Memory detection: The effects of emotional stimuli

klein Selle, N.; Verschuere, B.; Kindt, M.; Meijer, E.; Nahari, T.; Ben-Shakhar, G.

DOI
10.1016/j.biopsycho.2017.07.021

Publication date
2017

Document Version
Final published version

Published in
Biological Psychology

Biological Psychology 129 (2017) 25–35

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Memory detection: The effects of emotional stimuli

Nathalie klein Selle[a,b,*], Bruno Verschuere[b], Merel Kindt[b], Ewout Meijer[c], Tal Nahari[a], Gershon Ben-Shakhar[a]

[a] Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel
[b] Department of Clinical Psychology, University of Amsterdam, Postbus 15933, 1001 NK Amsterdam, The Netherlands
[c] Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

ARTICLE INFO

Keywords:
The concealed information test (CIT)
External validity
Arousal
Valence
Skin conductance response (SCR)
Respiration line length (RLL)
Heart rate (HR)

ABSTRACT

The Concealed Information Test (CIT) aims to detect the presence of crime-related information in memory. In two experiments, we examined the influence of stimulus emotionality on the outcomes of the CIT. In experiment 1, each participant was tested immediately or after one week, on a series of neutral and either negative arousing or negative non-arousing pictures. CIT detection efficiency was unaffected, but physiological and recognition data did not support the manipulation's effectiveness. In experiment 2, each participant was tested after a week on a series of neutral versus negative arousing pictures. Importantly, stimulus arousal was increased and memory ceiling effects were prevented. This time, both memory and CIT detection efficiency using the skin conductance, but not the respiration and heart rate measures, were enhanced for emotional compared to neutral pictures. Taken together, these results indicate that the use of emotional stimuli does not deteriorate and may even improve CIT validity.

1. Introduction

More than a century ago, William James (1890, p. 670) described the impact of emotion on memory as leaving "*a scar upon the cerebral tissues*". James's classical statement depicts the common intuition that highly emotional events are remembered with extensive clarity and detail and this holds particular relevance for the memory detection approach. This knowledge-based approach using the Concealed Information Test (CIT; Lykken, 1959; Verschuere, Ben-Shakhar, & Meijer, 2011) aims to uncover the presence of crime-related information in memory by measuring physiological and/or behavioral responses. In a typical CIT, examinees are presented with a series of multiple choice-like questions, each designed to examine suspects' knowledge of a distinctive crime-detail assumed to be known only to individuals involved in the crime and the investigative authorities. For each question, this critical feature of the crime (i.e., the crime-related item) is intermixed among several other plausible control items (e.g., Where was the victim's body found?... basement?... graveyard?... dumpster?... barn?... river?...). Provided that the crime-related information has been kept from the public, an innocent suspect will not be able to differentiate between the different alternatives and is therefore expected to show similar responses to all items. A guilty suspect, on the other hand, will recognize the crime-related alternatives and will show differential responses to these items (e.g., an increased skin conductance response (SCR), a shorter respiration line length (RLL), and a larger heart rate (HR) deceleration – Gamer, 2011). This pattern of differential responses elicited by the crime-related items has been labeled as the CIT effect.

While extensive research has demonstrated large CIT effect sizes with both autonomic nervous system measures and event-related potential measures (Meijer, klein Selle, Elber, & Ben-Shakhar, 2014), the external validity of this research has been questioned (e.g., Ben-Shakhar, 2012; Meijer, Verschuere, Gamer, Merckelbach, & Ben-Shakhar, 2016). Unfortunately however, it is extremely difficult to conduct methodologically sound field studies in this area (see Ginton, Daie, Elaad, & Ben-Shakhar, 1982; Iacono, 1991). CIT researchers have therefore used an alternative approach and systematically manipulated various factors that differ between the typical experimental and the forensic setting: the time between processing the crime related items and the CIT (e.g., Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Gamer, Kosiol, & Vossel, 2010; Nahari & Ben-Shakhar, 2011), the motivation to avoid detection (see Meijer et al., 2014) and the free choice to commit a crime (Nahari, Breska, Elber, klein Selle, & Ben-Shakhar, 2017). Another factor differentiating the experimental from the real-world set-up is the experienced emotional arousal – which is expected to be higher in the field than in the laboratory (see also Verschuere, Meijer, & De Clercq, 2011). A number of previous studies have examined this factor, either by manipulating the level of arousal

* Corresponding author at: Nathalie klein Selle Department of Psychology Hebrew University of Jerusalem, Mt. Scopus Jerusalem 91905 Israel.
E-mail addresses: nathalie.kleinselle1@mail.huji.ac.il, nkleinselle@gmail.com (N. klein Selle).

http://dx.doi.org/10.1016/j.biopsycho.2017.07.021
Received 6 February 2017; Received in revised form 26 July 2017; Accepted 28 July 2017
Available online 03 August 2017
0301-0511/

experienced during the mock-crime (Peth, Vossel, & Gamer, 2012) or by manipulating the level of arousal experienced during the CIT (Bradley & Janisse, 1981; Kugelmass & Lieblich, 1966), and found little to no effect on CIT detection efficiency.

Importantly however, in real-life forensic cases, not only the subjective experience, but also the crime-related stimuli themselves may be emotionally arousing. While the stimuli may become emotional only through their connection to the event (e.g., think of the otherwise neutral hammer that has received an emotional loading through its use in a murder), they may also be intrinsically arousing (e.g., the way a victim was raped). Naturally, it may be expected that the more violent the crime, the higher the emotionality of the crime-related items. Let's consider for example the sensational O.J. Simpson case that has captivated America for more than two decades. O.J. was a former NFL star suspected of murdering his ex-wife, Nicole Brown Simpson, and her friend, Ronald Goodman. Several important crime-related details, which could have been used in a CIT, were emotional in nature: the bloody glove, the victims' bodies and the modus operandi of the murder − the modus operandi (e.g., By beating?... By stabbing?... By drowning?... By strangling?... By poisoning?...) is actually a central feature in real life CIT investigations in Japan where the test is extensively applied (see Osugi, 2011). Japanese CIT examiners prefer these type of arousing stimuli because of their presumed high memorability. As the CIT is typically administered few weeks after a crime was committed, one of the key challenges facing CIT practitioners is the selection of memorable items that will induce a stable CIT effect.

A large body of memory research has shown that the emotional value of the stimuli affects their memorability. Most memory studies have however focussed on arousal and there is only some initial evidence that valence can influence how well a stimulus is remembered (see Kensinger, 2009). Studies that manipulated arousal have shown that all three stages of memory processing (i.e., encoding, storage and retrieval) are improved for arousing stimuli. Specifically, arousal leads to a "narrowing of attention" around the central emotional details of a stimulus, reducing the range of cues to which an organism is sensitive (Easterbrook, 1959; Kensinger, Garoff-Eaton, & Schacter, 2007; Sharot & Phelps, 2004). Further, by controlling for the influence of attention (encoding), several studies have shown that emotionally arousing stimuli are more likely to be stored in memory (e.g., Bradley, Greenwald, Petry, & Lang, 1992; Sharot & Phelps, 2004). Specifically, these studies found that memories for neutral stimuli worsen over time, whereas memories for arousing stimuli remain or even improve over time. Finally, several studies have shown that arousing stimuli (both scene-images and words) lead to better recollection than neutral stimuli (e.g., Kensinger & Corkin, 2003; Ochsner, 2000). In addition to laboratory research on the retrieval of arousing stimuli, research on the retrieval of arousing autobiographical events supports the positive influence of arousal on recollection (e.g., Bahrick, Parker, Fivush, & Levitt, 1998; Brown & Kulik, 1977; Budson et al., 2004, 2007; Sharot, Martorella, Delgado, & Phelps, 2007). Taken together, although real-life and laboratory stimuli may differ in both valence and arousal, arousal is most likely to influence memory, and consequently, the CIT effect.

The foremost aim of the two experiments presented here was to examine whether and how emotional stimuli influence CIT detection efficiency. This was accomplished by using pictures that differ in valence and arousal. To the best of our knowledge, there are no prior published studies on the topic and the vast majority of CIT studies were conducted using neutral (e.g., playing cards, mock-crime related) or even positive (e.g., personal/autobiographical) low-arousing stimuli. In Experiment 1 the emotional stimuli were either negative arousing or negative non-arousing, while in Experiment 2 all emotional stimuli were negative arousing. Based on the extensive memory literature, we formulated three separate, but related, predictions: (1) the emotional crime-related stimuli will be better remembered than the neutral crime-related stimuli; (2) the enhanced memory of the emotional crime-related stimuli will heighten CIT detection efficiency; (3) this emotional heightening effect on both memory and CIT detection efficiency will be most clear when the CIT is delayed.

