Reading the criminal mind

Exploring novel methods of memory detection

Van der Cruyssen, I.A.

Publication date
2023

2

Modality effects in memory detection

This chapter is published as:
Van der Cruyssen, I., Regnath, F., Ben-Shakhar, G., Pertzov, Y., & Verschuere, B. (2021). Is a Picture Worth a Thousand Words? Congruency Between Encoding and Testing Improves Detection of Concealed Memories. *Journal of Applied Research in Memory and Cognition, 10*(4), 667-676. doi:10.1016/j.jarmac.2021.03.001

Abstract

The current study addressed modality effects in a web-based Concealed Information Test (CIT) by asking participants to encode, and later conceal, crime-related details. Items were encoded and tested verbally or pictorially. A pilot (N = 73) and a preregistered study (N = 158) showed a robust interaction between encoding and testing modality: Items that were encoded and tested in the same modality were associated with better detection. Moreover, recognition of verbally encoded items could not be detected in a pictorial test. Our findings support the existence of a modality-congruency effect when subjects try to conceal their knowledge. In applied scenarios, the modality of test items should be matched to the modality in which crime-related details were encoded. Furthermore, a pictorial CIT might protect informed innocents if leakage happened verbally.

Pictures are better remembered and more easily recognized than their verbal counterparts. This so-called picture superiority effect is well-established and has been repeatedly demonstrated in the memory literature (e.g., Ensor et al., 2019; Kirkpatrick, 1894; Madigan, 2014). This phenomenon has been demonstrated in both recall and recognition tests (Bevan & Steger, 1971; Shepard, 1967) and is very robust. For example, in a study by Shepard (1967), participants studied over 500 words or pictures and were respectively 88% versus 97% accurate in a subsequent recognition test. Over the years, the picture superiority effect has received considerable attention from researchers, and the phenomenon has been applied in a range of areas requiring easy and accurate recall of information such as instructional design, advertising, and marketing communications (Childers & Houston, 1984; Sansgiry et al., 1997). However, despite the many demonstrations confirming the picture superiority effect, its underlying mechanism is still a subject of discussion. Theories explaining pictures' mnemonic advantage can be roughly divided into those attributing the effect to a perceptual processing advantage of pictures and those attributing it to a conceptual processing advantage of pictures (for a review, see Stenberg, 2006).

According to perceptual theories, the picture superiority effect can be attributed to a richer perceptual encoding of pictures (Stenberg, 2006). While words are constrained to letters, phonemes, and orthographic conventions, pictures have no such constraints or limits (Nelson et al., 1979). As a result, there is more representational variability from picture to picture than from word to word, leading to enhanced distinctiveness of pictures and easier recall or recognition. Paivio's (1971) dual-code theory, one of the earliest and most influential accounts of this effect, distinguishes between verbal and nonverbal (e.g., visual, auditory, haptic) memory pathways. According to this theory, words tend to be processed only via the verbal pathway, whereas pictures can be encoded through both routes. That is, words only evoke a verbal code, but pictures can activate both the image and its verbal label (e.g., “dog”) and are therefore more easily retrieved and recalled.

Conceptual theories, on the other hand, claim that the processing of pictures involves greater semantic elaboration than the processing of words (Stenberg, 2006). Accordingly, the transfer-appropriate processing theory assumes that cognitive operations performed during encoding are reinstated during retrieval (Rugg et al., 2015). Since a typical memory test such as a recall or recognition test requires deliberate recollection of prior experience, they have a mostly conceptual nature, which invokes subjects’ semantic knowledge. As picture encoding also invokes greater semantic elaboration, pictures have an advantage at test as compared to words, resulting in an overall better performance for pictures in most recall and recognition tests (Weldon et al., 1989).

The effect of presentation modality is particularly interesting in the context of memory detection tests. For example, in a situation where one tries to detect whether a suspect has implicating crime knowledge, it is important to know how the test items should ideally be presented (i.e., verbally or pictorially) and whether this is moderated by the way the perpetrator originally encoded the crime items. This question is especially important because in memory detection, people are deliberately suppressing or hiding their memories instead of trying to retrieve them (as in classical research about the picture superiority effect). Although the picture superiority effect is a reliable and reproducible phenomenon, it has been shown to be constrained by some limiting conditions. The test circumstances, for example, have been shown capable of abolishing or even reversing the picture superiority effect (Weldon & Roediger, 1987).

The Concealed Information Test (CIT) is a promising technique used for memory detection when subjects try to conceal their knowledge. Concealed knowledge is detected by comparing the examinees' responses to a crime-relevant item (e.g., the murder weapon) to multiple, equally plausible control options (e.g., “A revolver? A rope? A lead pipe? A candlestick?”). A guilty suspect is assumed to have specific crime-related knowledge and is therefore expected to respond differently to the crime-related item than to the control items. An innocent examinee, on the other hand, will not recognize the crime-related item and should therefore respond similarly to all presented items. After decades of research, it is well established that the CIT provides a valid tool for the detection of "knowledgeable" individuals (e.g., Meijer et al., 2014; Suchotzki et al., 2017) and is nowadays used in criminal investigations in Japan (Osugi, 2011).

