostic for the RT-CIT is similar to the physiological-based CIT, with a large CIT effect size $d = 1.3$, and an $AUC = .82$ [.77 - .87 CI$^{95\%}$], (Suchotzki et al., 2017).

To date, RT-CIT studies have typically used word stimuli (Eom, Sohn, Park, Eum, & Sohn, 2016; Hu, Evans, Wu, Lee, & Fu, 2013; Kleinberg & Verschuere, 2016; Kleinberg & Verschuere, 2015; Noordraven & Verschuere, 2013; Seymour & Kerlin, 2008; Seymour, Seifert, Shafto, & Mosmann, 2000; Seymour & Fraynt, 2009; Verschuere, Kleinberg & Theocharidou, 2015; Verschuere, Crombez,

Degrootte, & Rosseel, 2010; Visu-Petra, Miclea, & Visu-Petra, 2012; Visu-Petra, Varga, Miclea, & Visu-Petra, 2013; and Visu-Petra, Miclea, Buş, & Visu-Petra, 2014), with only a handful having used images. Moreover, those that have presented images have only used pictures of discrete objects that can be easily recognized and labeled (Visu-Petra, Jurje, Ciornei, & Visu-Petra, 2016; Varga, Visu-Petra, Miclea, & Visu-Petra, 2015; Suchotzki, Verschuere, Peth, Crombez, & Gamer, 2015). For example, Visu-Petra and colleagues (2016) used images of objects (backpacks, watercolors, pencils, and erasers) to test the effectiveness of the RT-CIT in children. In studying the effects of emotional valence, social factors and individual differences in the RT-CIT, another study presented pictures of objects (e.g., 'memory sticks', 'laptop bag', 'mobile phone', 'wireless mouse' and an 'agenda') and found that responses to crime and control items differed with a large effect size, Cohen's $d =$ 1.05 (Varga et al., 2015). Similarly, another experiment used images of objects (e.g. '50 euro note', 'laptops', 'CDs', 'markers', 'water crates', 'suitcase' and so on) and again found that crime and control items differed with a large effect size, $d = 1.24$ (Suchotzki et al., 2015).

Clearly, images of objects can be used effectively in the RT-CIT to detect recognition. However, being able to detect the concealed recognition of scenes might also greatly assist investigations by linking the suspect to a crime scene rather than to an object. This would not only open up more crime scene related details for use within a CIT but also allow investigators to generate a more complete picture of what the suspect recognizes about a particular crime. The RT-CIT is one potential tool for achieving this, however, to the author's knowledge the effectiveness of the RT-CIT for scene recognition has not been examined. However, this is not to say that scene stimuli have not previously been used in physiological-based CITs in both research

and the field, for example, a recent lab-based study used scene stimuli in their CIT (Norman, Wade, Williams & Watson, 2020).

Whilst there is extensive research on object recognition, (Ganis & Kutas, 2003) the nature of scene memory and its underlying mechanisms are under debate (Oliva & Torralba, 2006; Behrmann & Plaut, 2013). However, it is clear that scenes differ from objects in terms of processing, encoding, and recognition, all of which could change their effectiveness within the RT-CIT (Behrmann & Plaut, 2013). For example, memory for scenes is remarkably robust with participants able to encode and recall thousands of scenes that are previously novel to them (Standing, 1973; Standing, Conezio, & Haber, 1970; Kent, Lamberts, & Patton, 2018). Furthermore, whereas objects can be encoded in a single exposure, whole scenes seem to be encoded into memory over several glances building the memory incrementally over seconds. For example, one study found that approximately 60% of scene details were remembered after a one second viewing duration and over 80% of details recalled after ten seconds viewing (Melcher, 2006). Another key object-scene difference is that scenes usually contain a complex mixture of objects and features that could capture attention in differing ways. Some attention-grabbing objects/features (e.g., people of objects in the scene) might not be related to the crime thus rendering those scene stimuli undiagnostic. Finally, compared to objects, scenes cannot always be easily semantically labeled, with those that can, e.g., a 'beach', being more familiar than those that cannot. Relatedly, scene recognition is possible even without being able to identify exactly what or where the scene is; Recognition Without Identification (Cleary & Reyes, 2009). RTs to scene stimuli are seldom compared with objects (Ganis & Kutas, 2003), despite being an important factor when considering the use of scenes in the CIT. Global scene information is believed to be

processed rapidly and parallel to local object processing and requires fewer attentional resources than objects (Munneke, Brentari, & Peelen, 2013) with scene color processed very rapidly (~50ms, Wichmann, Sharpe & Gegenfurtner, 2002). This ability is potentially attributable to the Parahippocampal Place Area that appears to respond only to scenes and not to objects (Oliva & Torralba, 2006). This rapid initial understanding of a scene is called 'scene gist' and is achieved very quickly (20ms, Oliva & Torralba, 2006).

Testing whether a suspect recognizes a scene can be beneficial and the RT-CIT could be an appropriate test for doing so. However, as described there are differences in how scene and object stimuli are processed and the possible effects this can have on response times and detection in the RT-CIT is currently untested. In the current study, Experiment 1 established whether scenes produce a comparable RT-CIT effect to those found for object stimuli in the literature. Participants completed an RT-CIT in which they were instructed to conceal knowledge of autobiographic University campus scenes. Experiment 2 compared scenes and objects by having participants watched a mock crime video before completing an RT-CIT that contained both object and scene stimuli. Experiment 3 tested the susceptibility of scene-based RT-CITs to a physical countermeasure strategy aimed at slowing response to control items.

Experiment 1: RT-CIT with scene stimuli

Previous work has demonstrated that images of objects and scenes are processed, encoded and recognized differently and any of these factors could influence the effectiveness of scene stimuli in the RT-CIT. In Experiment 1, a scene-based RT-CIT was tested to determine whether it produced a RT-CIT effect similar to that found for object-based RT-CITs. Using autobiographic scenes of the participants' University campus, a scene-based RT-CIT was given to participants who were instructed to conceal recognition of their University.

Method

Participants

Previous RT-CIT experiments which have used images of objects as stimuli have shown large CIT effect sizes ranging from $d = 1.05$ to 1.24 (Suchotzki et al., 2015; Varga et al., 2015; and Visu-Petra et al., 2016). Given that the current study was the first reported scene-based RT-CIT, a smaller, but still relatively large, effect size was estimated[1]. A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), with a CIT effect size (crime minus control) of $d = 0.8$, and $\alpha = 0.05$ for a single group, suggested that 23 subjects would be sufficient for a power of 0.95. Thirty-six participants (25 women), aged between 18-32 years (*Mean* = 20.4, *SD* = 2.8) were recruited through a University of Warwick online participant panel. Participants received £3 payment for taking part in the 30-minute testing session.

---

[1] On reflection our initial estimation of an RT-CIT within-subject effect size of $d = 0.8$ for scene stimuli may have been optimistic given that our review of the literature suggested that the RT-CIT effect size with scenes might have been smaller than with objects. However, a sensitively analysis for this experiment was computed using G*Power which revealed a minimum detectable CIT effect size of d = 0.618 – note this was much smaller than the actual CIT effect size revealed for this experiment, d = 1.48 suggesting our design was suitable.

Materials

The image stimuli were photographs of scenes that typically contained landscapes, buildings, and other structures. The autobiographic images ('crime' items), were images of various scenes of the participants' university campus. For each crime item, four matched control scene stimuli were sourced using Google's Reverse Image Search function with the crime items as reference images. This resulted in a selection of structurally similar scenes based on low-level local features such as color, contrast/brightness, texture and shape at specific parts of the images (Chechik, Shalit, Sharma, & Bengio, 2009; Horváth, 2015). From this selection, we chose four control images based on their content similarity to the crime images. This allowed for the matching of appropriate control items for all crime scenes. In addition to the crime and control stimuli, five images of another University were used as target items. At the start of the study participants chose five scenes of their university campus, out of a selection of twenty, that they felt were most familiar to them which became the crime items. This was to ensure optimal encoding of these items prior to testing more ecologically valid scenarios. All images were open source, cropped to remove potential noise (e.g., people), were resampled to 1366 x 768 pixels and presented full-screen on a 21" LCD monitor, 16:9 aspect ratio at a resolution of 1920 x 1080 pixels (Figure 2). Participants sat approximately 40cm from the screen with the center of the screen at approximately eye level.