2. Experiment 1

In experiment 1, each participant was tested, either immediately or after one week, on a series of neutral versus emotional pictures. To clarify whether arousal rather than valence would be key in enhancing memorability, the emotional pictures were either negative arousing or negative non-arousing.

3. Method

3.1. Participants

One hundred and thirty-six undergraduate students (91 women) of the Hebrew University of Jerusalem (HUJI) with an age range of 18–32 ($M = 24.1$, $SD = 2.7$ years) participated in this experiment. All participants were native speakers of Hebrew and received either course credits or an average payment of 35 NIS (equivalent to approximately 10 USD) for their participation. Each participant read and signed a consent form indicating that participation was voluntary and that they could withdraw from the experiment at any time without penalty. The experiment was approved by the ethical committee of the Faculty of Social Sciences of the HUJI.

3.2. Design

Experiment 1 was constructed in a 2 × 2 × (2 × 2) mixed design with both arousal (negative arousing vs. negative non-arousing) and the time delay between encoding and CIT (immediate vs. 1 week delayed) as between-subjects factors and both item-type (crime-related vs. control) and valence (negative vs. neutral) as within-subject factors. This resulted in four experimental between-subjects conditions: (1) Arousing immediate; (3) Arousing delayed; (2) Non-arousing immediate; (4) Non-arousing delayed. Participants were randomly assigned to one of the four conditions. Sixty-eight participants were tested immediately with 38 of them receiving negative arousing images (i.e., arousing immediate condition) and 30 receiving negative non-arousing images (i.e., non-arousing immediate condition). Sixty-eight participants were tested after a one week delay with 34 of them receiving negative arousing images (i.e., arousing delayed condition), and 34 receiving negative non-arousing images (i.e., non-arousing delayed condition). These conditions did not differ in gender ($\chi^2(3, N = 136) = .49$, $p = .922$), but did differ in age, $F(3,132) = 2.93$, $f = .21$, $p = .036$; nonetheless, participants' mean age in the different conditions varied between 23.3 and 24.8.

3.3. Items

The arousal level and valence of the CIT items were manipulated by using two well validated and well researched picture sets (see Appendix A Table A1): the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and the Karolinska Directed Emotional Faces (KDEF; Goeleven, De Raedt, Leyman, & Verschuere, 2008; Lundqvist, Flykt, & Öhman, 1998). The selected pictures were either neutral or negative non-arousing or negative arousing. Importantly, within each CIT question, the crime-related item and all control items were equal in their arousing potential. Thus, stimuli from the selected categories (i.e., faces and scenes) were chosen on the basis of their arousal (low vs. high) and valence (neutral vs. negative) ratings, such that every question contained one crime-related item and several control items that were matched with respect to semantic category, arousal and valence levels.

The normative arousal ratings for the emotional (negative arousing and negative non-arousing) IAPS and KDEF items used in the CIT were as follows: The negative arousing pictures had a mean arousal rating of

6.29 (faces: $M = 6.72$, $SD = .54$; scenes: $M = 5.86$; $SD = .28$; on a scale of 1–9) and the negative non-arousing pictures had a mean arousal rating of 3.94 (faces: $M = 3.61$, $SD = .59$; scenes: $M = 4.26$, $SD = .82$; on a scale of 1–9). An independent samples t test on the difference between the ratings of the negative arousing and the negative non-arousing pictures was statistically significant, $t(26) = 9.05$, $p < .001$, $d = 3.42$.

We also checked whether the arousal ratings generalized to the sample under investigation (undergraduates of the HUJI) in a pilot study.[1] In this pilot, twenty-nine undergraduates rated the emotional (negative arousing and negative non-arousing) CIT items on arousal. The pilot ratings were as follows: The negative arousing pictures received a mean arousal rating of 4.88 (faces: $M = 6.15$, $SD = .27$; scenes: $M = 3.62$, $SD = .52$; on a scale of 1–10) and the negative non-arousing pictures received a mean arousal rating of 2.10 (faces: $M = 1.85$, $SD = .28$; scenes: $M = 2.35$, $SD = .73$; on a scale of 1–10). An independent samples t test on the difference between the ratings of the negative arousing and the negative non-arousing pictures was statistically significant, $t(56) = 6.08$, $p < .001$, $d = 1.60$.

3.4. Procedure

All participants arrived at the laboratory and were welcomed by Experimenter 1. Then, after signing an informed consent form, participants were tested on blood phobia using the Medical Fear Survey (MFS; Kleinknecht, Kleinknecht, Sawchuk, Lee, & Lohr, 1999). When the score on the MFS was below the cut-off value, participants started the first part of the experiment, i.e., the study phase.

3.4.1. Study phase

In this part of the experiment all participants received a fabricated police case-file concerning a hypothetical murder. The case-file contained a description of the crime, which was exactly the same for all examinees, and a set of four pictures: two face pictures (i.e., face of victim 1 and face of victim 2) and two scene pictures (i.e., crime scene and location of murder weapon), which differed in arousal level and valence. One of the face pictures and one of the scene pictures were neutral and the other two pictures were either negative arousing or negative non-arousing, depending on the experimental condition. These pictures later served as crime-related items in the CIT and were chosen from five different sets of pictures, such that each set served as the crime-related set for 20% of the participants. The four other sets of pictures served as control items in the CIT. Experimenter 1 familiarized participants with the case-file and instructed them to envision the situation and to imagine that they themselves were the murderer. Then, all participants were requested to take a few minutes to carefully go over the case-file and memorize the pictures. When ready, participants assigned to the immediate conditions continued to the second part of the experiment, whereas participants in the delayed conditions returned after a week to take the CIT.

3.4.2. The CIT

In the second part of the experiment, Experimenter 2, who was unaware of the crime-related items, attached the SCR and HR electrodes as well as the RLL band and conducted the CIT examination. Before the CIT, it was explained to all participants that they were suspects in a murder and had to undergo a "polygraph"-examination. They were further motivated to hide their knowledge of the crime-related items (i.e., come out as innocent) and were promised a bonus of 10 NIS (about 3 USD) as an incentive for successful concealment. The bonus was paid when the average SCR Z score, computed across all crime-related items (i.e., SCR detection score), was below 0.1. Following an initial rest period of 2 min, participants were presented with the first block of four CIT questions, each targeting one of the pictures from the case-file. To maintain participants' attention, a short break was inserted after the first block. Following the break, participants were presented with the same four questions. Thus, the CIT consisted of 4 questions × 2 repetitions = 8 questions. Each question (e.g., How does the first victim look like?) was presented on the computer monitor for 10 s and at the same time played through the computer's loudspeakers. The order of question presentation was random in both blocks, except that two face questions (examining victim 1 and victim 2) and two scene questions (examining the crime scene and the location of the murder weapon) did not appear successively. Following question presentation, the different items appeared for 5 s each, with an inter-stimulus interval of 10–14 s (Breska, Maoz, & Ben-Shakhar, 2011). The first item was always a neutral, buffer item designed to absorb the initial orienting response. Next, 1 crime-related item, 4 control items and 1 catch item were presented in a random order. Catch items were included as an extra means of assuring that participants' attention remained focused on the items presented (see also Verschuere, Crombez, Degrootte, & Rosseel, 2010) and were pictures from the same question-category with a large number presented in the middle. When presented with a catch item, participants were requested to say the number out loud. In response to all other items, participants were requested to say "NO".

3.4.3. Subjective ratings

Following the CIT, a multiple-choice recognition memory test was administered. The memory test consisted of the same four questions used in the CIT, however all possible answer-items were presented simultaneously for each question (correct answers were coded as 1, incorrect answers were coded as 0). After completing the recognition test, participants were given two questionnaires (see Appendix B) and were further asked to rate their level of motivation during the experiment (on a scale of 1–6) and, although not instructed to, whether and what kind of countermeasures they applied. Finally, all participants were debriefed and compensated for their participation in the experiment.

3.5. Data acquisition and reduction

The experiment was conducted in an air-conditioned laboratory. The apparatus included a HP Compaq DC 5800 Microtower computer that was used to control stimulus presentation and compute skin conductance, respiration and heart rate.

Electrodermal activity was recorded using a constant voltage system (0.5 V ASR Atlas Researches, Hod Hasharon, Israel), two Ag/AgCl electrodes (0.8 cm diameter) filled with a 0.05 M NaCL electrolyte (TD-246, Discount Disposables) and an A/D (NB-MIO-12) converter with a sampling rate of 50 Hz. Electrodes were placed on the distal phalanges of the left index and left ring finger. The mean tonic skin conductance level (SCL) was obtained throughout both blocks of the CIT and SCRs were defined as the maximal increase in conductance obtained from 1 s to 5 s after stimulus onset.[2]

The electrocardiogram (ECG) was recorded by placing three Ag/AgCl electrodes filled with an electrode paste in a standard Einthoven lead I configuration: one electrode attached to the distal phalange of the left index finger (i.e., one of the SCR electrodes), one electrode attached to the right wrist and the ground electrode attached to the left wrist. The ECG signal was sampled at 500 Hz, digitized at 12-bit resolution and filtered using a band pass of 0.5–35 Hz. Matlab was used to detect the R peaks, calculate the distance between them and apply a semi-automatic artefact detection and rejection procedure (similar to e.g., De

[1] The pilot study included 31 stimuli, 25 of which were selected for the CIT. Three additional items that were not tested in the pilot study were used in the CIT.

