JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2018, Vol. 63, No. 5
doi: 10.1111/1556-4029.13718
Available online at: onlinelibrary.wiley.com

PAPER

PSYCHIATRY & BEHAVIORAL SCIENCE

Katja Volz,[1] M.Sc.; Lisa-Marie Bahr,[1] M.Sc.; Markus Heinrichs,[2] Ph.D.; Dieter Vaitl,[1] Ph.D.; and Wolfgang Ambach,[1] Ph.D.

The Concealed Information Test is Susceptible to Misleading Information*

ABSTRACT: An approach toward detecting hidden knowledge is the Concealed Information Test (CIT). It relies on the memory of crime-relevant information. This study investigated whether its validity is susceptible to memory distortion by misleading information. A misleading information paradigm was employed to distort memory prior to an interrogation with a CIT. Forty-one participants watched a video with specific crime-related information. After a 1-week retention interval, misleading information was introduced. Afterward, a CIT was performed, followed by a threefold memory test. When misleading information was presented, memory performance was reduced, and no physiological response differences between crime-relevant and crime-irrelevant information were found. Without presenting misleading information, physiological responses differed between responses to crime-relevant and crime-irrelevant information. However, responses in all physiological measures also differed between misleading and irrelevant information. The results indicate that the CIT is susceptible to misleading information, which reduces its validity in specific constellations.

KEYWORDS: forensic science, forensic anthropology, memory, concealed information test, false memory, eyewitness

The question of whether hidden knowledge in forensic investigations can be detected validly using physiological measures has been examined by various research groups for about 60 years (1). Routinely applied in Japan (2–7), the Concealed Information Test (CIT, 8) is a well-examined approach for detecting concealed knowledge (for reviews, see 1,9). It relies on physiological responses to specific crime-relevant knowledge that only a guilty suspect would recognize, making memory a key factor. As memory distortion is an extensively studied phenomenon in crime investigations, it is questionable whether the CIT remains valid if the suspects' memory of the crime is somehow distorted.

Detecting Concealed Knowledge Using Physiological Measures

In the Concealed Information Test (CIT, 8), physiological responses to one crime-relevant and several crime-irrelevant items of the same category are compared (e.g., "Was the money stolen out of a (a) red, (b) blue, (c) yellow, (d) black, or (e) brown envelope?"). It is assumed that only guilty subjects distinguish between crime-relevant and crime-irrelevant items. Guilty suspects' reactions to crime-relevant items are enhanced in comparison with crime-irrelevant items. The CIT is routinely used in forensic investigations in Japan (2–7) and is regarded as a highly valid technique of detecting concealed information (1,9,10). Typically, the CIT uses measures of differential peripheral physiological responses, such as phasic heart rate (pHR), skin conductance response (SCR), respiration line length (RLL), and finger pulse waveform length (FPWL), to detect concealed information (11). SCR yields the greatest effect sizes and is regarded as the most valid parameter in the CIT (1).

Usually, the enhanced physiological responses to crime-relevant items are explained as reflecting the orienting response (OR, 12). The OR is defined as the physiological, cognitive, and behavioral response to a given stimulus. It is modulated by its novelty, significance, and intensity. Stimulus significance is defined as the special importance and meaning a subject attributes to an item (13; see also 14) and is therefore the most relevant for detecting guilty knowledge. Recently, it was proposed that the reactions of different peripheral parameters in the CIT correlate with different subprocesses of it (3,15,16). It was demonstrated that SCR mainly relies on the orienting response, whereas pHR and RLL are closely associated with concealment-related processes (17,18), among which arousal inhibition was underpinned most recently (19). Yet, it is unknown whether parameters underlying different processes are distinctly susceptible to memory distortions.

Countermeasures and the Vulnerability of Memory as Limiting Factors

Meta-analyses have indicated that the CIT is a stable and valid method for detecting concealed information (1,9,10). Nevertheless, the method's susceptibility to countermeasures is an important issue that has been investigated extensively (for an overview, see 20) and is still being researched (21). Several

[1]Institute for Frontier Areas of Psychology and Mental Health, Wilhelmstraße 3a, D-79098 Freiburg, Germany.
[2]Laboratory for Biological and Personality Psychology, Department of Psychology, University of Freiburg, Stefan-Meier-Straße 8, D- 79104 Freiburg, Germany.