A few studies have explored modality effects in the CIT (Ambach et al., 2010; Ben-Shakhar et al., 1996; Rosenfeld et al., 2015; Seymour & Kerlin, 2008). Both Ambach et al. (2010) and Seymour and Kerlin (2008) manipulated modality at test (but not at encoding) using autonomic and brain electrical measures and Reaction Times (RT) respectively. They both found no differences between verbal versus pictorial presentation. However, since these studies did not manipulate modality at encoding, they could not examine possible effects of encoding modality and congruency between encoding and testing modalities. Rosenfeld and colleagues (2015) applied the CIT with brain electrical measures and recommended that CIT examiners should always use pictorial presentations at testing, because they found that the use of pictorial presentation was always more effective, regardless of the encoding modality. However, these suggestions were based on data from only 11 participants. Lastly, Ben-Shakhar et al. (1996) used a modified version of the CIT based on electrodermal measures and found a complete generalization across modalities. Neither encoding modality nor test modality affected electrodermal responsivity in the

CIT.

The current study is the first preregistered, highly powered study to examine modality effects in memory detection. Participants encoded several verbal and pictorial crime-related details (e.g., the getaway vehicle “bike,” an image of the murder weapon). Subsequently, they completed a web-based RT CIT that incorporated a verbal block and a pictorial block. Based on the dual-code theory, we anticipated a larger CIT effect (i.e., larger RTs to the crime-related than control items) for items that were encoded as images than for items that were encoded as words (H1). In line with the transfer-appropriate processing theory, we also expected that modality congruence between encoding and testing would induce a larger CIT effect than modality incongruence (H2).

Pilot

Method

This study was approved by the ethical committee of the Social and Behavioral Sciences faculty at the University of Amsterdam and registered as number 2020-CP-11832. The protocol was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. The pilot was not preregistered. All materials, data, and analytic scripts have been made publicly available and can be accessed at https://osf.io/84eas/.

Participants

Ninety-six participants took part in the experiment via the online crowdsourcing platform Prolific (https://www.prolific.co). Because of possible age-related differences in lying proficiency and executive control (Debey et al., 2015), only participants between 18 and 45 years old were eligible to participate in the study. They received £2.50 for their participation and had the chance to obtain a bonus of £1 if they would beat the lie detector test (i.e., if they would have a dCIT < 0.2; see Kleinberg & Verschuere, 2015).

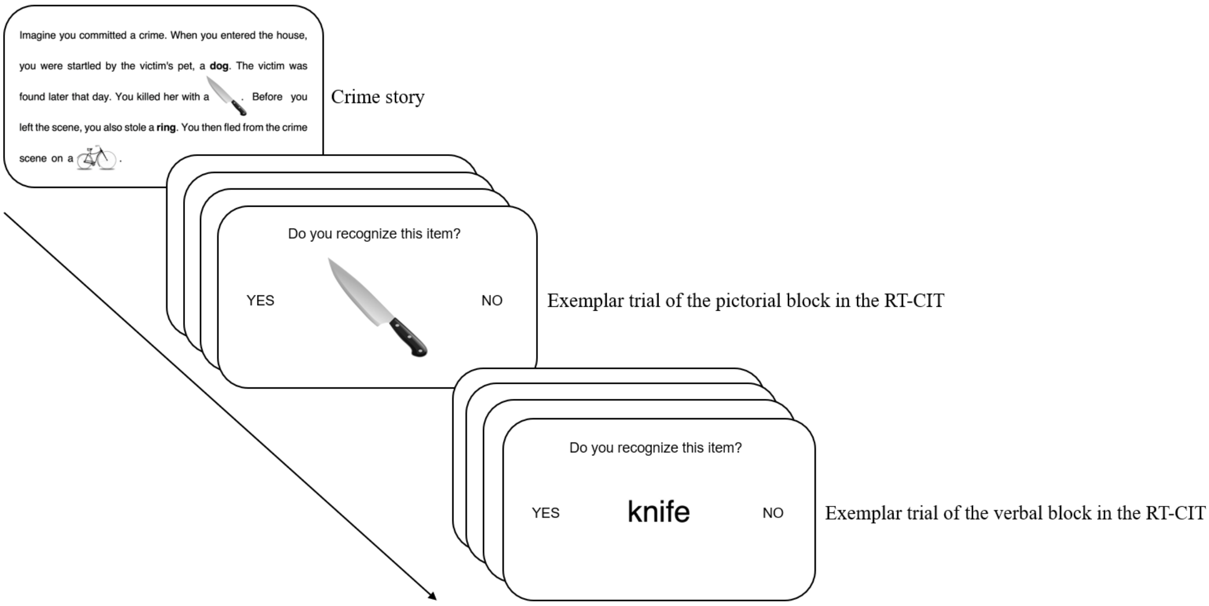
Out of these 96 participants, we excluded 4 participants because they terminated the experiment before it ended, 1 participant who did not recognize any of the crime-related items during the crime recognition check (see below), 1 participant who responded incorrectly to all items during the verbal test block, and another 17 participants because they had error rates above 50% for at least one stimulus type (i.e., target, crime-related, or control item). The final sample size contained 73 participants (70% male, 29% female, 1% other) with a mean age of 24 years (*SD* = 6 years, range 18-43 years). Participants' nationality was mainly Portuguese (27%), Polish (16%), or Italian (10%), and most participants indicated Portuguese (29%), Polish (16 %), or English (16%) as their native language.