[2] In Experiment 1 we mixed arousing and neutral pictures within each block of the CIT, while in Experiment 2 we used an arousing and a neutral question block. As tonic changes in the SCL occur more slowly than phasic changes in the SCR, the mixed format in Experiment 1 did not allow us to examine the SCL. We therefore used this measure only in Experiment 2.

Clercq, Verschuere, De Vlieger, & Crombez, 2006). Prior to analysis, the inter-beat intervals (IBI) were converted to HR in beats per minute (bpm) per real-time epoch (1 s). These second-by-second post-stimulus HR values were baseline-corrected by subtracting the average HR value in the third and second s preceding stimulus onset (i.e., the pre-stimulus baseline value), resulting in 13 post-stimulus difference scores (ΔHR).[3] The average of all ΔHR scores has been found to outperform the minimum of all ΔHR scores as a detection measure (Gamer, Verschuere, Crombez, & Vossel, 2008) and was therefore the preferred statistic when analyzing the data. This index perfectly captures the HR pattern, as is illustrated by Fig. 1 for Experiment 2.

Respiration was recorded using a pneumatic tube positioned around the thoracic area. Respiration responses were defined on the basis of the total RLL, which is a composite measure of respiratory amplitude (depth of breathing) and respiratory cycle (rate of breathing), during the 0.5-s to 13.5-s interval following stimulus onset. Following Elaad, Ginton, and Jungman (1992), we defined each response as the mean of ten length measures (0.1 s after stimulus onset through 13.1 s after stimulus onset, 0.2 s through 13.2 s after stimulus onset, etc.). In other words, ten 13-s windows were created, each beginning 0.1 s later than the previous window, and the RLL was defined as the mean of the ten length measures computed for the ten windows.

For all three measures, individual responses were removed if excessive movements were made during the measurement window or if the response was an outlier (Z score larger than 5 or smaller than -5). Further, similar to klein Selle, Verschuere, Kindt, Meijer, and Ben-Shakhar (2016, 2017) electrodermal non-responsivity was based on the within-participant standard deviation of the raw SCR scores. Participants whose standard deviation was below 0.01 μS in both blocks of the CIT were considered to be skin conductance non-responders and their SCR data were eliminated from all analyses. In case of non-responsivity in one of the blocks, only the SCR data from the respective block were removed.

Due to individual differences in physiological responsiveness and baseline activation, within-subject standard scores were calculated for each physiological channel separately (Ben-Shakhar, 1985). Further, considering that several physiological measures (especially SCR) are sensitive to habituation, the standard scores were computed for each item within a block of four questions. Thus, the within-subject standard scores were computed by subtracting the mean response of all control and crime-related items within one block of questions (buffer and catch items were excluded from the standardization) from each response to an individual item and dividing this difference by the respective standard deviation (see Ben-Shakhar & Elaad, 2002; Elaad & Ben-Shakhar, 1997). In the present study, each block contained a total of 20 pertinent items (4 crime-related and 16 control items) and consequently the Z scores were computed relative to the mean and standard deviation of the examinees' responses to these 20 items. For each participant and for each physiological measure a detection score was created by averaging the respective Z scores of all eight crime-related items.

A rejection region of $p < .05$ was used for all statistical tests and Cohen's f values values were computed as effect size estimates (Cohen, 1988). According to Cohen (1988), the values of $f = 0.1$, $f = 0.25$, and $f = 0.40$ correspond to small, medium, and large effects, respectively. One-tailed tests were used to test directional, a priori formulated hypotheses that were examined in our manipulation checks. For all main analyses we used two-tailed tests. Further, all analyses of the physiological data were supplemented by a Bayesian analysis and JZS Bayes factors (BFs) were computed. The JZS BF is a numerical value quantifying the odds ratio between the null and the alternative hypothesis given the data (Rouder, Speckman, Sun, Morey, & Iverson, 2009); a default JZS prior with scaling factor $r = 0.707$ was used for the alternative hypothesis. Importantly, the BFs are reported as either favouring the null or the alternative hypothesis and a BF of 3 or more is taken as substantial evidence for the respective hypothesis (Jeffreys, 1961). Finally, we evaluated CIT detection efficiency using the area under the Receiver Operating Characteristic (ROC) curve, which is now the standard detection efficiency measure used in CIT research (e.g., Green & Swets, 1966; National Research Council, 2003; Swets, Tanner, & Birdsall, 1961).

4. Results

4.1. Subjective ratings

A one-way ANOVA comparing the motivational level (scores ranging from 0 to 6) in the four between-subjects experimental conditions (arousing immediate vs. non-arousing immediate vs. arousing delayed vs. non-arousing delayed) yielded no statistically significant effect (F (3,128) = 1.00, $f = .00$, $p = .395$). In all conditions, participants reported high motivation during the experiment (Arousing immediate: $M = 5.29$, $SD = 0.86$; Non-arousing immediate: $M = 5.23$, $SD = 0.68$; Arousing delayed: $M = 5.03$, $SD = 1.05$; Non-arousing delayed: $M = 5.39$, $SD = 0.79$). Further, there was no significant difference between conditions in the number of subjects who reported using countermeasures ($\chi^2(3, N = 136) = 4.04$, $p = .258$).

4.2. Manipulation checks

4.2.1. Memory

To test whether the emotional value of the stimuli and a time delay affected recognition memory, a 3-way ANOVA (Time x Arousal x Valence) was performed on the memory data (scores ranging from 0 to 2). The ANOVA yielded no significant effects; in all conditions participants remembered the crime-related items almost perfectly, regardless of their arousal level or valence (neutral crime-related items − Arousing immediate: $M = 1.95$, $SD = .23$; Non-arousing immediate: $M = 1.97$, $SD = .18$; Arousing delayed: $M = 1.94$, $SD = .24$; Non-arousing delayed: $M = 1.97$, $SD = .17$; emotional crime-related items − Arousing immediate: $M = 2$, $SD = .00$; Non-arousing immediate: $M = 1.97$, $SD = .18$; Arousing delayed: $M = 1.97$, $SD = .17$; Non-arousing delayed: $M = 2$, $SD = .00$).

4.2.2. Physiology

All skin conductance data of 7 participants and the skin conductance data within the first block of 4 participants as well as the skin conductance data within the second block of an additional 21 participants were removed due to non-responsivity. Further, the skin conductance data of 6 participants and the heart rate data of 15 participants were lost due to technical issues (i.e., electrode failure). Thus, while the RLL data were based on 136 participants, the SCR and the HR data were based on 123 and 121 participants, respectively. For these participants, 2.5% of all SCRs, 1.1% of all RLL and 3.1% of all HR responses to the individual stimuli were removed due to excessive movements and outliers. A power analysis was conducted to determine the required sample size for detecting a difference between the responses to neutral and emotional (either negative arousing or negative non-arousing) stimuli; the analysis indicated that the sample size should be 34 in order to detect a medium effect size (i.e., Cohen's d of 0.50) with a statistical power of at least 0.8. The sample size in both the arousing ($n_{SCR} = 63$, $n_{RLL} = 72$, $n_{HR} = 63$) and non-arousing ($n_{SCR} = 60$, $n_{RLL} = 64$, $n_{HR} = 58$) conditions exceeded this number by far.

Enhanced SCRs and larger HR deceleration are expected to negative arousing than neutral pictures, regardless of whether they are crime-related or control items (for a review see Bradley, 2009). Hence, in order to verify our stimulus-induced arousal manipulation we conducted a 2-way ANOVA (Item Type: crime-related versus control x Arousal: negative arousing versus neutral) on the raw SCR as well as on

[3] The second before and the second after stimulus presentation as well as the fifteenth second after stimulus presentation were discarded due to measurement problems.

Table 1
Standardized Means and SDs of the Physiological Responses to Negative and Neutral Crime-related Items in the Four Between-subjects Conditions of Experiment 1.