Figure 2. Examples of control, crime and target scenes used in Experiment 1.

The CIT

The RT-CIT consisted of 450 images with 30 images (one block) repeated 15 times. Each block of 30 images contained five CITs and each CIT consisted of six images: a crime item, a target, and four control items. There was a short break of 3000ms after each block and a longer break of 30 seconds after every three blocks. The image duration was 800ms with a randomly selected inter-stimulus interval of either 500, 750 or 1500ms.[2] Items within each block were presented sequentially in a random order with the constraint that two crime items could not occur consecutively. The targets were randomly presented within each block and did not change irrespective of what crime items were selected by participants. Target items were not analysed as they were only used to ensure participant engagement with the stimuli.

---

[2] Due to a technical error, the first half of the participants saw an extended inter-stimuli interval of 1000, 1500 or 3000ms (instead of 500, 750 or 1500ms). However, ISI length (intended vs. extended) did not interact with control and crime item RTs, $F(1, 34) = 3.363$, $p = .075$, or % error rates $F(1, 34) = .114$, $p = .738$, and therefore results were collapsed over ISI length.

The data from participants with error rates (i.e., pressing "Yes" to a crime item) above 50% were removed from further analysis as it is unlikely that they were following the task instructions. Responses faster than 200ms or slower than 800ms were removed, as recommended in the literature (Verschuere et al., 2015). Incorrect responses were also removed from the RT analysis.

Procedure. Participants completed the experiment in a computer lab in two groups of 18. Participants were provided with an overview of the procedure, given the opportunity to ask questions and provided consent and demographic information. They were informed of their right to withdraw at any point without penalty or reason. Participants were then asked to imagine that they *"are an undercover spy from Warwick University and have infiltrated New York University to steal their latest research. New York University Security suspects a mole and are therefore requiring all staff to sit a 'lie detection test.' Their 'lie detection test' assumes that spies will be slower to recognise and make more mistakes when they respond to images of New York University[3]. They are also hoping to catch spies that accidentally respond "Yes" to images of Warwick University who they believe are the prime suspects."* Participants were then given five images of 'New York University' and told to memorise these to help them beat the lie detection test. Participants were then told that *"during the 'Lie Detection Test' you will be shown a series of items of scenes. Many of these items will be unfamiliar to you except the ones relating to Warwick University (which you must keep secret) and the scenes of New York University which you have just memorised. Each image will appear for around 1 second with*

---

[3] Although this is not how the RT-CIT test works as the target items, here the New York University images, are discarded from the analysis, this instruction was provided to participants to ensure that they were attentive to the task without revealing how the test worked.

less than a second gap between them. Using the keyboard, please respond to these

images as fast as you can making as few errors as possible! The question to consider

for each image is 'Do you recognise this scene?'"

Participants were instructed to press the LEFT KEY for "Yes" responses, i.e. New York University Images (targets), and the RIGHT KEY for "No" i.e. University of Warwick (crime items) and any other images (control items). Participants were given the opportunity to ask any questions before completing a practice test consisting of two blocks of trials (60 images). During the practice stage only, if the response was incorrect e.g. a "Yes" response to a crime item, the words "Wrong" were displayed until the start of the next trial. If a response time exceeded 800ms the words "Too Slow" were displayed until the start of the next trial. Participants were aware that this information would not be provided following the practice stage. Participants then completed the main test followed by debriefing. All studies were approved by the departmental ethics committee at the authors' institution.

Results

Reaction Times. No participant's data were removed due to error rates above 50%. Including target items, trials that exceeded the response deadline (1.3%), were faster than 200ms (0.19%) and incorrect trials (3.61%) were removed from the analysis (incorrect responses are used for the error analysis). Mean correct RTs were calculated for crime and control items for each participant and overall means are shown in Figure 3. A paired t-test on Item Type revealed that RTs were significantly slower for crime items compared to the control items, $t(35) = 8.87$, $p < .001$, $d = 1.48$, $(MD = 29.4)$.

Error Rates. Error rates were low overall (*Mean* = 1.34%, *SD* = 2.35 and *Mean* = 0.97%, *SD* = 2.47 for crime and control items respectively) and did not differ significantly, $t(35) = 1.26$, $p = .215$, $d = .279$, $(MD = .926)$.
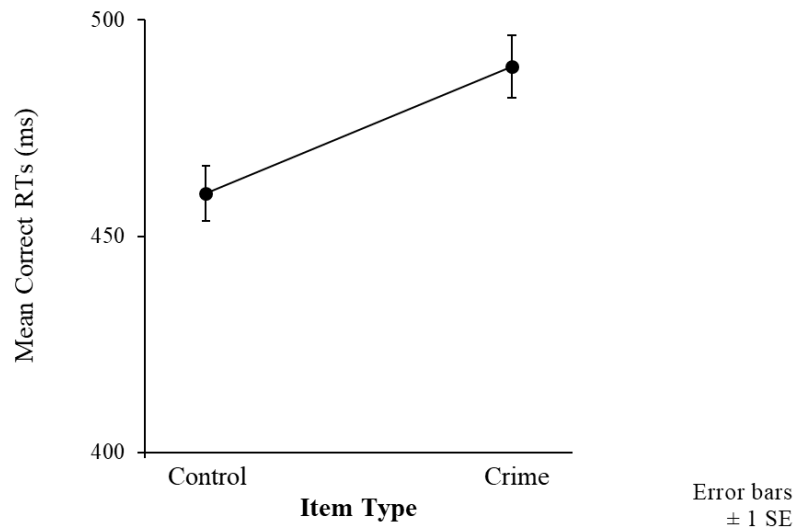
Figure.3. Experiment 1. Mean correct RTs as a function of Item Type.

Signal Detection Analysis

To assess the efficiency of detection, signal detection analysis was used to determine the degree of separation between the participants in our experiment who were considered 'guilty' and an equivalent innocent group. First, responses to each trial from each guilty participant were converted to within-subjects standardised scores (z-scores) (Ben-Shakhar, 1985). Given that no innocent participants were tested, data for innocent participants were simulated by the standard method used in the CIT literature (e.g. Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Visu-Petra et al, 2013; and Meijer, Smulders, Johnston, & Merckelbach, 2007). This approach assumes that innocent participants, not knowledgeable about the crime items, respond in the same manner to all items. Therefore, the procedure for simulating innocent participant data involves drawing random RTs from a standard normal distribution. This was conducted for each trial with one trial in five then randomly chosen to represent the simulated crime item. Once calculated for each participant, an ROC was generated to approximate signal detection using the within-subject scored RT-CIT effect (crime minus control item) for the 'guilty' group and for the normalized simulated 'innocent' group. ROCs are based on a comparison of

two detection score distributions, where detection score of guilty was defined as the mean normalized difference between crime and control items and the detection score of innocents was similarly defined but using the simulated crime and control responses.

As shown in Figure 4, the curve is close to the upper left-hand corner of the ROC, which indicates a high overall accuracy (Zweig & Campbell, 1993). The area under this curve (AUC) allows an objective measure of the accuracy trade-off between the test sensitivity and specificity. In our scene-based RT-CIT the AUC = .919 (Figure 4) and meshed with the large guilty-innocent effect size d = 1.93. Note that this effect size is the between-subjects effect size for guilty verses innocent participants as opposed to the within-subject mean RT difference between crime and control items for guilty participants.
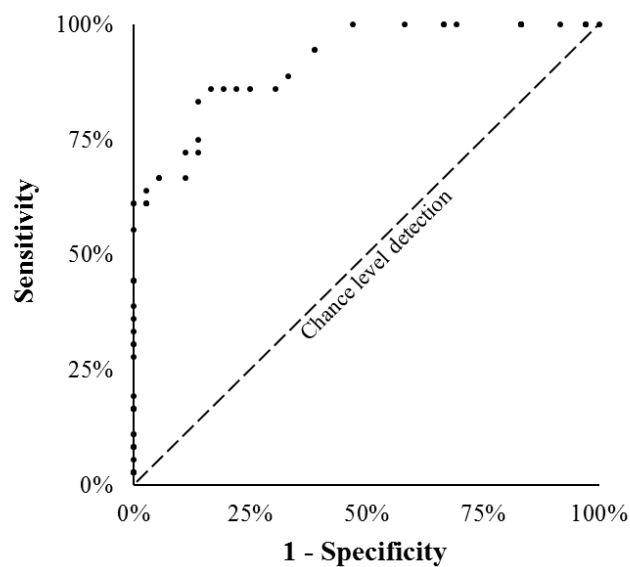

Figure 4. Experiment 1. Signal detection curve (ROC) showing the detection sensitivity and specificity between guilty and simulated innocent participants

Discussion

The finding from this experiment suggests that scene stimuli can be as effective as object picture stimuli when used within an RT-CIT. To our knowledge, this is the first attempt to experimentally test scene stimuli in the RT-CIT, however, there are some limitations. First, the scenes used represented autobiographic details which would rarely be used in a field setting. Second, participants were given the choice of five out of a set of twenty scenes to act as the crime item. Clearly this limits the generalizability of these initial findings as, in a field CITs, participants would not have this choice. This compromise on ecological validity was chosen to ensure maximal encoding of these scenes had taken place to allow for initial testing of the scene-based RT-CIT under optimal conditions.