Procedure

All participants gave informed consent before taking part in the study. The task was computerized (programmed in Inquisit 5.0), completed online, and can be found at https://osf.io/sqv67/. In the first phase of the experiment (i.e., the encoding phase), participants were asked to read and imagine a crime story that contained two verbal and two pictorial crime-related items (see below). Subsequently, all participants completed a memory game to ensure that all crime-related items were sufficiently encoded. In this memory game, participants had to flip virtual cards containing the crime-related items from the crime story in order to find the four matching pairs. In the second phase of the experiment (i.e., the test phase), participants had to try to hide their crime knowledge in an RT-CIT task (see below). This task consisted of two blocks: A verbal block in which the items were presented as written words and a pictorial block in which the items were presented as pictures. An illustration of a typical encoding and test phase can be found in Figure 1. After completing both blocks, participants were informed that the lie detection test had ended and were asked to respond honestly to all subsequent questions. To guarantee that participants still remembered the actual crime-related items, recognition was assessed by asking the participants to select each item they had seen in the crime story from a set of five items (the actual crime-related item and four control items) from the same category (e.g., all murder weapons), all presented in the modality in which the crime-related item was encoded. At the end of the experiment, participants provided demographic information about their age, gender, home country, and native language. The entire study was conducted in English.

Figure 1

Illustration of the experimental procedure consisting of an encoding phase (crime story) and a test phase (including a pictorial and a verbal block)

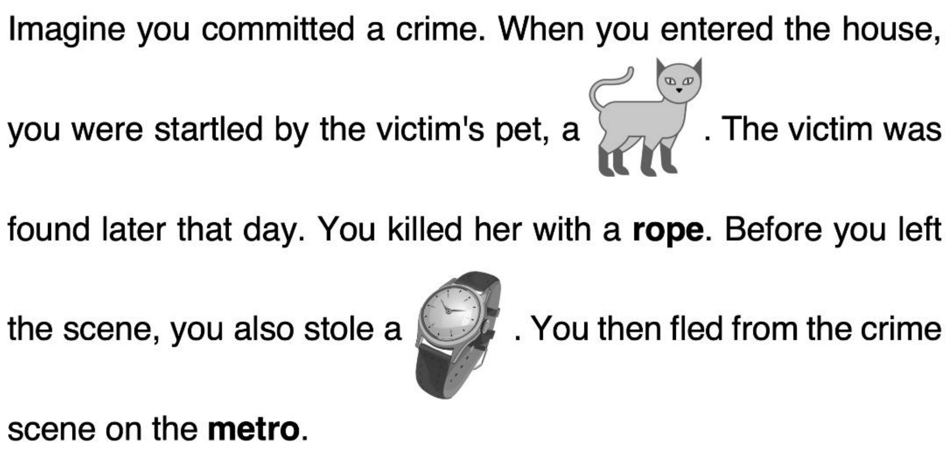


Materials

Crime Stories. To avoid stimulus-specific effects, we created four different profiles. These profiles determined which items were crime-related, target, or control items and which items were presented verbally or pictorially during the encoding phase for each participant (i.e., items that were crime-related/presented verbally for one participant were control items/presented pictorially for another). Participants were counterbalanced over these four profiles and received their crime story accordingly. The four crime stories included information about four crime-related categories: Murder weapon, stolen item, pet of the victim, and mode of transport. Of these crime-related items, two were presented as written words and two as images (see Figure 2 for an example; the four crime stories can be found at https://osf.io/bwc5d/). Participants were asked to read the story carefully. Afterward, there was a 20 seconds time interval in which participants were prompted to recall the four crime-related items. Subsequently, participants read the crime story again and filled possible memory gaps.

Figure 2

Illustration of one of the four crime stories



RT-CIT. Before the beginning of the RT-CIT, participants were instructed to conceal their knowledge of the crime-related items presented in the crime story. We informed them that if they would succeed to do this, they would obtain a monetary bonus. Then, a new set of four target items from the same categories as the crime-related items (i.e., one murder weapon, one stolen item, one pet, and one mode of transport) was introduced. Participants were instructed to memorize these targets.

During the RT-CIT task, crime-related items and targets were presented pseudo-randomly (see below) within a series of control items. Participants were prompted with the question “Do you recognize this item?” and instructed to respond ‘yes’ to target items and ‘no’ to all other stimuli (i.e., crime-related items and control items) by pressing the corresponding buttons on their keyboard as quickly as possible (i.e., ‘q’ and ‘p’ respectively). This procedure introduces a so-called stimulus-response incompatibility: Participants have to override their automatically activated response to the crime-related items (i.e., report crime-related items as recognized) and initiate the correct response instead (i.e., treat crime-related items as unrecognized, control items), a process that is known to be time-consuming and error-prone (see Verschuere & De Houwer, 2011). However, note that labeling an item as (un)recognized might not be a crucial aspect in the RT-CIT (Lukács & Ansorge, 2019). The targets are used to ensure attention to the stimuli and ensure that participants would not answer ‘no’ to all items.

As in Kleinberg and Verschuere (2015), all items were presented for 1500ms on the screen with randomly varying inter-stimulus intervals of either 250ms, 500ms, or 750ms to reduce anticipation and prepared responding. If no button was pressed 800ms after an item was displayed, the words “TOO SLOW” appeared for 200ms above the stimulus. If the participant pressed the wrong key (e.g., indicating ‘yes’ for a crime-related item), the words “WRONG” appeared for 200ms below the displayed item.