Measure	*N*	Arousal	Time	Mean *Z* score (*SD*) crime-related items	
				Negative	Neutral
SCR	32	Arousing	Immediate	0.54 (0.81)	0.79 (0.67)
	30	Arousing	Delayed	0.86 (0.77)	0.75 (0.75)
	31	Non-arousing	Immediate	0.84 (0.79)	0.74 (0.79)
	30	Non-arousing	Delayed	0.90 (0.73)	0.66 (0.62)
RLL	38	Arousing	Immediate	−0.58 (0.54)	−0.40 (0.60)
	30	Arousing	Delayed	−0.28 (0.57)	−0.30 (0.53)
	34	Non-arousing	Immediate	−0.28 (0.50)	−0.40 (0.48)
	34	Non-arousing	Delayed	−0.37 (0.66)	−0.41 (0.57)
HR	31	Arousing	Immediate	−0.41 (0.61)	−0.22 (0.52)
	26	Arousing	Delayed	−0.13 (0.44)	−0.34 (0.59)
	31	Non-arousing	Immediate	−0.45 (0.58)	−0.41 (0.54)
	32	Non-arousing	Delayed	−0.27 (0.58)	−0.28 (0.57)

the raw ΔHR in the arousing conditions. Both ANOVAs revealed a significant main effect of item type, reflecting larger SCRs and decelerative HR responses to crime-related than control stimuli; SCR: $F(1,62) = 42.43, f = .83, p < .001$; HR: $F(1,62) = 32.88, f = .73, p < .001$. All other effects failed to reach significance. In order to assess whether the null results for the main effect of arousal are supported by the data we computed BFs. Values of 10.93 and 12.94 in favor of the null hypothesis were found for the SCR and the HR, respectively.

4.3. Main analysis

Means and standard deviations of the detection scores were computed for each condition and each physiological measure and are displayed in Table 1. An inspection of Table 1 shows that in all cases the detection scores were considerably larger than 0, meaning that the crime-related items consistently led to enhanced responses (larger SCRs, shorter RLLs and greater HR deceleration) as compared to the control items.

A mixed 3-way ANOVA (Time × Arousal × Valence) conducted on the SCR detection scores revealed no statistically significant effects (all F values < 1.75). In order to assess whether the null results for the main effects of arousal and valence are supported by the data, we computed BFs. Values of 6.24 and 2.97 in favour of the null hypothesis were found for the arousal and valence factors, respectively. Thus, there is substantial evidence for the null hypotheses with the SCR measure.

A mixed 3-way ANOVA (Time × Arousal × Valence) conducted on the RLL detection scores failed to reveal any significant effect (all F values < 2.51). Just as for the SCR, we computed BFs for the main effects of arousal and valence and obtained values of 4.70 and 4.36, respectively. Thus, there is substantial evidence for the null hypotheses with the RLL measure.

A similar mixed 3-way ANOVA (Time × Arousal × Valence) conducted on the HR detection scores also revealed no statistically significant effects (all F values < 2.91). Just as for the SCR and RLL, we computed BFs for the main effects of arousal and valence and obtained values of 7.21 and 6.89, respectively. Thus, there is substantial evidence for the null hypotheses with the HR measure.

For sake of completion, we also evaluated CIT detection efficiency using the area under the ROC curve (e.g., Green & Swets, 1966; Swets et al., 1961). A description of the ROC analysis and its results can be found in Appendix C.

5. Discussion

The present experiment aimed to examine whether emotional stimuli influence memory and, consequently, CIT detection efficiency. Unexpectedly, explicit memory performance was unaffected by both time and the emotional value of the stimuli (predictions 1 & 3 not confirmed). Similarly, although the crime-related items consistently induced larger SCRs, shorter RLLs, and greater HR deceleration than the control items, CIT detection efficiency remained stable over time (immediate vs. delayed) and neither negative valence nor arousal had an effect (predictions 2 & 3 not confirmed). Taken together, all three of our a-priori formulated predictions could not be confirmed.

The absence of the expected arousal effect on memory (as well as on CIT detection efficiency) may however be accounted for by a ceiling effect; in fact, we observed nearly perfect recognition for both neutral and emotional stimuli in all experimental conditions. This ceiling effect may, in hindsight, be explained by the fact that all stimuli were central to the hypothetical murder and participants were given ample time to memorize them. Several studies demonstrated robust memory of central crime-related items when the CIT was delayed; memory loss occurs mostly for the peripheral items (Gamer et al., 2010; Gronau, Elber, Satran, Breska, & Ben-Shakhar, 2015; Nahari & Ben-Shakhar, 2011; Peth et al., 2012). Moreover, it should be acknowledged that we found no support for the effectiveness of our arousal manipulation: The arousing stimuli were not successful in increasing phasic physiological reactivity. Consequently, this may raise the question of whether our stimuli would have been arousing enough to influence memory and detection efficiency if no ceiling effect was at play. Although the IAPS is a widely used and validated stimulus set (Lang, Öhman, & Vaitl, 1988), the database is somewhat outdated, and has been criticized for the low quality of its pictures (Marchewka, Zurawski, Jednoróg, & Grabowska, 2014).

In order to increase stimulus-induced arousal in Experiment 2, we selected arousing pictures from a more recent high quality picture database − the Nencki Affective Pictures System (NAPS) − in addition to the IAPS and KDEF. Moreover, to prevent ceiling effects for the explicit memory measure, we increased the number of crime-related stimuli, reduced encoding-time, and included a more sensitive memory test.

6. Experiment 2

Experiment 2 re-investigated the effect of emotional stimuli on the detection efficiency of the CIT by using a more recent picture database and by preventing a memory ceiling effect. In Experiment 2 we focused on the conditions where the effect was expected to be largest. As research has shown that arousal rather than valence influences memory, we decided to include neutral and negative arousing, but not negative non-arousing stimuli. Further, as the effect of arousal on memory has been shown to increase over time, we administered only a delayed CIT. Clearly, the delayed CIT is also more ecologically valid as in real life cases there is always some delay between crime and test. Finally, the arousal effect was maximized by using an arousing question-block (with only arousing pictures) and a neutral question-block (with only neutral pictures) in the CIT (see Brouwer, van Wouwe, Mühl, van Erp, & Toet, 2013). The design, hypotheses, and main analyses of Experiment 2 were preregistered on aspredicted.org: https://aspredicted.org/8rh8c.pdf.

7. Method

7.1. Participants

Thirty-nine undergraduate students (29 women) of the Hebrew University of Jerusalem (HUJI) with an age range of 19–30 ($M = 22.8$, $SD = 1.7$ years) participated in this experiment. All participants were native speakers of Hebrew and received either course credits or an average payment of 55 NIS (equivalent to approximately 14 USD) for their participation. Each participant read and signed a consent form indicating that participation was voluntary and that they could withdraw from the experiment at any time without penalty. The experiment was approved by the ethical committee of the Faculty of Social Sciences of the HUJI.

7.2. Items

All pictures were either negative arousing or neutral. These pictures were selected from three different databases: the IAPS and KDEF (as in Experiment 1) as well as the NAPS (see Appendix A Table A1). A total of 56 pictures (of which 7 were used in both Experiment 1 and Experiment 2) were selected. The normative arousal ratings were as follows: the arousing pictures had a mean rating of 6.93 (victim: $M = 6.86$, $SD = .74$; animal: $M = 7.13$, $SD = .37$; body part: $M = 6.98$, $SD = .67$; crime scene: $M = 6.73$, $SD = .42$; on a scale of 1–9) and the neutral pictures had a mean rating of 2.96 (male witness: $M = 2.57$, $SD = .24$; female witness: $M = 2.51$, $SD = .21$; street: $M = 4.11$, $SD = .20$; kitchen object: $M = 2.65$, $SD = .41$; on a scale of 1–9). An independent samples t test revealed a statistically significant difference between these arousal ratings, $t(54) = 22.85$, $p < .001$, with a very large effect size ($d = 6.11$).

7.3. Procedure

The procedure of Experiment 2 was similiar to that of Experiment 1, with a number of methodological changes. Each of these changes are specified below.

7.3.1. Study phase

In this phase of the experiment we increased the number of to-be-studied pictures and reduced the time that each picture was studied. Specifically, the number of studied pictures was increased from 4 to 8 (i.e., *arousing*: murder victim, crime scene, body part and animal; *neutral*: male witness, female witness, street of the crime scene and kitchen object). Further, while participants in Experiment 1 were given ample time to memorize the pictures, participants in Experiment 2 were shown each picture for 10 s on the computer monitor. This short presentation time was expected to weaken encoding. Further, in contrast to Experiment 1 where delay was manipulated, all participants returned after a week to take the CIT.

7.3.2. The CIT

In the CIT we used an arousing question-block and a neutral question-block; the arousing question-block focussed only on arousing pictures, while the neutral question-block focused only on neutral pictures (unlike in Experiment 1, where the arousing and neutral questions were mixed within a block). The presentation-order of these blocks was counterbalanced across participants. Furthermore, the inter-stimulus interval was increased to 14–18 s.