Despite a large within-subject effect size based on RT differences, there was no CIT effect for error rates. Some studies have found differences in error rates for crime and control items however this is not always the case (Visu-Petra et al, 2016; Hu et al, 2013; Noordraen & Verschuere, 2015). In Experiment 1, the lack of error rate CIT effect could be caused by the relatively low overall error rates obtained in our study (approximately 3.6%). Alternatively, the lack of an error rate effect might be due to our use of scene stimuli; further study could clarify this. Finally, although our findings suggest that scene stimuli allow for a diagnostic RT-CIT, they do not directly tell us whether there is a difference between scene and object-based RT-CITs. Accordingly, in Experiment 2 this is investigated by presenting participants with a mock crime video (rather than relying on autobiographical memory) containing both scene and object stimuli allowing a direct comparison between the two.

Experiment 2: Scenes vs. objects in RT-CITs

Experiment 1 validated the use of scenes as stimuli in the RT-CIT. However, the scene images were autobiographic in nature and scene stimuli were not directly compared with object stimuli. To address these issues, in Experiment 2, participants watched a mock crime video before completing an RT-CIT that contained both object and scene images. The use of a mock crime video technique not only allowed a mixture of object and scene crime items to be tested but also allowed scene stimuli to be tested in a more realistic context, thereby increasing generalizability.

Method

    Participants

Initially the number of participants from Experiment 1 was simply increased for Experiment 2 to account for the reduced reliability as a result of halving the number of trials to account for the additional within subject condition i.e. object stimuli. Furthermore, due to use of an undergraduate participant pool for course credit, control over the exact number of participants was limited. Forty-four participants (38 women, aged 18-21, *Mean* = 18.8, *SD* = 0.8) were recruited from an Undergraduate Psychology course and took part in the 30-minute testing session in return for course credit[4].

---

[4] Given that there was no previous literature to indicate the size of a possible scene-object RT-CIT difference, the authors referred to a previous study which, during a post-hoc analysis, found no significant difference between object and scene stimuli in the physiological CIT (Norman et al., 2020). Analysis of that data revealed no significant interaction between Item (Crime vs Control) and Stimuli (Object vs Scene) and a medium within-subject effect size of, η_p^2 = .046. A post-hoc power analysis using G*Power (Faul et al., 2007), the effect size above, and α = 0.05 for a repeated measures ANOVA, suggested that 46 subjects would be sufficient for a power of 0.95.

Materials

Instead of the autobiographic scenes of the University campus used in Experiment 1, a three-minute, 1st person perspective mock crime video was shown to participants (Figure 5). In the following text-based description of the video *italics* indicate crime items. Participants (observing from the perspective of the thief) identified a locked bike outside the *Humanities building entrance*. The participants covered up a nearby CCTV camera using *shaving foam* and then used *bolt cutters* to break the bike lock. The culprit then met an accomplice in a multi-story *carpark* to hand over the bike for cash. Four matched control items were selected for each crime item (Figure 6). For Target items (highlighted), a false alibi was constructed: *"It wasn't me who committed that crime as I was with my friend at his home gardening all day. We only left his house to buy some garden clippers and weed killer from a nearby DIY store."*. Therefore, in total there were two object crime items and two scene crime items each with four controls and one target. The remaining experimental set up was the same as Experiment 1.
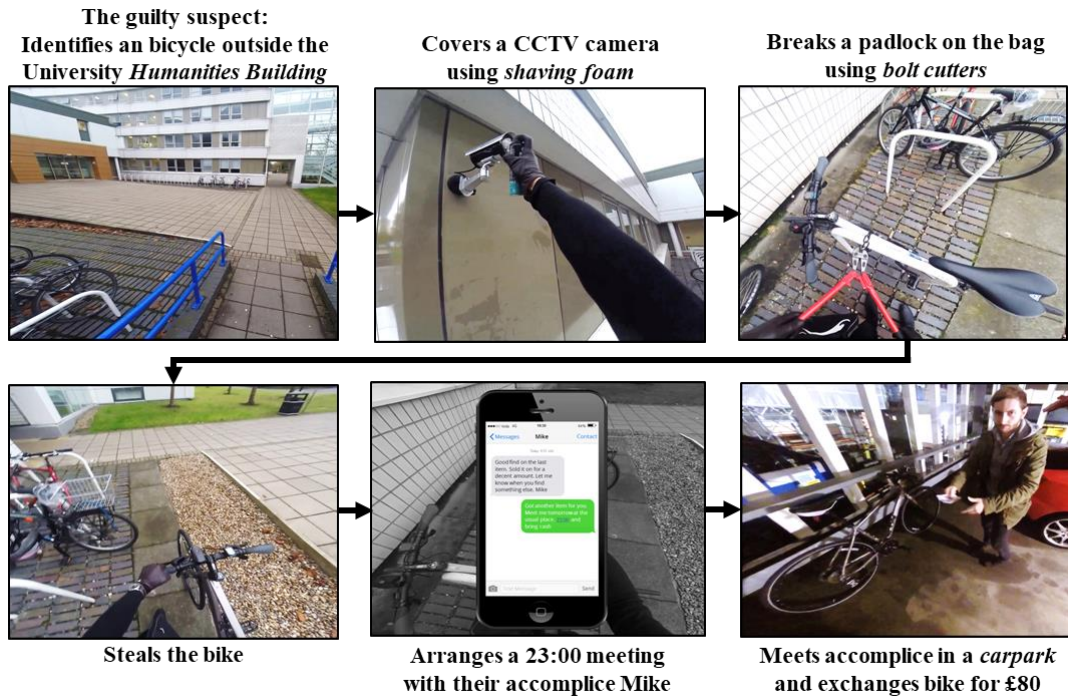
Figure 5. Key events in the 1ˢᵗ person-perspective mock crime video that guilty suspects view (crime items in italics).
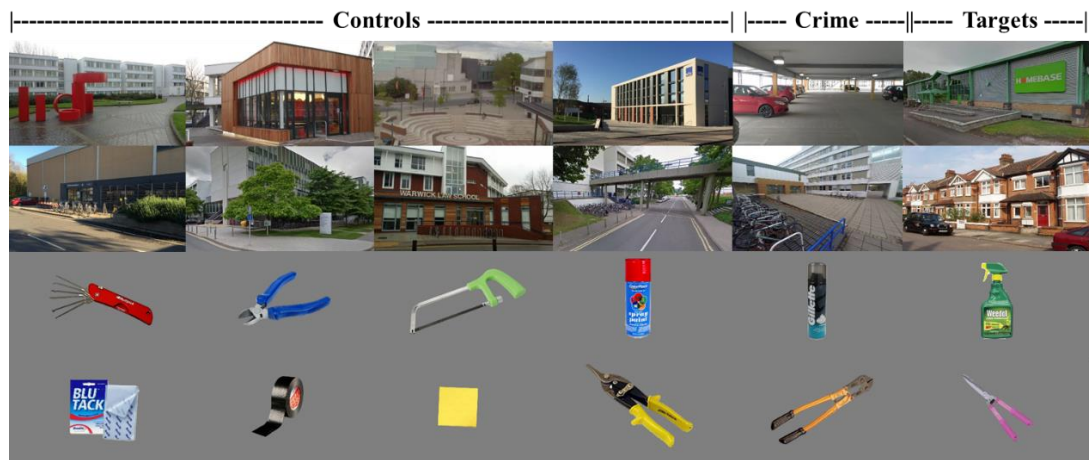

Figure 6. All scene and object images used in Experiment 2.