*Presented at the 41st Meeting "Psychologie und Gehirn" "Psychology and Brain", June 2015, in Frankfurt, Germany.

Received 6 July 2017; and in revised form 20 Nov. 2017; accepted 27 Nov. 2017.

studies found that physical (22–24) and mental (22–25) countermeasures reduced detection accuracy in CITs with peripheral measures. Interestingly, not every peripheral measure was affected in the same way: SCR effects were found to be reduced by physical and mental countermeasures, whereas effects of respiration were much less affected by both measures, and effects of heart rate were found to be reduced by mental countermeasures only (22,23). Two other studies also revealed attenuated CIT validity for SCR measures only (24,25). However, it is important to differentiate between countermeasures, which enhance reactions to neutral control items and countermeasures, which reduce reactions to crime-relevant items. Several studies using event-related potentials examined mental countermeasures that manipulated the memory of crime-relevant information, that is, reduced reactions to crime-relevant items: Detection accuracy in CITs was decreased when participants successfully suppressed memory-related activity to crime-relevant information (26,27) or when their memories were distorted (28). However, when few methodological details were altered (e.g., the CIT protocol; see 29,30), event-related potentials achieved high validity, even when mental countermeasures using memory suppression were applied. It can be further questioned whether a CIT that uses peripheral measures is compromised when the memory of a crime is distorted instead of just intentionally suppressed.

Yet, only few studies investigated the effect of memory distortions on the performance of a CIT using peripheral measures. For example, in one study (31), reactions in a CIT to true and false memories in a Deese-Roediger-McDermott (32,33) task were compared. Except for SCR, no differences in physiological responses associated with true memories compared to false memories were found. With respect to the applied context, these results imply that suspects with distorted memories react with enhanced physiological responses to both crime-relevant and falsely memorized irrelevant information. Thus, the validity of the CIT might be limited to situations with unaltered memory. However, in the described study (31), subjects did not gain any guilty knowledge by committing a mock crime or watching a crime-related video. It is therefore necessary to replicate these results in a typical CIT experiment in which subjects gain guilty knowledge before further implications can be drawn.

Another recent study investigated whether the validity of the CIT decreases when memory is distorted by retroactive interference (34): Participants first committed a mock crime. The experimental group then learned and rehearsed another mock crime. In the experimental group, SCR response differences were significantly reduced, whereas those in respiration volume remained stable. The authors concluded that retroactive interference has a major influence on SCR response differences. Nevertheless, the validity of the CIT using a combined measure of SCR and respiration remained high. The vulnerability of memory appears to be a limiting factor regarding the validity of the CIT, but such vulnerability does not affect all peripheral measures to the same extent. However, a question remains as to the extent that memories of a mock crime can be distorted by retroactive interference. Instead of using retroactive interference, we addressed this problem by implementing a misleading information paradigm, which is frequently used in applied forensic research to distort crime-related memory (35,36).

The misinformation paradigm is frequently used in eyewitness research and aims at distorting memory performance by postevent misleading information (for a review, see 35). In this paradigm, subjects first watch a video typically showing a crime. After a retention phase, misinformation hidden in a narrative text is introduced. Finally, subjects complete memory tasks about the event (35). Typically, memory performance for items in the video is decreased, whereas memory performance for misinformation hidden in the text is enhanced (35).

Aims of this Study

In practice, implementing a CIT shortly after a crime is committed is rarely possible. During the time interval between crime and interrogation with a CIT, misleading information can alter memory. We wanted to examine the susceptibility of the CIT to memory distortion by implementing a misinformation paradigm prior to the interrogation.