The test part of the RT-CIT was divided into two blocks, a verbal block, and a pictorial block. Each block contained instructions, training, and testing in one modality. The order of the blocks (first the verbal or first the pictorial) was counterbalanced across participants. Each block started with a training phase to familiarize participants with the RT-CIT task. This training phase consisted of three stages of 12 trials that became increasingly similar to the testing phase in terms of response limitations and feedback (see Table 1). In order to pass each stage of the training phase, participants had to meet certain accuracy and/or RT criteria. In stage 1 of the training phase, participants had to reach an overall target accuracy of 50%. In stage 2, we also required responses to be slower than 150ms on at least 20% of the trials to prevent participants from pressing the same button continuously. In stage 3, participants also needed to obtain an average RT that was faster than 800ms before they could move on to the testing phase. Participants had one re-attempt per stage if they failed to meet the criteria in their first attempt. After two unsuccessful attempts within the same training stage of the first modality section, the experiment was automatically terminated, since this indicated bad compliance with the task.

Table 1

Overview of Properties in the Training and Test Phase of the RT CIT

Stage	Trials	Stimulus duration	Accuracy feedback	Response time feedback
Training 1	12	Self-paced	“WRONG”	No feedback
Training 2	12	1500ms	“WRONG”	No feedback
Training 3	12	1500ms	“WRONG”	“TOO SLOW”
Test	480	1500ms	“WRONG”	“TOO SLOW”

After the training phase, the test trials started. Forty-eight unique stimuli were used (i.e., 24 written words and 24 images), of which eight were crime-related (i.e., four verbal, four pictorial), eight were targets (four verbal, four pictorial), and 32 were control items (16 verbal, 16 pictorial). Participants completed 240 trials in each of the two modality blocks. That is, all 24 items of the respective modality were presented in a random order and this process was repeated ten times. An overview of the test stimuli per crime story can be found at https://osf.io/p3xyf/.

General Overview of Data Analyses

In all data analyses, we only included trials with a correct response (i.e., 'no' for crime-related and control items) with RTs between 150ms and 800ms. We employed a within-subjects design: Each participant encoded verbal and pictorial stimuli and encountered both modalities in the test phase. That is, item modality (i.e., verbal versus pictorial) at encoding and at testing were the independent variables. The dependent variable was the standardized difference in RT (in milliseconds) between the crime-related and control items per participant (i.e., the CIT effect; dCIT = $[M_{RT(crime\text{-}related)} - M_{RT(control)}]/SD_{RT(control)}$; see also Kleinberg & Verschuere, 2015).

We tested our hypotheses by performing a 2 by 2 Bayesian repeated-measures (RM) ANOVA on the dCIT scores, with Encoding modality (verbal versus pictorial) and Test modality (verbal versus pictorial) as the two factors. In a (Bayesian) RM ANOVA, the same set of participants is tested multiple times under different conditions, after which is assessed whether there is a difference in means between the conditions. The Bayesian approach of an RM ANOVA compares the predictive performance of different models. The models under consideration are the null model, where there are no differences between the conditions (i.e., no effect), and the alternative models, where there are differences between the conditions (i.e., there is an effect). There are four alternative models in our analysis: The model with only a main effect of encoding, the model with only a main effect of testing, the model with two main effects (encoding and testing), and the model with two main effects (encoding and testing) and their interaction. For each model, a Bayes Factor (BF) is computed. This BF reflects how likely the data are under that model compared to the null model. BF_{01} shows evidence in favor of the null model and BF_{10} shows evidence in favor of an alternative model. Subsequently, we performed follow-up post hoc comparisons (Bayesian t-tests controlled for multiplicity). After comparing the different models, we compared the different effects. For each effect (i.e., both main effects and the interaction effect), we computed a $BF_{inclusion}$ reflecting a comparison of all models containing a particular effect to those without the effect. In other words, $BF_{inclusion}$ can be interpreted as the evidence in the data for including an effect or interaction, similar to BF_{10} in the case of simple comparisons (see also van den Bergh et al., 2020). Lastly, we checked whether there was a CIT effect present in each modality condition. Therefore, we performed both a one-tailed paired samples t-test and a Bayesian one-tailed paired samples t-test on the crime-related-control item difference (in ms). For all analyses, we used JASP's default prior *r* scale of 0.5 for fixed effects in the Bayesian RM ANOVA and the default Cauchy scale of 0.707 for all follow-up Bayesian paired sample *t*-tests.

In the RT-CIT task, error rates tend to be less reliable and valid than RTs due to floor effects (i.e., often <5% errors; see Kleinberg & Verschuere, 2015). For the sake of completion, results for error rates can be found at https://osf.io/7gh89/ for the pilot and https://osf.io/ru8eb/ for study 1.