The CIT

The RT-CIT of Experiment 1 was used except that there were 360 images were presented in 15 blocks of 24 images. Each block consisted of four CITs, (two object CITs and two scene CITs), with each CIT containing a crime item, target, and four control images (Figure 6). Image order was randomised with two exceptions; i)

crime images were always preceded by control images and, ii) each block always

started with a control item.

Procedure

Participants were provided with an overview of the study procedure, given

the opportunity to ask questions and then provided consent and demographic

information. Participants were informed of their right to withdraw at any point

without penalty or reason. Participants were then told *"You'll now watch a three*

minute, 1st person perspective, mock crime video, of a thief, (you!) stealing a bicycle

from outside the humanities building on campus. It's really important you pay

attention throughout and really try to imagine yourself as the person whose

perspective you're seeing in the video. There will also be a memory test at the end."

Participants then put on headphones and watched the mock crime video. Following

the video, participants were asked to *"Now imagine you have been contacted by the*

local police station and have been informed that you are now a potential suspect in a

recent crime. They explain that during their investigation they would like to

administer a lie detection test to all potential suspects to help narrow down their

investigation. The lie detection test will use the crime images below which you

should now recognize from the video". Participants were then shown the four crime

images that would be used in the test to ensure sufficient encoding (note this would

not be appropriate in an applied setting however was done to ensure optimal

encoding).

Participants were then told, *"You have asked your good friend to be a false*

alibi for the time of the events and you have told the police this alibi story..."

Participants were then given their false alibi and four images related to it to

remember. Participants were then told: *"During the 'Lie Detection Test' you will be*

shown a series of items consisting of objects and scenes. Many of these items will be

unfamiliar to you except the ones relating to the crime you just 'committed' (in the

video) and the Alibi items you have just memorised. Each image will appear for

around 1 second with less than a second gap between them. Using the keyboard,

please respond to these images as fast as you can making as few errors as possible!

The question to bear in mind for every image is: "Do you recognise this item?".

Participants were instructed to press the LEFT KEY for "Yes" responses, that is,

False Alibi Images (targets), and the RIGHT KEY for "No", that is, Mock Crime

Images (crime items) and any other random Images (control items). The remaining

instructions were the same as in Experiment 1 until after the RT-CIT when

participants completed a short memory check. Finally, participants were debriefed.

Results

No participant data were removed from the analysis due to error rates above

50%. Including target items, trials that exceeded the response deadline (1.6%), were

faster than 200ms (0.36%) and incorrect responses (5.0%) were removed from the

analysis (Incorrect responses were used for the error analysis).

Manipulation Checks. When asked to identify the correct crime items, 40

out of the 44 (91%) participants correctly recalled all crime-relevant items with the

other four participants forgetting one item each.

Reaction Times. A 2 (Image Type: Object vs Scene) × 2 (Item Type: Control

vs. Crime) within-subjects ANOVA on the mean correct RTs revealed a significant

main effect of Image type, $F(1, 43) = 14.8$, $p < .001$, $MSE = 10031$, $\eta_p^2 = .256$, (RT-

CIT effect was $d = .784$ collapsed over Image type) and of Item type, $F(1, 43) = 26.5$,

$p < .001$, $MSE = 16885$, $\eta_p^2 = .381$. As shown in Figure 7, RTs were slower overall

with scene stimuli than with object stimuli and were slower on crime trials than on

control stimulus trials. The Image type × Item type interaction type was not significant, $F(1, 43) = .102$, $p = .751$, $MSE = 41.1$, $\eta_p^2 = .002$.
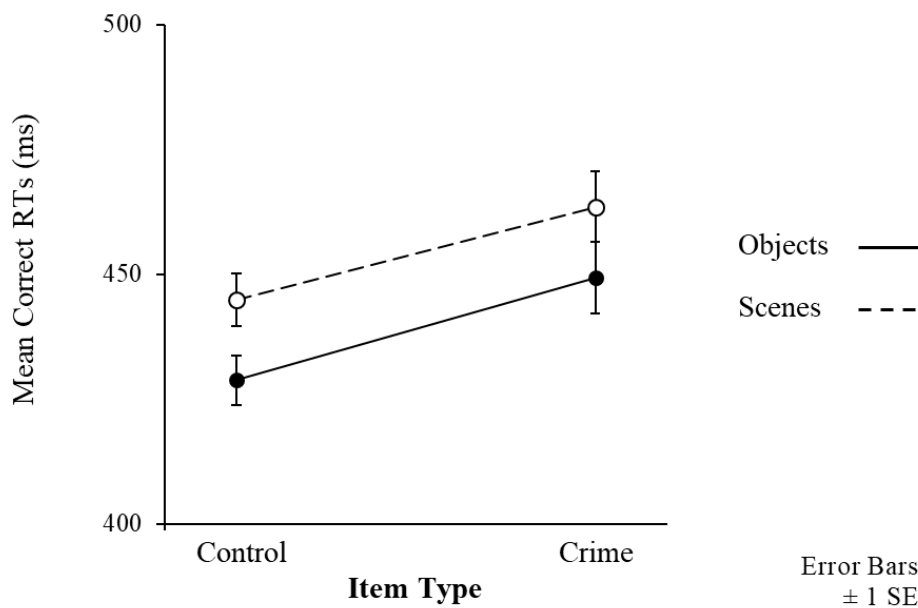

Figure 7. Experiment 2 - Mean correct RTs as a function of Item and Image Type.

Bayesian Analysis. A sensitivity analysis for this experiment was computed using G*Power which revealed a minimum detectable interaction effect between stimuli type (Objects, Scenes) and item type (Control, Crime) of $\eta^2 = 0.048$. This was larger than the actual interaction effect size found in this experiment, $\eta^2 = 0.001$ suggesting this experiment was underpowered. Therefore, a Bayesian Analysis was conducted. Where frequentist analysis reveals a non-significant difference, the Bayes factor BF_{01} is sometimes reported to quantify the degree to which the data support the null hypothesis (Wagenmakers et al, 2018). Therefore, to further assess the interaction between Image type and Item type reported above, the RT-CIT effect (the difference between the crime and control item) was compared for both scene and object stimuli using a Bayes t-test with JASP software (JASP Team, 2018). With a default Cauchy prior width of 0.7 this revealed a BF_{01} of 5.8, implying 'substantial

evidence for the null hypothesis' (Jefferys, 1961). This suggests that scenes and objects produce an equivalent CIT effect.

Errors. A 2 (Image type: object vs scene) × 2 (Item type: Control item vs. crime item) within-subjects ANOVA on mean error rates revealed a significant main effect of Item type, $F(1, 43) = 6.28$, $p = .016$, $MSE = 65.2$, $\eta_p^2 = .127$, but not for Image type, $F(1, 43) = .412$, $p = .524$, $MSE = 11.3$, $\eta_p^2 = .009$. As shown in Figure 8, error rates were higher for crime trials than for control stimulus trials but there was no difference between scene and object stimuli. The Image type × Item type interaction was not significant, $F(1, 43) = .166$, $p = .686$, $MSE = 8.39$, $\eta_p^2 = .004$. A Bayes t-test calculated using the crime-control item differences for objects and scenes revealed a BF_{01} of 5.7, implying 'substantial evidence for the null hypothesis'.
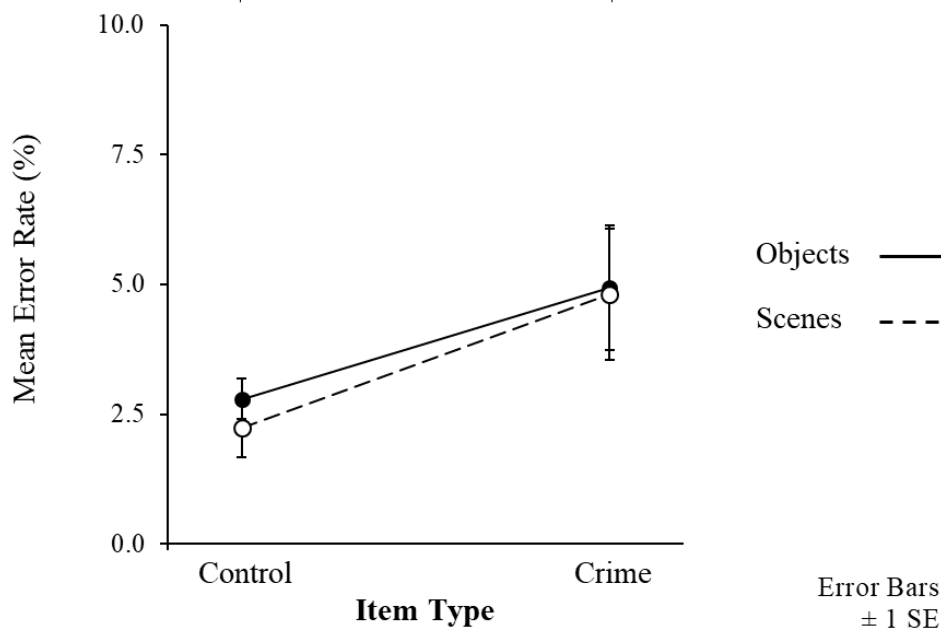

Figure 8. Experiment 2 - Mean % errors as a function of Item and Image Type.