7.3.3. Subjective ratings

Following the CIT, we used a more sensitive memory test and also asked participants to provide significance, arousal and valence ratings. In the memory test, participants were presented, one by one, with all crime-relevant and 16 randomly chosen control pictures (two from each category). Participants were asked whether they remembered having seen the picture last week and rated their certainty level in remembering the picture on a scale ranging from 1 to 9. In addition, they rated each picture on significance, arousal and valence (on a scale of 1–9). Thus, significance, arousal and valence ratings were obtained from the participants themselves (unlike in Experiment 1, where these ratings were obtained from a pilot sample).

7.4. Data acquisition and reduction

The apparatus and measurement methods were the same as in Experiment 1 with two exceptions. The ECG signal was filtered using a band pass of 1–35 Hz and the second-by-second post-stimulus HR values were baseline corrected by subtracting the average HR value in the three seconds preceding stimulus onset (i.e., the pre-stimulus baseline value), resulting in 15 post-stimulus difference scores (ΔHR).

8. Results

8.1. Item ratings

The significance, arousal and valence ratings of both crime-related (arousing and neutral) and control (arousing and neutral) items obtained during the experiment were analyzed using a two-way repeated measures ANOVA (Item Type × Arousal). All three rating-types revealed a significant main effect of arousal (significance: $F(1,38) = 56.31$, $f = 1.22$, $p < .001$; arousal: $F(1,38) = 69.61$, $f = 1.35$, $p < .001$; valence: $F(1,38) = 252.32$, $f = 2.58$, $p < .001$), indicating that the negative arousing pictures were considered more negative, more arousing and more significant than the neutral pictures. Further, a significant main effect of item type was revealed for the significance and arousal ratings (significance: $F(1,38) = 77.28$, $f = 1.43$, $p < .001$; arousal: $F(1,38) = 53.78$, $f = 1.19$, $p < .001$), indicating that the crime-related pictures were considered more significant and arousing than the control items. Finally, the arousal ratings revealed a significant Item Type x Arousal interaction, $F(1,38) = 9.31$, $f = .50$, $p = .004$, indicating that the arousal difference between crime-related and control pictures was smaller when these pictures were negative arousing compared to when they were neutral. All other effects were not statistically significant.

8.2. Manipulation checks

8.2.1. Memory

To test for differences in memory between the various pictures, a 2-way repeated measures ANOVA with item type (crime-related versus control) and arousal (arousing versus neutral) as within-subject factors was performed on the memory data.[4] The ANOVA revealed a significant main effect of arousal, $F(1,38) = 39.18$, $f = 1.02$, $p < .001$, indicating higher memory ratings for negative arousing compared to neutral pictures. In addition, the Item Type × Arousal interaction was statistically significant, $F(1,38) = 11.64$, $f = .55$, $p = .002$. A follow-up paired sample t test revealed that the arousing-neutral difference in memory was larger for crime-related compared to control pictures, $t(38) = 3.41$, $p = .002$, $d = .73$. Nonetheless, for both crime-related and control pictures, higher memory ratings were observed for negative arousing than for neutral stimuli (crime-related: $t(38) = 5.82$, $p < .001$, $d = 1.13$, control: $t(38) = 3.07$, $p = .002$, $d = .37$). These results were strengthened by BFs favouring the alternative: a BF of 34081.26 was found for the arousing vs. neutral crime-related comparison and a BF of 18.16 was found for the arousing vs. neutral control comparison. Thus, participants were more certain when rating their memory for negative arousing pictures, particularly when the pictures were crime-related, but also when they were controls.

8.2.2. Physiology

The skin conductance data of 2 participants within the first block of the CIT as well as the skin conductance data of 2 additional participants within the second block of the CIT were removed due to non-responsivity. Further, the skin conductance data of 1 participant, the respiration data of 1 participant and the heart rate data of 1 other participant were lost due to either measurement errors or equipment failure. Thus, while the RLL and HR analyses are based on data of 38 participants, the SCR analysis is based on data of 34 participants. As noted in Experiment 1, a power-analysis had shown that the sample size should be 34 in order to detect a medium effect size (i.e., Cohen's d of 0.50) with a statistical power of at least 0.8. For the remaining participants, 1.6% of all SCRs, 0.8% of all RLL and 2.5% of all HR responses to the individual stimuli were removed due to excessive movements and outliers.

[4] Because better memory of control items is reflected by lower ratings, the scale was reversed for these items before analysis.

Table 2
Standardized Means and SDs of the Physiological Responses to Negative Arousing and Neutral Crime-related Items in Experiment 2.

Measure	*N*	Mean *Z* score (*SD*) crime-related items	
		Negative arousing	Neutral
SCR	34	0.73 (0.58)	0.45 (0.48)
RLL	38	−0.44 (0.46)	−0.41 (0.58)
HR	38	−0.26 (0.49)	−0.35 (0.52)

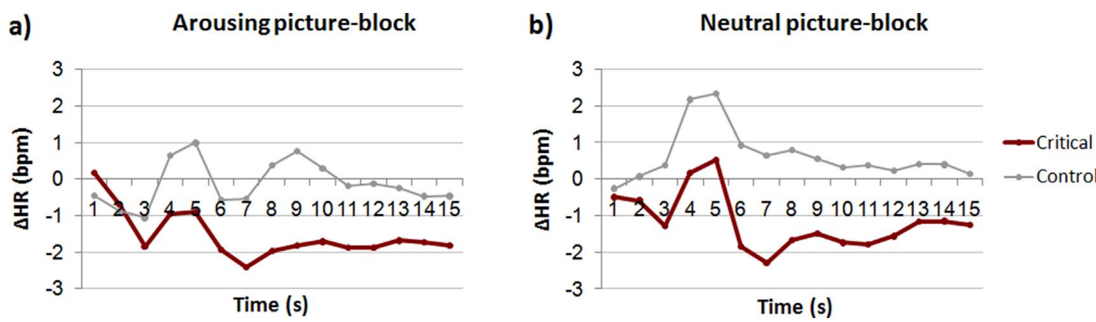

Fig. 1. HR-change to Critical and Control Items in the Arousing Picture-block (a) and the Neutral Picture-block (b).

The use of a blocked design allowed us to check the manipulation by measuring tonic arousal (SCL) in addition to phasic arousal (SCR, ΔHR). Further, as the raw physiological responses were expected to be affected by the interaction between arousal and block order (i.e., faster habituation is expected when the first block is arousing than when the first block is neutral), we added block order as a factor. Specifically, we performed three 2-way mixed ANOVAs, with arousal (arousing versus neutral) as a within-subject factor and block order (arousing first versus neutral first) as a between-subjects factor on the raw SCR, raw ΔHR[5] and the SCL. A significant main effect of arousal was observed for the SCL and SCR and a trend towards significance was observed for the HR (SCL: $F(1,36) = 3.60$, $f = .32$, $p = .033$; SCR: $F(1,32) = 4.69$, $f = .38$, $p = .019$; HR: $F(1,36) = 2.75$, $f = .28$, $p = .053$). These effects reflect higher physiological arousal in the arousing picture-block compared to the neutral picture-block. It should however be mentioned that the computed BFs provide only moderate evidence for the alternative hypothesis: a BF of 1.37 was found for the SCL, a BF of 1.85 was found for the SCR and a BF of 1.28 was found for the HR, all in favour of the alternative. Further, for both the SCL and SCR, a significant Arousal x Block Order interaction-effect was observed (SCL: $F(1,36) = 16.84$, $f = .68$, $p < .001$; SCR: $F(1,32) = 9.57$, $f = .55$, $p = .004$). Follow-up paired sample *t* tests revealed that the difference in raw SCR and SCL between blocks was significantly larger for participants who started with an arousing block compared to those who started with a neutral block, SCL: $t(36) = 1.90$, $p = .033$, $d = .62$; SCR: $t(32) = 2.17$, $p = .019$, $d = .74$. Thus, both tonic and phasic skin conductance remained relatively stable when the second block in the CIT was arousing, but decreased when the second block was neutral. These results were strengthened by BFs: a value of 2.40 was found for the SCL and a value of 3.67 was found for the SCR, both in favour of the alternative. All other effects failed to reach significance.

8.3. Main analysis

Means and standard deviations of the detection scores were computed for each condition and each physiological measure and are displayed in Table 2. An inspection of Table 2 shows that in all cases the detection scores were considerably larger than 0, meaning that the crime-related items consistently led to enhanced responses (larger SCRs, shorter RLLs and greater HR deceleration) as compared to the control items.