Results

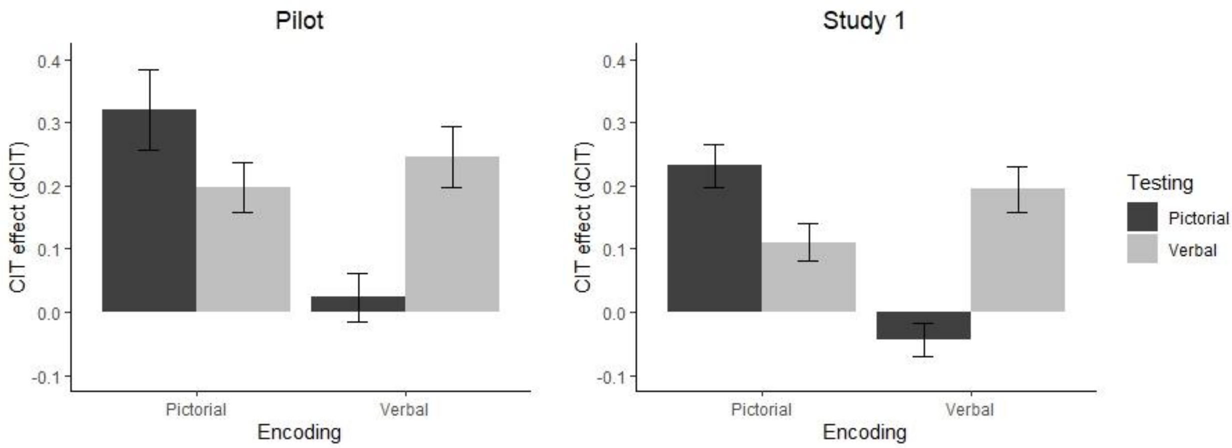
Modality Effects

Using a Bayesian RM ANOVA, the alternative model that included the two main effects Encoding and Testing and their interaction turned out to be the best performing model. The BF indicated that the data are 82.27 times more likely under that model than under the null model. The data showed substantial evidence for the main effect of Encoding ($BF_{inclusion}$ = 3.65), and also our post hoc comparison showed anecdotal evidence for a difference between verbally and pictorially encoded items (BF_{10} = 2.08), such that items encoded as pictures (M_{dCIT} = 0.26; SD_{dCIT} = 0.45) yielded larger dCIT scores than items encoded as words (M_{dCIT} = 0.13; SD_{dCIT} = 0.39). However, a Bayesian comparison between the Pictorial-Pictorial and the Verbal-

Verbal condition showed substantial evidence for the absence of a difference (BF_{01} = 4.83). We found decisive evidence for the presence of an interaction between Encoding and Testing ($BF_{inclusion}$ = 104.82). dCIT scores were larger when there was a match between encoding and test modality (M_{dCIT} = 0.28; SD_{dCIT} = 0.48) than when there was a mismatch in modalities (M_{dCIT} = 0.11; SD_{dCIT} = 0.34). Lastly, we found substantial evidence for the absence of an effect of Testing ($BF_{exclusion}$ = 4.55), which was also confirmed in our post hoc comparison (BF_{01} = 6.75). The results are summarized in Figure 3.

Figure 3

Standardized differences between response times to crime-related and control stimuli (dCIT scores) for verbal versus pictorial encoding and testing in the pilot and in study 1.



Note. Error bars represent standard errors.

CIT Effect for each Modality Combination

We examined whether RT-based CIT effects were present in each modality condition and which modality condition yielded the largest CIT effect (i.e., Pictorial-Pictorial, Verbal-Verbal, Verbal-Pictorial, or Pictorial-Verbal). Descriptive statistics (means and standard deviations of RTs of crime-related items and control items, the corresponding effect size, its 95% confidence interval (CI), and the corresponding BF) can be found in Table 2. Our results indicate the presence of a CIT effect in the Pictorial-Pictorial condition (M_{dCIT} = 0.32; SD_{dCIT}

= 0.54), the Verbal-Verbal condition (M_{dCIT} = 0.25; SD_{dCIT} = 0.41), and the Pictorial-Verbal condition (M_{dCIT} = 0.20; SD_{dCIT} = 0.33), but not in the Verbal-Pictorial condition (M_{dCIT} = 0.02; SD_{dCIT} = 0.33), suggesting that participants did not convert verbal information to pictorial information.

Table 2

Overview of Reaction Time-Based Crime-related-Control item Differences for each Modality Combination in the Pilot Study (N=73)

	RTs		Differences between crime-related and control items	
Condition	Crime-related	Control	*d* [95% CI]	*BF*$_{10}$
	M (*SD*)	*M (SD)*		
Pictorial – Pictorial	482 (59)	457 (37)	0.60 [0.38, ∞]	11 210.80
Verbal – Verbal	504 (65)	480 (49)	0.60 [0.39, ∞]	11 922,80
Pictorial – Verbal	499 (65)	480 (49)	0.59 [0.38, ∞]	8 802.46
Verbal – Pictorial	459 (49)	457 (37)	0.08 [-0.11, ∞]	4.17[a]
Across modalities	486 (62)	468 (45)	0.48 [0.37, ∞]	7.96×10^{11}

Note. Reaction times are given in milliseconds. Cohen's d expresses the size of the RT difference between crime-related items and control items. A Cohen's d of 0.20 can be interpreted as a small effect, a Cohen's d of 0.50 as a medium effect, and a Cohen's d of 0.80 as a large effect (Cohen, 1988). The BF_{10} expresses how much more likely the alternative model is (participants were slower for crime-related items than for control items) than the null model.
[a]Reported number is BF_{01} which shows evidence for the null.

Study 1

The pilot showed that congruency in modality between encoding and testing was associated with better detection than incongruency and that it was easier to detect recognition of items that were encoded as pictures than items that were encoded as written words. Importantly, our results suggest that participants did not convert verbal information to pictorial information: Participants did not show recognition when items were tested as pictures but encoded as words. To test the robustness of these findings and to rule out the possibility that the pilot's results were due to the use of a mostly non-native English-speaking sample, we recruited a new sample that consisted of 50%

native English speakers and 50% non-native English speakers in study 1.

Method

The method of study 1 was identical to that of the pilot with the only exception that we now collected a sample comprised of 50% native English speakers and 50% non-English speakers. The study was preregistered (https://osf.io/vrkfq). All materials, data, and analytic scripts have been made publicly available and can be accessed at https://osf.io/84eas/.