Signal Detection Analysis

Using the same procedure as Experiment 1, a signal detection analysis was conducted for all participants using their RT responses for both scene, $AUC = .696$ and guilty-innocent effect size $d = .728$, and object stimuli, $AUC = .709$ and $d = .791$

(Figure 9). Note that collapsed over Image type the $AUC = .746$ and guilty-innocent
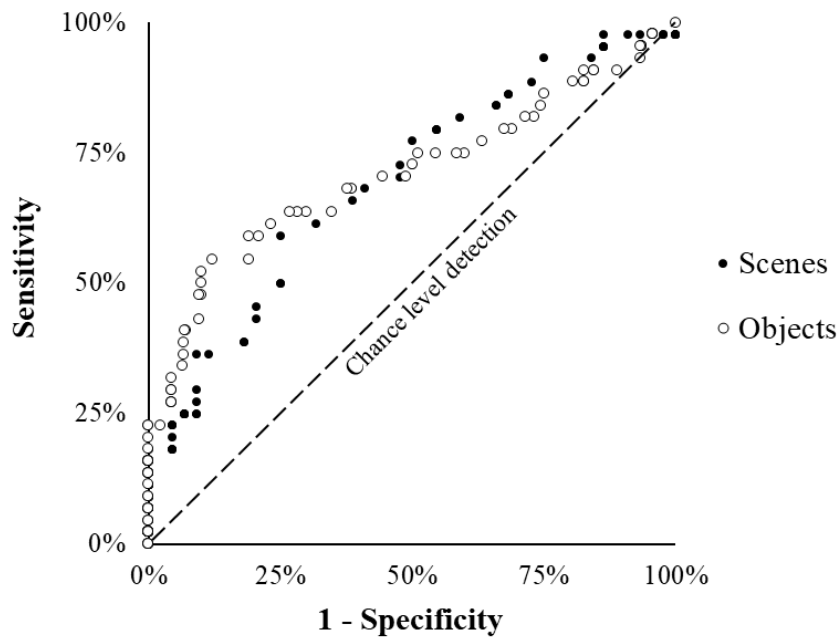
effect size $d = 1.01$.

Figure 9. Experiment 2 - Signal detection curve (ROC) showing the detection

sensitivity and specificity between guilty and simulated innocent participants

Discussion

In Experiment 2, both scene and object stimuli were used from a mock crime

video thereby allowing a direct contrast between stimulus type and removing the

reliance on autobiographic memory as used in Experiment 1. The main finding was

that, once again, scene stimuli successfully elicited a large within-subject CIT effect.

Moreover, a Bayesian analysis indicated that scenes and objects were equivalent in

terms of producing crime-control item RT differences. Signal detection analysis

based on the RT data revealed a lower AUC in Experiment 2 compared to

Experiment 1, likely due to both the smaller number of crime items used in

Experiment 2 (four instead of five) (Meijer et al., 2014) and the use of the mock

crime stimuli rather than autobiographic stimuli. A secondary finding was that

participants responded more slowly to scenes than to objects. This suggests that

scenes may be more cognitively demanding to process, perhaps because they are made up of multiple objects and have a generally higher complexity than pictures of isolated, single objects. In any case, the finding of an equivalent crime-control difference for pictures of objects and scenes suggests that this overall difference between scenes and objects does not impair the effectiveness of the test.

In contrast to Experiment 1, the RT-CIT effect was found for both RTs and error rates. This may be due to the higher number of errors in Experiment 2 compared to Experiment 1 (3.6% and 5.0% respectively, $p = .042$). Thus, the overall difficulty of the task may determine whether or not an RT-CIT effect is expressed in error rates as well as in RT measures. Either way, these findings indicate that although errors can be useful in detecting 'guilty' participants, they may be a less reliable measure than RT-based data. In Experiment 3 we test the robustness of the RT-CIT to countermeasures.

This experiment sought to determine whether there were any differences in the RT-CIT effect between object and scene stimuli under optimal conditions. Therefore, to reduce the chance of participants not adequately encoding the mock crime items, and at the cost of ecological validity, participants were briefly reminded of the key crime information after the mock crime video. Note however, that this would not be possible in a real CIT as it would compromise the results for innocent, unknowledgeable participants. This limitation may have resulted in inflated detection rates as participants were more likely to have had good memory for the crime details.

Participants recalled 91% of crime items from the first-person perspective mock crime videos used in this experiment. Although mock crime videos are a simple and effective medium for having participants encode crime details, they are less ecologically valid compared to committing a real-world mock crime and

unlikely to mimic real-world scene encoding. It is possible that any recognition differences between objects and scenes only become apparent when they are encoding in the real-world. Nevertheless, this was indirectly examined in a previous study where participants encoded both objects and scenes in the real-world before undergoing a physiological-based CIT – this study also revealed no significant difference in the CIT effect (Norman et al., 2020). However, in both the current study and the one by Norman and colleagues, no delay between encoding and testing were introduced. This clearly raises questions regarding the ecological validity of this work and therefore future work which manipulates delay between encoding and testing would be advisable.

Finally, although this experiment did not reveal any significant difference in the RT-CIT effect between object and scene stimuli, it is worth considering potential confounding variables. Scenes and objects can differ in saliency and saliency is known to modulate the CIT. Previous work (e.g., Kleinberg and Verschuere, 2015, see also klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2017; Jokinen, Santtila, Ravaja & Puttonen, 2006) has shown that items with higher personal salience (e.g., country of origin or birthday) produce a larger RT-CIT effect than less personally salient stimuli (e.g., favourite colour or animal). Therefore, it is conceivable that scene stimuli may in fact result in a differing RT-CIT effect to objects but that this effect is not seen in this study due to differences in item-specific saliency which was not controlled for in this work. Further work controlling for this possible modulating factor would be beneficial.

Experiment 3: The effect of countermeasures on the RT-CIT

Experiments 1 and 2 showed that, like object stimuli, scene stimuli can be used to generate a CIT effect. However, in both experiments participants were not instructed to use any form of countermeasure strategy (a strategy to try to fool the test and elicit a false negative result). One would expect guilty suspects in the real world to attempt to use some form of countermeasure strategy to avoid detection and, arguably, this could be simple to perform in an RT deception test (Gronau, Ben-Shakhar, & Cohen, 2005). Steps to mitigate against countermeasure strategies are therefore frequently used in standard RT-CIT procedures. For example, a response deadline of 800ms is used to prevent participants from intentionally delaying responses to control items and therefore negating the CIT-effect. Furthermore, target items that require a different response ("Yes") to crime and control items are used to ensure that participants are engaging with each stimulus as presented. If they were not, then this would yield a high error rate suggesting that the participant is either not paying attention or attempting some form of countermeasure.

A handful of studies have consider the effects of countermeasures on deception tasks similar to the RT-CIT. Ganis and colleagues (2011) gave twenty-six participants an fMRI-based RT-CIT, using word stimuli, and instructed them to covertly move the left index, middle finger, and toe to three out of the four control items during the CIT. The idea was that this might increase the RTs for control items and thus reduce the difference between the crime and control stimuli neutralizing the RT-CIT effect. The countermeasures were effective in increasing RTs to control items compared with crime items resulting in a negative RT-CIT effect (MD = -313ms, SE = 74). Compared to the control group (MD = 60ms, SE = 54) this suggested a large

countermeasure effect size, $d = 1.73$ for the RT differences with a similar result found for the f-MRI measure, $\eta_p^2 = 0.94$.