For this purpose, we combined a CIT with a misinformation paradigm. To make sure the misinformation paradigm and the CIT are combined in an ideal way to distort memory, some changes to the classical CIT procedure were made. A previous study examining misinformation effects revealed that a very deep encoding of original information weakens the effects of misinformation on memory (37). To avoid deep encoding, we abstained from employing a mock crime procedure and presented crime-relevant information in a video instead (see 18,24,38). Therefore, we interrogated the participants as informed innocents rather than as suspects of the crime. It was recently found that the memories of informed witnesses vanished faster than the memories of participants committing mock crimes (39). Because we were to investigate the validity of the CIT while crime-relevant memory was distorted, we laid more stress on a successful distortion of memory than on a deep encoding of crime-relevant information. After a 1-week retention phase, which is adequate to distort memory (39), and sufficient to produce a CIT effect (38–41), half of the crime-relevant information was replaced by misleading information in a narrative text.

Not all studies investigating the CIT or misinformation effects did ensure full randomization. To improve the stimulus materials used, we excluded effects of single items or categories using a differentially elaborated randomization procedure that will be described later on (see also 42).

Hypotheses

The CIT relies on memory of crime-relevant information. The main goal of this study was to investigate the validity of the CIT when memory is distorted. For this purpose, a misinformation paradigm was used to distort crime-relevant memory prior to an interrogation with the CIT. The hypotheses are as follows:

- We expect that the misleading information implemented for half of the crime-relevant information will (i) decrease memory performance of crime-relevant information and (ii) increase memory performance of misleading information in all three memory tests.
- We expect that physiological responses differ between crime-relevant and crime-irrelevant information when memory is not distorted (1,9). In the following, this difference is called a "CIT effect." It comprises enhanced skin conductance responses as well as more pronounced heart rate decelerations, smaller respiration line length, and smaller finger pulse waveform length after crime-relevant compared to irrelevant items.
- We expect that (i) the CIT effect decreases when misleading information was presented compared to when it is not; (ii) no CIT effect emerges when misleading information was presented; instead (iii) an effect similar to the CIT effect

emerges when misleading information was presented, that is, physiological responses to misleading information differ from those to irrelevant information.

Methods

Participants

Forty-one healthy student participants (21 f., 20 m., age: 22.59 ± 3.25 years.) from different departments (excluding psychology and neuroscience departments) were recruited via student services and university bulletins. Written consent was obtained prior to the experiment, which met all ethical requirements according to the Declaration of Helsinki. Sixteen Euros were paid for participation.

Materials and Procedure

The memory-related focus of the experiment was disguised using a cover story about moral sense. Each subject participated in two sessions 1 week apart. During the first session, which lasted about 20 min in total, a video introduced crime-relevant information. During the second session, which lasted about 60 min, a narrative describing the video plot implemented misleading information. Afterward, physiological reactions on crime-relevant, misleading and irrelevant (new) information were measured in a CIT. Finally, data from three different memory tests were collected. One experimenter guided the video and narrative; a second experimenter guided the CIT and the memory tests to separate the information acquisition from the testing sections of the experiment.

The crime-relevant, misleading, and neutral items were drawn from eight categories with five items each. Categories were *name on a door plate*, *picture of a landscape*, *box*, *fruit*, *key pendant*, *color of an envelope*, *drink,* and *playing card*. A list of all items used can be found in the Appendix 1.

During the first session, crime-relevant knowledge was introduced in a video lasting about 6.50 min. It showed a woman stealing a ring and money prior to a job interview. For each participant, the video was composed of a basic plot in which eight scenes were inserted. Each of these scenes was filmed five times with different items of the same category (e.g., the woman was filmed opening a red, blue, yellow, black, and brown envelope). The combination of items used in the video was selected using a randomization procedure. The participants watched the video twice. Hereby, it was intended to reduce the possibility participants missed an item. They were asked to divide it into meaningful sections and to rate every section on a scale of 0 (*morally not reprehensible*) to 100 (*morally reprehensible*). This task was used to assure that participants watched the video attentively.