Participants

As described in our preregistration, we first opened up the experiment to 100 participants on Prolific. For 50 of these, we required their native language to be English, for the other 50, we allowed every native language except English. Eligibility was restricted to participants between the ages of 18 and 45. Then, after applying our preregistered exclusion criteria (see below), we ran a 2 (encoding: verbal versus pictorial) by 2 (testing: verbal versus pictorial) Bayesian RM ANOVA on the dCIT scores in JASP. The decision to stop collecting data was based on the BF of the main effect of encoding and the BF of the interaction effect (by adding the main effects of encoding and testing as nuisance variables to the null model). We planned that once we would reach substantial evidence for either the alternative hypothesis (i.e., BF_{10} larger than 5) or the null hypothesis (i.e., BF_{10} smaller than 1/5) for both models (i.e., the main effect of encoding and the interaction between encoding and testing), we would stop testing, otherwise we would open up the experiment for another 100 participants (Schönbrodt et al., 2017). We intended to repeat this procedure until we would reach substantial evidence for both models for either the null or the alternative hypothesis or until we would have tested N = 500. We reached substantial evidence favoring the alternative hypotheses for both models after opening up the experiment two times for 100 participants. In total, we recruited 196 participants. Participants received £2 as compensation and an additional £1 bonus payment when they would beat the lie detector test (i.e., dCIT < 0.2; see

Kleinberg & Verschuere, 2015).

Based on our preregistered exclusion criteria, we excluded 6 participants because they terminated the experiment, 4 participants because they did not recognize any crime-related items by the end of the experiment, 2 participants due to high (>50%) error rates on at least one experimental test (i.e., verbal or pictorial, both participants had less than 50% accuracy on the verbal block), and another 26 participants due to high (>50%) error rates on at least one item type (i.e., target, crime-related, or control item; additionally, 1 participant who had less than 50% accuracy on the verbal block also had less than 50% accuracy for an item type). Our final sample consisted of 158 participants (49% men, 48% women, 2% other gender, and 1% preferred not to say) with a mean age of 26 years (*SD* = 7 years, range 18-45 years). Most participants indicated to be of British (32 %), Portuguese (13%), or Polish (13%) nationality, and most reported English (47%), Portuguese (13%), or Polish (13%) as their native language.

Results

Confirmatory Analyses

The BF indicated that the data were 2.98 x 10^7 times more likely under the alternative model that included both main effects of Encoding and Testing and their interaction than under the null model. The $BF_{inclusion}$ indicated strong evidence for an effect of Encoding ($BF_{inclusion}$ = 10.77) and the post hoc comparison indicated substantial evidence for a difference between verbal and pictorial encoding (BF_{10} = 8.01). Items that were encoded as pictures yielded a higher CIT effect (M_{dCIT} = 0.17; SD_{dCIT} = 0.40) than items that were encoded as words (M_{dCIT} = 0.08; SD_{dCIT} = 0.41). However, a Bayesian comparison between the Pictorial-Pictorial and the Verbal-Verbal condition showed substantial evidence for the absence of a difference (BF_{01} =

8.71). In addition, the analysis indicated decisive evidence for the model with the interaction effect between Encoding and Testing ($BF_{inclusion}$ = 5.27 × 10^6). That is, modality congruency between encoding and testing items (i.e., the Pictorial-Pictorial and Verbal-Verbal conditions) resulted in larger dCIT scores (M_{dCIT} = 0.21; SD_{dCIT} = 0.43) than incongruent modality combinations (i.e., the Pictorial-Verbal and Verbal-Pictorial conditions; M_{dCIT} = 0.03; SD_{dCIT} = 0.36). Lastly, we found anecdotal evidence for the absence of an effect of Testing ($BF_{exclusion}$ = 1.89), which was also confirmed in our post hoc comparison (BF_{01} = 2.23). Our findings are summarized in Figure 3.

Exploratory Analyses

CIT Effect for each Modality Combination. We also tested whether each item modality combination induced a CIT effect and which combination yielded the largest CIT effect. Descriptive statistics (means and standard deviations of RTs of crime-related items and control items, the corresponding effect size, its 95% CI, and the corresponding BF) can be found in Table 3. The results indicate the presence of a CIT effect in the Pictorial-Pictorial condition (M_{dCIT} = 0.23; SD_{dCIT} = 0.42), the Verbal-Verbal condition (M_{dCIT} = 0.19; SD_{dCIT} = 0.45), and the Pictorial-Verbal condition (M_{dCIT} = 0.11; SD_{dCIT} = 0.38), but not in the Verbal-Pictorial condition (M_{dCIT} = -0.04; SD_{dCIT} = 0.32), suggesting that participants did not convert verbal to pictorial information.

Table 3

Overview of Reaction Time-Based Crime-related-Control item Differences for each Modality Combination in Study 1 (N=158)

	RTs		Differences between crime-related and control items	
Condition	Crime-related	Control	*d* [95% CI]	BF_{10}
	M (SD)	*M (SD)*		
Pictorial – Pictorial	476 (56)	457 (42)	0.52 [0.38, ∞]	1.97×10^{7}
Verbal – Verbal	485 (66)	469 (46)	0.40 [0.27, ∞]	19 261.05
Pictorial – Verbal	478 (60)	469 (46)	0.26 [0.13, ∞]	28.71
Verbal – Pictorial	453 (52)	457 (42)	-0.14 [-0.27, ∞]	33.33[a]
Across modalities	473 (60)	463 (44)	0.28 [0.21, ∞]	8.53×10^{8}

Note. Reaction times are given in milliseconds. Cohen's d expresses the size of the RT difference between crime-related items and control items. The BF_{10} expresses how much more likely the alternative model is (participants were slower for crime-related items than control items) than the null model. [a] Reported number is BF_{01} which shows evidence for the null.