In another study, Huntjens, Verschere, and McNally (2012) presented a word-based autobiographic RT-CIT to a control group, countermeasure-instructed ("*do not respond any faster or slower to the crime items*") group, and a dissociative identity disorder patient group ($n = 27$, 23 and 11 respectively). They found no difference between the control and countermeasure group, who still showed a crime-control item difference ($d = .48$). However, in Huntjens and colleagues' (2012) study, participants were simply instructed not to respond any faster or slower to crime items which, without further instruction is unlikely to have been a sufficient countermeasure strategy.

An ERP-based CIT, using word stimuli (Mertens & Allen, 2008) tested seventy-nine participants across five conditions: control, innocent, and three countermeasure instructed groups. Countermeasure instructed groups were given directions to: "*think about being slapped by a bully*" (Group 1), "*apply pressure to the toe*" (Group 2) and "*think about being slapped by a bully after the first distracter, tightening of the sphincter muscle after the second, distracter, applying pressure to toes after the third distracter, and not responding at all to the fourth distracter*" (Group 3). Although this study found countermeasure differences for EPR amplitude and latency, the behavioral RT data only showed a countermeasure effect for Group 3. Note however that all other groups still showed a significant RT-CIT effect.

In another ERP-based CIT study, with a RT measure and also using word stimuli (Rosenfeld, Soskins, Bosh & Ryan, 2004), thirty-three participants were asked to perform various covert tasks when control items were presented. The covert task were: i) pressing the left forefinger, ii) pressing the left middle finger, iii)

wiggling the big toe in the left shoe, iv) wiggling the big toe in the right shoe, and v) imagining the experimenter slapping the participant. Compared to a control group where 91% of participants were correctly classified based on their RTs, only 45% of the countermeasure instructed group were detected which demonstrated a significant countermeasure effect. Rosenfeld and colleagues found a significant difference between RT differences (crime minus control) for their guilty (no countermeasure) and countermeasure group, $t(10) = 2.19$, $p < .05$, with a large countermeasure effect size $d = .932$.

Finally, using an Autobiographical Implicit Association Test (aIAT), one study found that 39-78% of its eighteen guilty participants were able to remain undetected by being informed on how the aIAT works and instructing participants to slow down in the confession-true task (Verschuere, Prati & Houwer, 2009). Over the three experiments, participants given countermeasure instructions could significantly lower their test score to appear innocent (average countermeasure $d = .98$). The studies described suggest that countermeasures can be effective in reducing detection, however few studies have tested the impact of countermeasures in RT-CIT tasks that do not also use ERPs or fMRI (see Suchotzki et al., 2017, for a small meta-analysis of these). Furthermore, none of the above studies, or any others to our knowledge, have tested the susceptibility of scene stimuli to countermeasures.

Accordingly, in Experiment 3, a physical countermeasure strategy was tested in a scene-based RT-CIT. The most obvious approach to reduce the RT-CIT effect, i.e. the difference in RT's between crime and control items, is to slow responses to the irrelevant items. As described previously there are difference approaches to this which can be broadly categorized as either mental or physical countermeasure strategies. In this experiment we chose a simple and easy to perform physical

countermeasure that requires little practice - press upon or wiggle a toe to every control item. Indeed, this had been used in previous research (e.g. Rosenfeld et al, 2004; Mertens & Allen, 2008) which showed that RTs slow with increased motor response complexity (Henry & Rogers, 1960; Anson, 1982; Klapp, 2010). By preforming an additional task i.e. pressing a toe, for control items only, RT's should increase thereby reducing the RT-CIT effect.

Method

Participants

Guided by the literature described above, the average RT countermeasure effect (the difference in RT-CIT effect between control and countermeasure groups), when found, was large $d = 1.03$. Assuming a large countermeasure effect, a power analysis using G*Power, with an effect size of $d = 0.8$, and $\alpha = 0.05$ for a single group, suggested that 42 subjects per group would be sufficient for a power of 0.95. Ninety-eight participants (58 women and 4 undisclosed, aged between 18 - 42, *Mean* $= 22.7$, *SD* $= 5.1$), 48 in the control group, were recruited through a University online participant panel at the authors' institution and took part in the 30-minute testing session in return for £3 payment. Participants were assigned to each condition based on the experiment session they signed up to (there were four sessions with approximately twenty places available.) Participants, without knowledge of the different experimental conditions, chose which one to attend.

Procedure

The materials, RT-CIT, and procedure for the control group were identical to those of Experiment 1. For the countermeasure group, participants were told how the RT-CIT worked and instructed to "*perform a toe-tap or a toe wriggle when responding to unfamiliar images*" to try and fool the test. To ensure that participants

were indeed carrying out the instructed countermeasure, the experimenter visually observed participants during the experiment.

Results

The data from two participants were removed from all analyses due to errors rates greater than 50% (58% and 89% from the control and countermeasure condition respectively) and one participant from the countermeasure group due to a technical error with the program. Of the remaining participants, trials (including target items) that exceeded the response deadline (2.3%), were faster than 200ms (0.73%) and incorrect (7.38%) were removed from the analysis (incorrect responses were used for the error analysis).

Reaction Times. Mean correct RTs were analyzed using a 2 (Item Type: crime item vs. control item) × 2 (Condition: control vs. countermeasure) mixed-ANOVA with Item Type as the within-subjects factor and Condition as the between-subjects factor. This revealed a significant main effect of Item Type, $F(1, 93) = 112.2$, $p < .001$, $MSE = 34623$, $\eta_p^2 = .547$, (RT-CIT effect was $d = 1.1$ collapsed over condition) with RTs on crime item trials longer than those on control item trials (Figure 10) However, neither the main effect of Condition, $F(1, 93) = .15$, $p = .696$, $MSE = 2178$, $\eta_p^2 = .004$, nor the Condition × Item Type interaction, $F(1, 93) = .02$, $p = .882$, $MSE = 3.14$, $\eta_p^2 = .001$, approached significance.

A sensitivity analysis for this experiment was computed using G*Power which revealed a minimum detectable interaction effect between countermeasure group (Control, Countermeasure) and item type (Control, Crime) of $\eta^2 = 0.033$. This was larger than the actual interaction effect found in this experiment, $\eta^2 = 0.001$ suggesting this experiment was underpowered. Therefore, a Bayes t-test was computed to evaluate the difference between the countermeasure and control

condition using the crime item/control item RT difference. This revealed a BF_{01} value of 4.6 implying 'substantial' evidence for the null hypothesis suggesting that the countermeasure was ineffective.

Error Analysis. A 2 (Condition: Control vs. Countermeasure) × 2 (Item Type: Control item vs. Crime item) repeated-measures ANOVA on the mean error rates revealed no main effect of Item Type, $F(1, 93) = .061$, $p = .805$, $MSE = 1.52$, $\eta_p^2 = .001$, or Condition, $F(1, 93) = 1.35$, $p = .101$, $MSE = 430$, $\eta_p^2 = .029$. Mean error rates for all trials were low ($M = 5.76$) with no difference between the Control and Countermeasure groups. The Condition × Item Type interaction was not significant, $F(1, 93) = 1.35$, $p = .248$, $MSE = 33.4$, $\eta_p^2 = .014$. A Bayesian t-test calculated using the crime item-control item differences for both Conditions revealed a BF_{01} of 2.6, implying 'anecdotal evidence for the null hypothesis'.
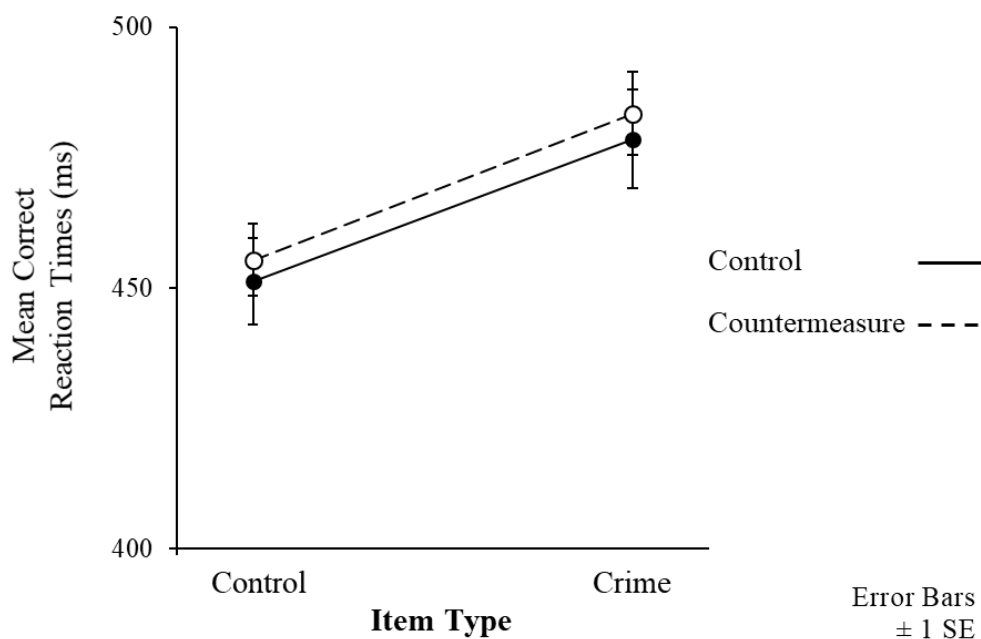