The misleading information was introduced 1 week later. In a narrative describing the video plot, four of the eight crime-relevant items were replaced by misleading items of the same category (e.g., a brown envelope in the video was replaced by a "red envelope" in the text). The remaining four items were described neutrally (i.e., without repeating the particular item) by naming the category (e.g., an ice tea shown in the video was described as "drink" in the text). Subjects were asked to read the text carefully and work on three tasks: Dividing the text into meaningful sections, rating every section on a scale of 0 (*morally not reprehensible*) to 100 (*morally reprehensible*), and assigning a brief title to each section.

Afterward, the second experimenter conducted the CIT. Participants were led to an experimental chamber and connected to the polygraph leads. The CIT consisted of eight blocks referring to eight categories. Each of the four categories without misleading information comprised one crime-relevant and four crime-irrelevant (unknown) items. Each of the four categories with misleading information comprised one crime-relevant item, one misleading item, and three irrelevant items. The first item of a block served as a buffer item; the according trials were discarded from analysis. In total, 32 trials per participant were analyzed. In each block, a question appeared 2-sec prior to the presentation of a particular item (e.g., "Did she take the money out of this envelope?"). Items were presented for 10 seconds as text on a 19-inch monitor at a distance of 90 cm from the subject, followed by a jittered blank screen lasting for 8–10 sec. Participants were instructed to conceal their knowledge of items seen the video. It was made clear that all questions in the CIT refer to the video. For each item, they were asked to react as quickly as possible by pressing a response key ("no") and vocally responding with "no." Subjects were told that they would receive an additional reward of 3 Euros if they succeeded in hiding their knowledge from the experimenter. In the end, all subjects received the additional reward independent from their actual performance in the CIT.

After the CIT, participants were detached from the leads and asked to complete a threefold memory task truthfully (i.e., without concealing their knowledge). In a level-of-confidence rating, participants were asked to indicate on a scale from −3 (*definitely not seen*) to +3 (*definitely seen*) whether or not an item was part of the video. In the following source identification task, they were asked to decide whether the first presentation of an item had occurred in the video, the narrative, or the CIT. Finally, in the multiple-choice task, all five items of a category were presented together. Participants were asked to decide which one of the five items was part of the video.

Randomization procedures warranted that the role of each item as crime-relevant, misleading, or irrelevant was balanced across subjects. All categories were equally often used as categories with or without misleading information. For the CIT, the sequence of categories was balanced, as well as the item sequence within categories; particularly, the relative position of crime-relevant and misleading items was balanced. Randomization was also applied to item sequences in the concluding threefold memory test.

Design

Two factors were varied within subject: First, in half of the eight categories, misleading information was introduced. This resulted in four *control* and four *misled* categories. Second, each category comprised different types of items. Control categories comprised one *crime-relevant* and four *irrelevant* items. Misled categories comprised one *crime-relevant*, one *misleading*, and three *irrelevant* items.

Physiological Recording

Physiological data were recorded in a dimly lit, electrically, and acoustically shielded experimental chamber (*Industrial Acoustics GmBH*, Niederkrüchten, Germany). The temperature was maintained by air conditioning and was set to approximately 22.2°C at the beginning with a maximum increase of 1.3°C throughout the recording. The subjects sat in an upright position so they could easily reach the keyboard and watch the 19″-monitor.

Skin conductance, electrocardiogram, respiratory activity, and finger plethysmogram were measured. Physiological data were logged using the Physiological Data System I 410-BCS (*J&J Engineering*, Poulsbo, Washington) and converted from analog to digital at a resolution of 14 bits, allowing skin conductance to be measured with a resolution of 0.01 μS. Stimulus on/offsets and physiological data were sampled at a rate of 510 Hz. Standard ag/AgCl electrodes (*Hellige*; diameter 0.8 cm), electrode paste of 0.5% saline in a neutral base (*TD 246 Skin Resistance, Mansfield R&D*, St. Albans, Vermont, UK), and a constant voltage of 0.5 V were used while recording skin conductance. Electrodes were placed over the thenar and hypothenar muscles of the nondominant hand. Electrocardiogram was measured using *Hellige* electrodes (diameter 1.3 cm) according to Einthoven II. The thoracic and abdominal respiratory activity was registered by two PS-2 biofeedback respiration sensor belts (*KarmaMatters*, Berkeley, CA) with a built-in length-dependent electrical resistance. An infrared pulse sensor in a cuff around the end phalanx of the middle finger of the nondominant hand was used to record the finger plethysmogram.