As both time and valence were not manipulated in this experiment, we now analyzed the standardized SCR, RLL and HR responses using paired sample *t* tests comparing the arousing and neutral picture-blocks. These *t* tests revealed significantly larger standardized SCRs in the arousing compared to the neutral picture-block ($t(33) = 2.71$, $p = .011$, $d = .52$), but no difference in the standardized RLL and HR responses between picture-blocks (RLL: $t(37) = -.31$, $p = .758$, $d = -.06$; HR: $t(37) = .96$, $p = .341$, $d = .19$). This pattern of HR results is also reflected in Fig. 1, showing the second-by-second HR changes. The above results were further strengthened by BFs: a BF of 4.11 in favour of the alternative was found for the SCR, while BFs of 5.47 and 3.72 in favour of the null were found for the RLL and HR, respectively. Thus, while there is substantial support for the alternative hypothesis with the SCR, there is substantial support for the null hypothesis with the RLL and HR measures.

For sake of completion, we also evaluated CIT detection efficiency using the area under the ROC curve (e.g., Green & Swets, 1966; Swets et al., 1961). A description of the ROC analysis and its results can be found in Appendix C.

9. Discussion

Experiment 2 re-examined the question of whether emotional stimuli influence CIT detection efficiency by increasing stimulus arousal and by preventing a memory ceiling effect. Both these methodological changes had the expected effect. First, in contrast to Experiment 1, memory was better for the emotional pictures than for the neutral pictures (prediction 1 is confirmed). Second, we found physiological evidence for the success of our arousal manipulation. The SCL, raw SCR and raw ΔHR responses were larger in the arousing compared to the neutral picture-block. After ensuring the success of our arousal manipulation, we returned to our main question and examined whether CIT detection efficiency was affected by the emotional value of the stimuli. Interestingly, an emotional heightening effect was observed only for the SCR. Thus, while SCR detection efficiency was stronger for

[5] As the expected increase in tonic HR would be influenced by the phasic decrease in HR, we decided to analyse only the phasic response of this measure.

emotional compared to neutral stimuli, RLL and HR detection efficiency were similar for emotional and neutral stimuli (prediction 2 is only partly confirmed). Taken together, these findings indicate that emotional stimuli are better remembered and better detected when relying on the SCR (but not on the RLL and HR) measure.

10. General discussion

In two separate experiments we examined whether CIT detection efficiency is affected by the emotional value of the stimuli. In both experiments, participants were requested to imagine that they are suspects in a murder and were familiarized with a number of crime-related pictures that differed in arousal level and valence. Experiment 1 revealed that neither arousal/valence nor a time delay impacts upon the CIT effect. Physiological and memory data, however, indicated that the arousal manipulation was unsuccessful. Experiment 2 therefore aimed to increase stimulus-arousal (by using more arousing pictures and a blocked design) and prevent a ceiling effect of memory (by increasing the number of studied stimuli, reducing encoding time and including a more sensitive memory test). Indeed, both the physiological and memory data of Experiment 2 showed that the arousal manipulation exerted its intended effect. Moreover, CIT detection efficiency was heightened for emotional compared to neutral stimuli when relying on the SCR, but not when relying on either the RLL or HR measures.

10.1. Emotional stimuli and memory

Both the emotional and neutral crime-related stimuli were remembered almost perfectly in Experiment 1. This ceiling effect may be explained by the low demand of the memory task; participants were given ample time to study four stimuli. In Experiment 2 on the other hand, participants studied eight different stimuli, each for only 10 s. Consequently, while memory of the emotional stimuli remained high, memory of the neutral stimuli was significantly reduced. This observed memory pattern resembles that of central and peripheral items in the CIT. Central items' resistance to forgetting has been associated with a more elaborated encoding which in turn increases the depth of processing (Gamer et al., 2010). Similarly, emotional stimuli have been shown to influence encoding, storage and retrieval. Clearly, in real-life, much longer time lags between crime and interrogation are common and in such cases memory may be reduced for all stimulus-types. However, as both the present study and previous memory studies (e.g., Bradley et al., 1992; Sharot & Phelps, 2004) indicate that this memory-deterioration may be smaller for emotional compared to neutral stimuli, emotional stimuli may be preferable in real-life cases.

10.2. Emotional stimuli and physiology

Unlike the emotional stimuli used in Experiment 2 (taken from the IAPS/NAPS), the emotional stimuli used in Experiment 1 (all taken from the IAPS) did not succeed to increase physiological activity. This was surprising as the IAPS (Lang et al., 1988) is a widely used and well-validated stimulus set in the experimental study of emotional processes by laboratories across the globe. Each picture in this database was rated by numerous participants on measures of valence, arousal and dominance. These ratings have been found to be associated with objective measures of physiological arousal such as heart rate and skin conductance (Bradley, Codispoti, Cuthbert, & Lang, 2001; Lang, Greenwald, Bradley, & Hamm, 1993; for a review of research on the IAPS see Bradley & Lang, 2007). One explanation for the manipulation check failure in Experiment 1 may be that the IAPS database is somewhat outdated, as is reflected by the low quality of its pictures (see Marchewka et al., 2014). Alternatively, the selected IAPS stimuli may have not been arousing enough for the present Israeli student population. As far as we know, there are no existing Israeli norms for the IAPS. There is however one study showing that Israeli students react differently (i.e., give higher valence ratings) to the IAPS pictures than the American population on which the norms are based (Okon-Singer, Kofman, Tzelgov, & Henik, 2011). In the pilot of Experiment 1 we also found differences between the normative ratings and Israeli ratings: the Israeli pilot sample gave lower arousal ratings than the normative American sample. Okon-Singer et al. (2011) suggested that this difference may be related to the compulsory military service in Israel, during which exposure to traumatic events is more likely to occur.

10.3. Emotional stimuli and CIT detection efficiency

Experiment 2 revealed that the usage of emotional stimuli in the CIT affects the skin conductance but not the cardiorespiratory measures. A similar fractionation of responses has been observed in previous studies examining a variety of factors (e.g., interfering task: Ambach, Stark, & Vaitl, 2011, question repetition: Ben-Shakhar & Elaad, 2002, countermeasure attempts: Ben-Shakhar & Dolev, 1996; Honts et al., 1996). This fractionation aligns with the hypothesis that different mechanisms drive the various response measures (klein Selle et al., 2016, 2017). While the SCR CIT-effect may primarily reflect an orienting mechanism, the RLL and HR seem more responsive to arousal inhibition. In line with this reasoning, klein Selle et al. (2017) found that stimulus salience (high versus low salient stimuli) influences the SCR (orienting factor), but not the RLL and HR (inhibition factor). Similarly, in the present study, stimulus emotion affected the SCR, but not the RLL and HR. The higher salience (i.e., significance) of the emotional stimuli might explain these parallel findings. Specifically, although emotion likely heightened the salience of both control and crime-related items, this increase might have been larger for the "emotional crime-related" compared to the "emotional control" stimuli. An alternative explanation in terms of memory (i.e., emotion increased memory and hence CIT detection efficiency), as we initially predicted, seems less likely as memory would be expected to affect all physiological measures, not only the SCR.

10.4. Applied implications

In its forensic application, the CIT aims to assess recognition of crime-related facts that remained hidden from the media and the public. It is reasonable to assume that the more serious the crime, the higher the chance that the available facts are emotional in nature. Although such emotional stimuli are typically ignored in the laboratory, they are actually preferred by CIT practitioners (Osugi, 2011). The present findings support this practitioners' intuition and provide preliminary evidence suggesting that emotional stimuli do not deteriorate and may in fact improve CIT detection efficiency. Importantly, the present results provide an answer to concerns expressed by several researchers (e.g., Ben-Shakhar, 2012; Ben-Shakhar & Furedy, 1990) that laboratory results cannot be generalized to the field because of differences in arousal. In fact, the findings of Experiment 2 show that as far as the emotional value of the stimuli is concerned, results of laboratory studies represent a lower bound of the real accuracy. Still, as this is only the first CIT study to manipulate stimulus arousal and valence, replication studies are crucial to examine the robustness of the current observations.

10.5. Limitations and future directions

Both physiological and recognition data showed that participants in Experiment 2 were aroused by our emotional stimuli. It bears mentioning, however, that even though the emotional stimuli were arousing, they were just pictures. Moreover, real-life stimuli may not be inherently arousing, but rather become arousing through the crime (e.g., a rope used in a murder). Care must therefore be taken when generalizing these results of passive perception to real-life situations. Future studies could resolve this issue by manipulating stimulus emotion through different activities. Osugi and Ohira (2005), for example, instructed their participants to either stab a pillow or a doll.

Second, all arousing stimuli in the present experiments were negative, while crime-relevant stimuli may in certain cases also be positive (e.g., the stolen object in a theft may be positive to the perpetrator). Hence, future studies should also clarify the effect of such positively arousing stimuli on detection efficiency. However, as memory is mostly affected by arousal, rather than valence, similar results could be expected (see Kensinger, 2009).