Figure 10. Experiment 3 - Mean correct RTs as a function of Item Type and Countermeasure Condition.

Signal Detection Analysis

Using the same procedure as in Experiments 1 and 2, a signal detection analysis was conducted for both the countermeasure, $AUC = .808$ and guilty-innocent effect size $d = 1.14$, and control group, $AUC = .878$ and $d = 1.52$, using a simulated innocent group (Figure 11). Note that collapsed over condition, the $AUC = .843$ and guilty-innocent effect size $d = 1.33$.
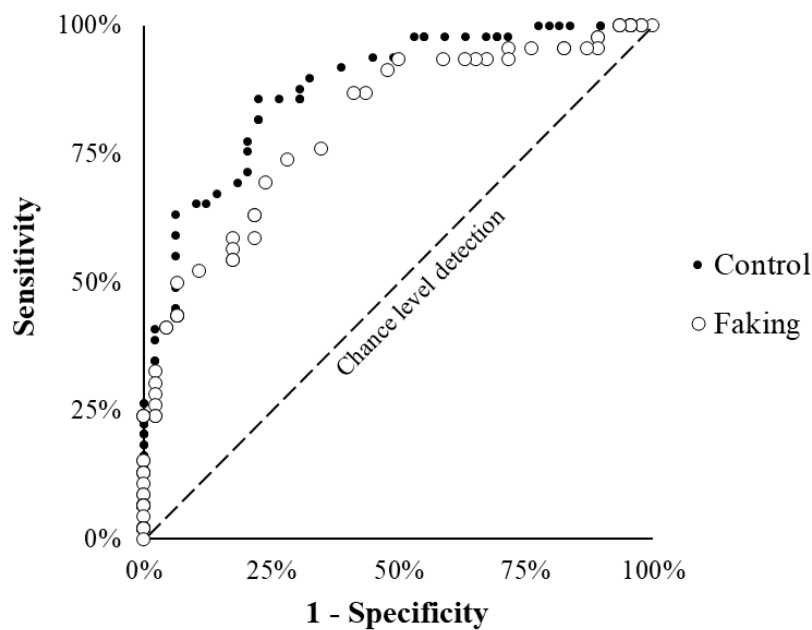

Figure 11. Experiment 3 - Signal detection curve (ROC) showing the detection sensitivity and specificity between guilty and simulated innocent participants

Discussion

As in Experiments 1 and 2, scene stimuli produced a robust RT-CIT effect. However, of most interest, Experiment 3 tested the susceptibility of the scene-based RT-CIT to a physical countermeasure strategy; specifically, participants were asked to "*perform a toe-tap or a toe wriggle when responding to unfamiliar images*". The logic behind this type of countermeasure is that performing an additional task on control stimulus trials might increase the RTs on those trials thus reducing the RT difference between crime and control stimuli; hence reducing the RT-CIT effect.

There is currently little work specifically investigating the effects of countermeasures on the RT-CIT (Suchotzki et al., 2017). The findings in this experiment suggest that there was no difference between the control and countermeasure group in terms of the crime-control RT difference - this lack of difference was supported by a Bayesian analysis.

Our findings contrast with those from ERP (Mertens & Allen, 2008; Rosenfeld et al., 2004), and fMRI (Ganis et al., 2011) and aIAT (Verschuere, Prati & Houwer, 2009) studies in which countermeasures were influential. This difference may be explained by the fact that the methodologies for an ERP, aIAT, fMRI-based CIT are quite different in terms of the stimulus duration time, interstimulus interval and the use of additional physiological measurements which requires the participant having to remain stationary throughout the experiment.

General Discussion

Determining whether a suspect recognizes crime-related information can be valuable and may be achieved using the RT-CIT. A substantial body of research has already established that the RT-CIT can be an effective means of revealing a suspect's knowledge when word stimuli are used. A smaller number of studies have also established that the RT-CIT works with images of discrete objects (Visu-Petra et al., 2016; Varga et al., 2015; Suchotzki et al., 2015). However, as well as objects, crime-related information can also take the form of scenes related to criminal activity. Linking a suspect with a crime scene will extend the range of situations in which the RT-CIT can be successfully applied. Importantly, knowledge of such scenes could not be easily tested by the presentation of a single word (or a limited number of words), nor by presenting images of single discrete objects. Although we know that the RT-CIT can determine whether a suspect recognizes one or more crime-related objects, the present study is the first to apply the test to the recognition of crime-related scenes.

At first glance, one might expect that scenes would work in an RT-CIT just as well as images of single objects. Indeed, this appears to be the case when using the physiological-base CIT (Norman et al., 2020). However, as detailed in the Introduction, due to differences in the way in which scenes and objects are encoded and processed we might expect the RT-CIT effect to differ between them. For example; scenes require encoding into memory over several seconds (Melcher, 2006), scene recognition is possible without complete identification (Cleary & Reyes, 2009), scene information is processed rapidly requiring fewer attentional resources than objects (Munneke et al, 2013), and the PPA brain region responds only to scenes and not to objects (Oliva & Torralba, 2006). Clearly, there are reasons

to question whether the RT-CIT will be effective (or at least as effective) with scene-based stimuli than with object-based stimuli. Furthermore, scenes typically contain many objects and focusing on a single object, either during the 'crime' or at the test phase, might reduce the extent to which the RT-CIT can detect differences between the crime and control items if those objects are different. Similarly, limits in attentional capacity might reduce what is remembered from a scene at both encoding and retrieval phase. In addition, scenes may well contain more simple, global features (Oliva & Torralba, 2006) which might interfere with the processing of the deeper meaning of the scene. This could again have an effect at the encoding stage if participants simply encode and remember the gist of a scene.

Nonetheless, despite these concerns, a robust RT-CIT effect was obtained across a variety of scene-based stimuli in the three experiments. Specifically, responses to crime items were slower than to control items when either autobiographic or more recent memory was tested, and the difference between crime and control responses was equivalent to those obtained with object-based stimuli. Overall this study suggests that RT-CIT effect sizes ($d = .784$ to 1.48) for scene stimuli were similar to those obtained in previous RT-CIT studies which used pictures of objects, ($d = 1.05$ to 1.24, Visu-Petra et al., 2016; Varga et al., 2015; Suchotzki et al., 2015) and words ($d = 1.05$ [.93 - 1.17, $CI^{95\%}$], Suchotzki et al., 2017). Consequently, this meant that RT-CIT diagnosticity ($AUC = .746 - 919$) was also similar to those reported for the RT-CIT a previous meta-analysis ($AUC = .82$ [.77 - .87, $CI^{95\%}$], Meijer et al., 2016).

These findings also suggest that the scene-based RT-CIT may be robust to at least one simple-to-implement countermeasure – a covert manual movement when responding to control stimuli. In the current study, it appears that making an

additional physical movement did not interfere with the basic difference between RTs to crime and control items. It is, of course, possible that participants simply did not apply the countermeasure and, although this was monitored by an experimenter, it would have been difficult to confirm thereby. Clearly if some participants in the countermeasure condition did not use the instructed countermeasure, then this would be a significant limitation with this experiment. However, the finding of a trend for error rates to be higher in the countermeasure condition than in the control condition provides some, albeit relatively weak, evidence that participants were experiencing a higher cognitive load, consistent with them attempting to implement the countermeasure.