Modality Effects and Language. To examine possible language effects, we performed a 2 (Encoding: verbal versus pictorial, within-subjects) by 2 (Testing: verbal versus pictorial, within-subjects) by 2 (Language: native versus foreign, between-subjects) Bayesian RM ANOVA on the dCIT scores. We found substantial evidence for the absence of a three-way interaction between Encoding, Testing, and Language ($BF_{inclusion}$ = 4.55), suggesting native versus non-native language did not moderate the effect. A more detailed discussion of the results of this exploratory analysis can be found at https://osf.io/zwfx8/.

Discussion

The goal of this study was to examine modality effects in concealed memory detection. Our first aim was to investigate whether item modality at encoding would influence the CIT effect. Next, we examined whether the CIT effect is affected by (in)congruency between item modalities at encoding and testing. Our results were qualified by a robust interaction between encoding and test modality. That is, congruent item modalities at encoding and testing yielded larger CIT effects, while incongruent modality combinations resulted in a comparatively lower (for Pictorial-Verbal) or even absent (for Verbal-Pictorial) CIT effect.

Perceptual Processing

Although the presence of the interaction limits its generalizability, items that were encoded pictorially were easier to detect than items that were encoded verbally, especially when testing was done using pictures. Moreover, no conversion across modalities was observed when items were encoded as words. These findings are in line with the dual code theory (Paivio, 1971), which suggests that pictures are processed via both the verbal and nonverbal memory pathways and are therefore stored as an image (e.g., picture of a dog) and its corresponding label (e.g., “dog”). In comparison, words only invoke a verbal code (e.g., “dog”). Therefore, items that were encoded as pictures were more accessible at testing, resulting in a larger CIT effect for encoded pictures than encoded words. Furthermore, if participants only store a word as its verbal code, but are then prompted with the corresponding picture, retrieval is hampered, resulting in the lack of a CIT effect in the Verbal-Pictorial condition. Similarly, the dual code theory also explains the CIT effect in the incongruent Pictorial-Verbal condition. Encoded pictures invoke both a visual and a verbal code: Even if participants are prompted with the corresponding word during testing, they can still retrieve the accompanying verbal code from memory.

Conceptual Processing

The finding that congruent item modality at encoding and testing resulted in a larger CIT effect than incongruent item modality is consistent with the transfer-appropriate processing account. This theory assumes that the cognitive processes performed during encoding are recapitulated at retrieval (see Rugg et al., 2015). Therefore, this theory predicts better memory performance when the type of processing is shared between encoding and retrieval. In the context of the current experiment, congruent item modalities at encoding and testing (Pictorial-Pictorial or Verbal-Verbal) resulted in higher CIT effects. Previous research using the RT-CIT have found similar results: Items that were encoded on a categorical (e.g., car) or exemplar level (e.g., Citroën) incurred the largest CIT effect if the test modality was congruent (i.e., category-category or exemplar-exemplar; see Geven et al., 2019). Also in memory literature, it has been shown that similarities between study and testing are important determinants of memory performance (e.g., Stenberg et al., 1995; Weldon & Roediger, 1987). The current results are inconsistent with the results reported by Ben-Shakhar et al. (1996) that revealed a complete generalization of the skin conductance CIT effect from the verbal to the pictorial mode and vice versa. This inconsistency may be related to the use of very different dependent measures (skin conductance versus response time) which could be driven by different underlying mechanisms (e.g., klein Selle et al., 2016, 2017). Future studies will be needed to resolve this issue.

Practical Implications

Our results suggest that examiners should consider in which modality a guilty suspect likely encoded crime-related information and present items in a congruent modality during the test. This way, the detection efficiency of the RT-CIT could be maximized. Since it can be challenging to find out how a suspect encoded crime information, future research could examine the possibility of presenting pictorial and verbal items simultaneously at testing (e.g., a picture of the victim’s dog with the word “dog” printed below or above the picture).

The current findings also bear significance to one of the biggest challenges in CIT: The leakage problem. The notion that only guilty suspects have knowledge about a crime is, unfortunately, not always the case. Through newspapers, social media, or even during interrogation itself, innocent suspects may be exposed to critical details of the crime. Since the CIT only detects whether a suspect has knowledge of particular aspects of a crime, but not how this knowledge was acquired, leakage of information can induce an enhanced response to the crime-related items in informed innocent participants (Bradley et al., 2011). Based on our results, it could be assumed that if the perpetrator encoded crime-related aspects pictorially (i.e., visually) at the crime scene, while an innocent suspect only read or heard about the crime in a verbal form, then a pictorial CIT could protect the innocent suspect from being falsely accused. It can be argued that leakage is not always limited to a purely verbal form and encoding of a crime is not always limited to a pictorial form. For instance, leakage can also occur through pictures or videos of the crime, and certain crime-related information could be encoded verbally by the perpetrator (e.g., what was said in an extortion or threat, or the name of a city). However, since it is much more likely that leakage occurs verbally, and, arguably, almost all crime-related details are encoded pictorially, the promising lack of conversion from verbal to pictorial information found in the current study could possibly protect informed innocents in a pictorial CIT. We argue that the better the match between test and encoding, the better the CIT will be able to detect the guilty and protect the innocent. Related, in a study by Norman et al. (2020), crime scenes that were revisited in virtual reality resulted in even better concealed recognition than 2D images of these crime scenes.