11. Conclusions

Taken together, our findings indicate that the emotionality of CIT stimuli does not weaken and may even strengthen CIT validity with the SCR. These results carry an important practical importance and foster our understanding of how arousal, memory and CIT detection efficiency interact. Our results are reassuring in demonstrating that, as far as stimulus emotion is concerned, laboratory experiments can be generalized to real-life situations and provide us with the lower bound accuracy estimates.

Author note

This research was funded by a grant, No. 238/15, from the Israel Science Foundation to Gershon Ben-Shakhar. The original data and analysis files are publically available on the Open Science Framework: https://osf.io/qxe58/. We wish to thank Adi Dan, Aya Navot, Noa Feldman, Shani Vaknine, Eli Rosner and Naama Agari for their assistance in data collection.

Appendix A

Table A1
Image-numbers from all stimuli used in Experiment 1 and Experiment 2.

Study	Emotion	Image–numbers
Experiment 1	Negative arousing	IAPS_3005.1, IAPS_3010, IAPS_3030, IAPS_3051, IAPS_3053, IAPS_3059, IAPS_3060, IAPS_3130, IAPS_6020, IAPS_9302, IAPS_9480, IAPS_9623, IAPS_9904, IAPS_9909
	Negative non-arousing	IAPS_6000, IAPS_7079, IAPS_7520, IAPS_9000, IAPS_9295, IAPS_9360, IAPS_9469; KDEF_AM04ANS, KDEF_AM06ANS, KDEF_AM07ANS, KDEF_AM25ANS, KDEF_AM26ANS, KDEF_AM29ANS, KDEF_AM30ANS
	Neutral	IAPS_7033, IAPS_7036, IAPS_7491, IAPS_7504, IAPS_7547, IAPS_7710, IAPS_9468; KDEF_AM08NES, KDEF_AM10NES, KDEF_AM14NES, KDEF_AM23NES, KDEF_AM28NES KDEF_AM31NES, KDEF_AM34NES
Experiment 2	Negative arousing	IAPS_3030, IAPS_3060, IAPS_3062, IAPS_3150, IAPS_3168, IAPS_3213, IAPS_3261; NAPS_Animals_016_h, NAPS_Animals_039_h, NAPS_Animals_056_h, NAPS_Animals_063_h, NAPS_Animals_073_h, NAPS_Animals_075_h, NAPS_Animals_077_h, NAPS_Faces_364_v, NAPS_Faces_365_v, NAPS_Faces_367_h, NAPS_Landscapes_177_h, NAPS_Objects_001_h, NAPS_Objects_013_h, NAPS_Objects_121_v, NAPS_Objects_126_h, NAPS_Objects_132_h, NAPS_Objects_149_h, NAPS_People_198_h, NAPS_People_216_h, NAPS_People_220_h, NAPS_People_221_h,
	Neutral	IAPS_7000, IAPS_7002, IAPS_7004, IAPS_7006, IAPS_7009, IAPS_7233, IAPS_7235; KDEF_AF01NES, KDEF_AF06NES, KDEF_AF09NES, KDEF_AF13NES, KDEF_AF18NES, KDEF_AF19NES, KDEF_AF29NES, KDEF_AM08NES, KDEF_AM10NES, KDEF_AM13NES, KDEF_AM14NES, KDEF_AM28NES, KDEF_AM30NES, KDEF_AM34NES; NAPS_Landscapes_016_h, NAPS_Landscapes_024_v, NAPS_Landscapes_037_v, NAPS_Landscapes_042_h, NAPS_Landscapes_071_h, NAPS_Landscapes_072_h, NAPS_Landscapes_083_h

Note: To indicate the database the images stem from, we added KDEF_, IAPS_ or NAPS_ as a prefix.

Appendix B

After completing the recognition test, participants were given two questionnaires: the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lusthene, 1970), and the Disgust Emotion Scale (DES; Walls & Kleinknecht, 1996). These questionnaires were included as the arousing pictures may have induced either anxiety (fear) or disgust in certain participants which in turn may have influenced their physiological responses and the CIT effect (Gross & Levenson, 1993; Lader, 1975; Vrana, 1993, 1994). However, as in both Experiment 1 and Experiment 2 no significant correlations were found between the STAI and DES scores and the physiological measures, we will not further report these results.

Appendix C

Adopted from Signal Detection Theory (SDT; e.g., Green & Swets, 1966; Swets et al., 1961), a standard measure used in memory detection research to evaluate CIT detection efficiency was computed (e.g., Ben-Shakhar & Elaad, 2003; National Research Council, 2003). This SDT measure, the area under the Receiver Operating Characteristic (ROC) curve, is typically derived by comparing the detection score distribution of knowledgeable (guilty) individuals with the detection score distribution of unknowledgeable (innocent) individuals. The area under the ROC curve describes the detection efficiency of the CIT across all possible cut-off points on the detection score and varies between 0 and 1, with a chance level of 0.5 (for a more detailed description of signal detection analysis as applied to the detection of concealed information, see Lieblich, Kugelmass, & Ben-Shakhar, 1970).

As the present experiment did not include a sample of unknowledgeable (innocent) participants, we used a simulation procedure to estimate their expected detection score distribution (see Meijer, Smulders, Johnston, & Merckelbach, 2007). The simulation procedure is based on the assumption that the individual crime-related items hold no special meaning for unknowledgeable participants and therefore there is no reason to expect that these items would elicit systematic differential responses. Since all responses are standardized within individuals, the expected responses of unknowledgeable participants to the individual crime-related items would have a mean of zero and a unit standard deviation. The average of these standardized responses (i.e., the detection score) would then have a mean of zero and a standard deviation of 1 divided by the square root of the number of crime-related items used for generating the detection score (here 8). As the simulation procedure further assumes that the detection scores are distributed normally, we created the hypothetical innocents distribution by taking random samples n (n equals the sample size of the relevant knowledgeable distribution) from a normal distribution; $N(0, 1/\sqrt{8})$. The simulated distribution was then compared with the empirical detection score distribution obtained for the knowledgeable participants and the area under the ROC curve was computed. This process was repeated 10.000 times for each physiological measure in each condition and the mean area, as well as the 95% confidence interval of the area across the 10.000 repetitions was computed.

Experiment 1. The empirical distributions in the arousing conditions were based on either neutral or negative arousing crime-related items and in the non-arousing conditions on either neutral or negative non-arousing crime-related items. Thus, a total of 8 areas under the ROC curves (4 conditions times 2) were computed per physiological measure. Detection efficiency with the SCR measure was well above chance in all conditions (i.e., the confidence intervals did not include a chance level of 0.5), with ROC areas ranging between 0.67 and 0.85. Detection efficiency with the RLL measure was highly similar to that of the SCR measure, with ROC areas ranging between 0.65 and 0.83. Finally, ROC areas with the HR measure varied between 0.58 and 0.74, with a non-significant ROC area for negative arousing stimuli in the arousing delayed condition. A comparison of the areas obtained for emotional (either negative arousing or negative non-arousing) and neutral items revealed no statistically significant difference for any of the physiological measures (all $Z < 1.60$).

Experiment 2. A total of 2 areas under the ROC curves, either based on the negative arousing or the neutral items, were computed per physiological measure. Detection efficiency with the SCR was well above chance with ROC areas of 0.84 and 0.75 for arousing and neutral items, respectively. Detection efficiency with the RLL was also well above chance with ROC areas of 0.76 and 0.70 for arousing and neutral items, respectively. Finally, detection efficiency of the HR measure was above chance with ROC areas of 0.68 and 0.71 for arousing and neutral items, respectively. A comparison of the areas obtained for emotional (negative arousing) and neutral items revealed no statistically significant difference for any of the physiological measures (all $Z < 1.30$).

References

Ambach, W., Stark, R., & Vaitl, D. (2011). An interfering n-back task facilitates the detection of concealed information with EDA but impedes it with cardiopulmonary physiology. *International Journal of Psychophysiology*, *80*, 217–226. http://dx.doi.org/10.1016/j.ijpsycho.2011.03.010.

Bahrick, L. E., Parker, J. F., Fivush, R., & Levitt, M. (1998). The effects of stress on young children's memory for a natural disaster. *Journal of Experimental Psychology: Applied*, *4*, 308–331. http://dx.doi.org/10.1037/1076-898X.4.4.308.

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*, 273–281. http://dx.doi.org/10.1037/0021-9010.81.3.273.

Ben-Shakhar, G., & Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination. *Journal of Applied Psychology*, *87*, 972–977. http://dx.doi.org/10.1037/0021-9010.87.5.972.

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*, 131–151. http://dx.doi.org/10.1037/00219010.88.1.131.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective.* New York, US: Springer Verlag Publishinghttp://dx.doi.org/10.1007/978-1-4612-3282-7.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*, 292–299. http://dx.doi.org/10.1111/j.1469-8986.1985.tb01603.x.

Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: An overview. *Frontiers in Psychology*, *3*, 342 [0.3389/fpsyg.2012.00342].

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and papillary measures. *Psychophysiology*, *18*, 307–315. http://dx.doi.org/10.1111/j.1469-8986.1981.tb03040.x.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford: Oxford University Press.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*, 379–390. http://dx.doi.org/10.1037/0278-7393.18.2.379.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and Motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*, 276–298. http://dx.doi.org/10.1037/1528-3542.1.3.276.

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, *46*, 1–11. http://dx.doi.org/10.1111/j.1469-8986.2008.00702.x.

Breska, A., Maoz, K., & Ben-Shakhar, G. (2011). Inter-stimulus intervals for skin conductance response measurement. *Psychophysiology*, *48*, 437–440. http://dx.doi.org/10.1111/j.1469-8986.2010.01084.x.

Brouwer, A., van Wouwe, N., Mühl, C., van Erp, J., & Toet, A. (2013). Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Frontiers in Human Neuroscience*, *7*, 1–10. http://dx.doi.org/10.3389/fnhum.2013.00295.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99. http://dx.doi.org/10.1016/0010-0277(77)90018-X.

Budson, A. E., Simons, J. S., Sullivan, A. L., Beier, J. S., Solomon, P. R., Scinto, L. F., Daffner, K. R., & Schacter, D. L. (2004). Memory and emotions for the September 11, 2001, terrorist attacks in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Neuropsychology*, *18*, 315–327.

Budson, A. E., Simons, J. S., Waring, J. D., Sullivan, A. L., Hussoin, T., & Schacter, D. L. (2007). Memory for the September 11, 2001, terrorist attacks one year later in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Cortex*, *43*, 875–888. http://dx.doi.org/10.1016/s0010-9452(08)70687-7.

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*, 261–269. http://dx.doi.org/10.1037/1076-898X.9.4.261.

Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

De Clercq, A., Verschuere, B., De Vlieger, P., & Crombez, G. (2006). Psychophysiological Analysis (PSPHA): A modular script-based program for analyzing psychophysiological data. *Behavior Research Methods*, *38*, 504–510. http://dx.doi.org/10.3758/BF03192805.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*, 183–201. http://dx.doi.org/10.1037/h0047707.

Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology*, *34*, 587–596. http://dx.doi.org/10.1111/j.1469-8986.1997.tb01745.x.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*, 757–767. http://dx.doi.org/10.1037/0021-9010.77.5.757.

Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology and Behavior*, *95*, 333–340. http://dx.doi.org/10.1016/j.physbeh.2008.06.011.

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

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

Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real life situation. *Journal of Applied Psychology*, *67*, 131–137. http://dx.doi.org/10.1037/0021-9010.67.2.131.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The karolinska directed emotional faces: A validation study. *Cognition and Emotion*, *22*, 1094–1118. http://dx.doi.org/10.1080/02699930701626582.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and Psychophysics.* New York, NY: John Wiley & Sons.

Gronau, N., Elber, L., Satran, S., Breska, A., & Ben-Shakhar, G. (2015). Retroactive memory interference. A potential countermeasure technique against psychophysiological knowledge detection methods. *Biological Psychology*, *106*, 68–78. http://dx.doi.org/10.1016/j.biopsycho.2015.02.002.

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, *64*, 970–986. http://dx.doi.org/10.1037//0022-3514.64.6.970.

Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, *33*, 84–92. http://dx.doi.org/10.1111/j.1469-8986.1996.tb02111.x.

Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology*Greenwich, CT: JAI Press [201–201.

James, W. (1890). *The principles of psychology*. New York: Henry Holthttp://dx.doi.org/10.1037/11059-000.

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

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*, 1169–1180. http://dx.doi.org/10.3758/BF03195800.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity. Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*, 575–591. http://dx.doi.org/10.1016/j.jml.2006.05.004.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*,

99–113. http://dx.doi.org/10.1177/1754073908100432.
Kleinknecht, R. A., Kleinknecht, E. E., Sawchuk, C. N., Lee, T. C., & Lohr, J. M. (1999). The medical fear survey: Psychometric properties. *The Behavior Therapist*, *22*, 109–119.
klein Selle, N., Verschuere, B., Kindt, M., Meijer, E. H., & Ben-Shakhar, G. (2016). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*, *53*, 579–590. http://dx.doi.org/10.1111/psyp.12583.
klein Selle, N., Verschuere, B. J., Kindt, M., Meijer, E. H., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the Concealed Information Test. *Psychophysiology*, *54*, 628–639. http://dx.doi.org/10.1111/psyp.12825.
Kugelmass, S. S., & Lieblich, I. (1966). Effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, *50*, 211–216. http://dx.doi.org/10.1037/h0023324.
Lader, M. H. (1975). The psychophysiology of anxious and depressed patients. In D. C. Fowles (Ed.), *Clinical applications of psychophysiology* (pp. 12–41). New York: Columbia University Press.
Lang, P., Öhman, A., & Vaitl, D. (1988). *The international affective picture system [photographic slides]*. Gainesville, FL: University of Florida, Centre for Research in Psychophysiology.
Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261–273. http://dx.doi.org/10.1111/j.1469-8986.1993.tb03352.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report A-8*. Gainesville, FL: University of Florida.
Lieblich, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR detection of information as a function of stimulus set size. *Psychophysiology*, *6*, 601–608. http://dx.doi.org/10.1111/j.1469-8986.1970.tb02249.x.
Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The karolinska directed emotional faces (KDEF)*. Stockholm: Department of Neurosciences Karolinska Hospital.
Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388. http://dx.doi.org/10.1037/h0046060.
Marchewka, A., Żurawski Ł, Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 596–610. http://dx.doi.org/10.3758/s13428-013-0379-1.
Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., & Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology*, *44*, 814–822. http://dx.doi.org/10.1111/j.1469-8986.2007.00543.x.
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*, 879–904. http://dx.doi.org/10.1111/psyp.12239.
Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, *53*, 593–604. http://dx.doi.org/10.1111/psyp.12609.
Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*, 733–744. http://dx.doi.org/10.1111/j.1469-8986.2010.01148.x.
Nahari, T., Breska, A., Elber, L., klein Selle, N., & Ben-Shakhar, G. (2017). The external validity of the Concealed Information Test: The effect of choosing to commit a mock-crime. *Applied Cognitive Psychology*, *31*, 81–90. http://dx.doi.org/10.1002/acp.3304.
National Research Council (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press, Committee to review the scientific evidence on the Polygraph, Division of Behavioral and Social Sciences and Education.
Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology*, *129*, 242–261. http://dx.doi.org/10.1037/0096-3445.129.2.242.
Okon-Singer, H., Kofman, O., Tzelgov, J., & Henik, A. (2011). Using international emotional picture sets in countries suffering from violence. *Journal of Traumatic Stress*, *24*, 239–242. http://dx.doi.org/10.1002/jts.20600.
Osugi, A., & Ohira, H. (2005). Effects of emotional arousal on lie detection: An ERP study. *Poster presented at the 45th annual meeting of the society for psychophysiological research*.
Osugi, A. (2011). Daily application of the concealed information test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: theory and application of the concealed information test* (pp. 253–275). Cambridge, UK: Cambridge University Press.
Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*, 381–390. http://dx.doi.org/10.1111/j.1469-8986.2011.01313.x.
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t–tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. http://dx.doi.org/10.3758/PBR.16.2.225.
Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 294–306. http://dx.doi.org/10.3758/CABN.4.3.294.
Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 389–394. http://dx.doi.org/10.1073/pnas.0609230103.
Spielberger, C. D., Gorsuch, R. L., & Lusthene, R. E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto: Consulting Psychologists Press.
Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. C. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340. http://dx.doi.org/10.1037/h0040547.
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*, 991–1002. http://dx.doi.org/10.1002/acp.1601.
Verschuere, B., Ben-Shakhar, G., & Meijer, E (Eds.). (2011). Memory detection: theory and application of the Concealed Information Test. Cambridge, UK: Cambridge University Press.
Verschuere, B., Meijer, E., & De Clercq, A. (2011b). Concealed information under stress. A test of the orienting theory in real-life police interrogations. *Legal and Criminological Psychology*, *16*, 348–356. http://dx.doi.org/10.1348/135532510X521755.
Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. *Psychophysiology*, *30*, 279–286. http://dx.doi.org/10.1111/j.1469-8986.1993.tb03354.x.
Vrana, S. R. (1994). *Startle reflex response* during sensory modality specific disgust, anger, and neutral imagery. *Journal of Psychophysiology*, *8*, 211–218.
Walls, M. M., & Kleinknecht, R. A. (1996). *Disgust factors as predictors of blood-injury fear and fainting*.