The lack of an effect of the countermeasure is inconsistent with others (Mertens & Allen, 2008; Ganis et al., 2011; Rosenfeld et al., 2004; Verschuere et al., 2009). However, as noted earlier, there appears to be large methodological differences between the studies that have found countermeasures to be effective and this current study. Determining which countermeasures are effective and under what conditions will be a useful goal for future research. In conclusion, the findings from this study suggest that the RT-CIT can be successfully applied to the recognition of scenes and produces an equivalent effect size to object-based tests.

Open Science Statement:

All data are available in the supplementary materials, materials will be provided on request and none of the experiments were preregistered.

References

Anson, G.J. (1982). Memory drum theory: Alternative tests and explanations for the complexity effects on simple reaction time. Journal of Motor Behavior, 14(3), 228-246.

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. Trends in cognitive sciences, 17(5), 210-219.

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

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. Journal of Experimental Psychology: Applied, 9(4), 261.

Chechik, G., Shalit, U., Sharma, V., & Bengio, S. (2009). An online algorithm for large scale image similarity learning. In Advances in Neural Information Processing Systems (pp. 306-314).

Cleary, A. M., & Reyes, N. L. (2009). Scene recognition without identification. Acta Psychologica, 131(1), 53-62.

Debey, E., Ridderinkhof, R. K., De Houwer, J., De Schryver, M., & Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: Delta plots reveal the role of response inhibition in lying. Consciousness and Cognition, 37, 148–159.

Eom, J. S., Sohn, S., Park, K., Eum, Y. J., & Sohn, J. H. (2016). Effects of Varying
Numbers of Crime items on RT-based CIT Accuracy. International Journal of
Multimedia and Ubiquitous Engineering, 11(2), 229-238.

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, 175–191.

Fisher, B. A. J. (2004). Techniques of crime scene investigation. (7th ed.) New York,
NY: CRC Press

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on
object identification. Cognitive Brain Research, 16(2), 123-144.

Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011).
Lying in the scanner: covert countermeasures disrupt deception detection by
functional magnetic resonance imaging. Neuroimage, 55(1), 312-319.

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

Gronau, N., Ben-Shakhar, G., & Cohen, A. (2005). Behavioral and physiological
measures in the detection of concealed information. *Journal of Applied
Psychology*, *90*(1), 147.

Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated
movements and a "memory drum" theory of neuromotor reaction. Research
Quarterly. American Association for Health, Physical Education and
Recreation, 31(3), 448-458.

Horváth, A. (2015, December). Object recognition based on Google's reverse image
search and image similarity. In Seventh International Conference on Graphic

and Image Processing (pp. 98170Q-98170Q). International Society for Optics
and Photonics.

Hu, X., Evans, A., Wu, H., Lee, K., & Fu, G. (2013). An interfering dot-crime item
task facilitates the detection of mock crime memory in a reaction time (RT)-
based concealed information test. Acta psychologica, 142(2), 278-285.

Huntjens, R. J., Verschuere, B., & McNally, R. J. (2012). Inter-identity
autobiographical amnesia in patients with dissociative identity disorder. PloS
one, 7(7), e40580.

JASP Team (2019). JASP (Version 0.8.5.1) [Computer software]

Jeffreys, H. (1961). Theory of probability, (3rd ed.) Oxford, UK: Oxford University
Press.

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.

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., 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.

Klapp, S. T. (2010). Comments on the classic Henry and Rogers (1960) paper on its
50th anniversary: resolving the issue of simple versus choice reaction time.
Research Quarterly for Exercise and Sport, 81(1), 108-112.

Kent, C, Lamberts, K & Patton, R, 2018, 'Cue quality and criterion setting in
recognition memory'. Memory and Cognition, vol 46., pp. 757-769

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

Kleinberg, B., & Verschuere, B. (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. Journal of Applied Research in Memory and Cognition, 5(1), 43-51.

Meijer, E. H., Selle, N. K., 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.

Meijer, E. H., Smulders, F. T., Johnston, J. E., & Merckelbach, H. L. (2007). Combining skin conductance and forced choice in the detection of concealed information. Psychophysiology, 44(5), 814-822.

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.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. Journal of vision, 6(1), 2-2.

Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, *45*(2), 286-298.

Munneke, J., Brentari, V., & Peelen, M. V. (2013). The influence of scene context on object recognition is independent of attentional focus. Frontiers in psychology, 4.

Noordraven, E., & Verschuere, B. (2013). Predicting the sensitivity of the reaction time-based Concealed Information Test. Applied Cognitive Psychology, 27(3), 328-335.

Norman, D. G., Wade, K. A., Williams, M. A., & Watson, D. G. (2020). Caught Virtually Lying—Crime Scenes in Virtual Reality Help to Expose Suspects' Concealed Recognition. Journal of Applied Research in Memory and Cognition.

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.

Peterson, J., Sommers, I., Baskin, D., & Johnson, D. (2010). The role and impact of forensic evidence in the criminal justice process. National Institute of Justice, 1-151.

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300‐based tests of detection of concealed information. Psychophysiology, 41(2), 205-219.

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. Applied psychophysiology and biofeedback, 34(3), 177-187.

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. Applied Cognitive Psychology, 22(4), 475-490.

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge". Journal of Applied Psychology, 85(1), 30

Standing, L. (1973). Learning 10,000 pictures. Quarterly Journal of Experimental Psychology, 25, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for picture: Single-trial learning of 2500 visual stimuli. Psychonomic Science, 19, 73-74.

Suchotzki, K., Verschuere, B., Peth, J., Crombez, G., & Gamer, M. (2015). Manipulating item proportion and deception reveals crucial dissociation between behavioral, autonomic, and neural indices of concealed information. Human brain mapping, 36(2), 427-439.

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. Psychological Bulletin, 143(4), 428.

Varga, M., Visu-Petra, G., Miclea, M., & Visu-Petra, L. (2015). The "Good Cop, Bad Cop" Effect in the RT-Based Concealed Information Test: Exploring the Effect of Emotional Expressions Displayed by a Virtual Investigator. PloS one, 10(2), e0116087.

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

Verschuere, B., Crombez, G., Degrootte, T., & Rosseel, Y. (2010). Detecting concealed information with reaction times: Validity and comparison with the polygraph. Applied Cognitive Psychology, 24(7), 991-1002.

Verschuere, B., Kleinberg, B., & Theocharidou, K. (2015). RT-based memory

detection: Item saliency effects in the single-crime item and the multiple-

crime item protocol. Journal of Applied Research in Memory and Cognition,

4(1), 59-65.

Verschuere, B., Suchotzki, K., & Debey, E. (2014). Detecting deception through

reaction times. Detecting deception: current challenges and cognitive

approaches, 269-291.

Verschuere, B., Prati, V., & Houwer, J. D. (2009). Cheating the lie detector: Faking in

the autobiographical Implicit Association Test. Psychological Science, 20(4),

410-413.

Visu-Petra, G., Miclea, M., & Visu-Petra, L. (2012). Reaction Time-based Detection

of Concealed Information in Relation to Individual Differences in Executive

Functioning. Applied Cognitive Psychology, 26(3), 342-351.

Visu-Petra, G., Miclea, M., Buş, I., & Visu-Petra, L. (2014). Detecting concealed

information: The role of individual differences in executive functions and

social desirability. Psychology, Crime & Law, 20(1), 20-36.

Visu-Petra, G., Varga, M., Miclea, M., & Visu-Petra, L. (2013). When interference

helps: increasing executive load to facilitate deception detection in the

concealed information test. Frontiers in psychology, 4, 146.

Visu-Petra, L., Jurje, O., Ciornei, O., & Visu-Petra, G. (2016). Can you keep a

secret? Introducing the RT-based Concealed Information Test to children.

Psychology, Crime & Law, 22(3), 276-301.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... &

Matzke, D. (2018). Bayesian inference for psychology. Part I: Theoretical

advantages and practical ramifications. Psychonomic bulletin & review, 25(1), 35-57.

Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(3), 509.

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