Limitations

A possible limitation of the current study could be the fact that the stimuli were not split into crime-related, target, or control item in a fully randomized manner. We prevented stimulus-specific effects by preparing four profiles so that items that were crime-related for one participant were control items for another. However, it is still theoretically possible that the items that were crime-related in each profile were simply more salient or took more processing time for some reason than the other items. Furthermore, as images have many varying properties, it is difficult to properly compare them to verbal equivalents in a manner that can be safely applied to other image types (e.g. from drawings in the current study to photographs or even virtual reality), limiting the generalizability of the reported results.

Conclusion

The current study demonstrated that successful detection of crime knowledge with the RT-CIT depends on the way items are encoded and tested. We found evidence for a robust effect of modality congruence. That is, our results suggest that the CIT effect can be maximized if the testing modality is congruent to the encoding modality. Interestingly, we also found no conversion from verbal encoding to pictorial testing. When participants encoded the crime-related items verbally and were tested with pictorial representation of the items, crime-knowledge was not detected. These results are in line with the principles of the dual code and transfer-appropriate processing account and highlight the involvement of both perceptual, but in particular conceptual memory processing in the RT-CIT.

References

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258-267. doi:10.1016/j.ijpsycho.2009.12.007

Ben-Shakhar, G., Frost, R., Gati, I., & Kresh, Y. (1996). Is an apple a fruit? Semantic relatedness as reflected by psychophysiological responsivity. *Psychophysiology, 33*(6), 671-679. doi:10.1111/j.1469-8986.1996.tb02363.x

Bevan, W., & Steger, J. A. (1971). Free recall and abstractness of stimuli. *Science, 172*(3983), 597-599. doi:10.1126/science.172.3983.597

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. *Memory detection: Theory and application of the Concealed Information Test,* 187. doi:10.1017/cbo9780511975196.011

Childers, T. L., & Houston, M. J. (1984). Conditions for a picture-superiority effect on consumer memory. *Journal of Consumer Research, 11*(2), 643-654. doi:10.1086/209001

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

Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior Pinocchio: A cross-sectional lifespan investigation of deception. *Acta psychologica*, *160*, 58-68. doi:10.1016/j.actpsy.2015.06.007

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory & Cognition, 47*(1), 182-193. doi:10.3758/s13421-018-0858-9

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

JASP Team (2020). *JASP* [Computer software]. https://jasp-stats.org/

Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychological Review, 1*(6), 602. doi:10.1037/h0068244

Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test. *PLOS ONE, 10*(4), e0118715. doi:10.1371/journal.pone.0118715

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: Orienting and inhibition processes in the CIT. *Psychophysiology, 53*(4), 579-590. doi:10.1111/psyp.12583

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: Orienting and inhibition in the CIT. *Psychophysiology, 54*(4), 628-639. doi:10.1111/psyp.12825

Lukács, G., & Ansorge, U. (2019). Information leakage in the response time-based Concealed Information Test. *Applied Cognitive Psychology, 33*(6), 1178-1196. doi:10.1002/acp.3565

Madigan, S. (2014). Picture memory. *Imagery, Memory and Cognition*, 65-89.

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

Nelson, D. L., Cermak, L., & Craik, F. (1979). Remembering pictures and words: Appearance, significance and name. *Levels of processing in human memory*, 45-76. doi:10.4324/9781315796192-11

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.* doi:10.1016/j.jarmac.2019.12.008

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. *Memory detection: Theory and application of the Concealed Information Test*, 253-275. doi:10.1017/CBO9780511975196.015

Paivio, A. (1971). *Imagery and Verbal Processes* (1st ed.)., New York: Holt, Rinehart and Winston.

Rosenfeld, J. P., Ward, A., Thai, M., & Labkovsky, E. (2015). Superiority of pictorial versus verbal presentation and initial exposure in the P300-based, complex trial protocol for concealed memory detection. *Applied Psychophysiology and Biofeedback*, *40*(2), 61-73. doi:10.1007/s10484-015-9275-z

Rugg, M. D., Johnson, J. D., & Uncapher, M. R. (2015). Encoding and retrieval in episodic memory. *The Wiley Handbook on the Cognitive Neuroscience of Memory*, 84-107. doi:10.1002/9781118332634.ch5

Sansgiry, S. S., Cady, P. S., & Adamcik, B. A. (1997). Consumer comprehension of information on over-the-counter medication labels: effects of picture superiority and individual differences based on age. *Journal of Pharmaceutical Marketing & Management*, *11*(3), 63-76. doi:10.3109/j058v11n03_05

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological methods*, *22*(2), 322. doi:10.1037/met0000061

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology: The Official*

Journal of the Society for Applied Research in Memory and Cognition, 22(4), 475-490. doi:10.1002/acp.1375

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156-163. doi:10.1016/s0022-5371(67)80067-7

Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology, 18*(6), 813-847. doi:10.1080/09541440500412361

Stenberg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory & Cognition, 23*(4), 425-441. doi:10.3758/bf03197244

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), 428-453. doi:10.1037/bul0000087

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Gupta, A. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique, 120*(1), 73-96. doi:10.3917/anpsy1.201.0073

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: Response latency-based measures. *Memory detection: Theory and application of the Concealed Information Test,* 46-62. doi:10.1017/CBO9780511975196.004

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15*(4), 269-280. doi:10.3758/bf03197030

Weldon, M. S., Roediger, H. L., & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition*, *17*(1), 95-105. doi:10.3758/bf03199561