Data Analysis

Skin conductance data (SCR) from one subject had to be discarded from analysis because of electrodermal nonresponse (more than 85% of responses were smaller than 0.01 μS). Responses in skin conductance were defined as increases in conductance that were initiated within a time period of one to 5-sec after image onset. The amplitude of the responses was automatically evaluated as the difference between response onset and the subsequent maximal value in the set time window (43). Phasic heart rate (pHR) data were notch-filtered at 50 Hz and underwent an automatic R-wave peak detection prior to visual inspection of the resulting data. The R–R intervals were transformed into heart rate and real-time scaled (44). The heart rate during the last second before trial onset served as the prestimulus baseline. The pHR was calculated by subtracting this value from each second-per-second poststimulus value. For extracting the trial-wise information of the phasic heart rate, the mean change in heart rate within 15 sec after trial onset compared with the prestimulus baseline was calculated (45,46). Respiratory data (RLL) were manually scanned and low-pass filtered for eliminating artifacts. RLL was computed over a time interval of 15 sec after trial onset (47,48). This integrates information on the frequency and depth of respiration. RLL data from both bolts were averaged. The finger pulse wave form length (FPWL) within the first 15s after trial onset was calculated from the finger pulse waveform and then subjected to further analyses (49). It comprises information about both heart rate and pulse amplitude. In all analyses, within-subject z-standardized physiological data were used (50–52).

Statistical analysis was conducted using SPSS 23 (*IBM Corp.*, Armonk, NY). Two subjects were discarded from analysis due to problems of understanding the instructions of the CIT. A combined measure was calculated by adding up the absolute value of z-scores of SCR, pHR, and RLL. For each hypothesis, a t-test for paired samples was conducted separately for each standardized physiological measure and the combined measure. Normal distribution was checked for all variables. If a variable was not distributed normally, a Wilcoxon test for paired samples was used instead of a *t*-test. For Wilcoxon tests, Cohen's *d* (52) was calculated using the formulas $r = z/\sqrt{N}$ (53) and $d = \frac{2r}{\sqrt{(1-r)^2}}$ (54). All tests were conducted with an alpha level of 0.05.

Binary logistic regression analyses were performed with inclusion of each of the measures, with an equal weights combination of EDA, pHR, and RLL, and with a fixed inclusion of these three, or all four measures in a binary logistic regression analysis (which, in contrast to a stepwise inclusion, prohibits the included measures from differing between groups). Because our study was comprised only of knowledgeable participants, data for an equal number of innocents were simulated according to Meijer and colleagues (55). Response values for innocent subjects were drawn randomly from a normal distribution; these data were then treated in the same way as the data from real participants. Subject classification statistics were calculated for a classification of guilty and (hypothetical) innocent participants. Besides correct classification rates based on a fixed criterion of 0.5 (according to the percentage of guilty participants), ROC areas for the classification of guilty and innocent subjects were computed by varying the criterion from 0 to 1 (see 56). This allows an estimation of the capability to differentiate guilty from innocent participants for all possible cutoff points and for different dependent measures and their combinations. The area under the ROC curve serves as an overall index of detection accuracy (9,57). Confidence intervals for these ROC areas were computed by bootstrapping. As shrinkage correction, a hold-two-out procedure was performed, similar to the more common hold-one-out procedure, which tends to bias results when unequal group sizes and bad predictors come together. Pairs of subjects (one guilty plus one innocent) were held out of the dataset and then classified on the basis of beta weights obtained with the remaining subjects.

Results

Manipulation Check

First, it was checked whether misleading information reduced explicit memory performance. Descriptive statistics of memory performance in the threefold memory task are summarized in Table 1.

Memory performance of crime-relevant items in misled compared to control categories was lower for all three memory tests: Level of confidence, $z = 2.78$, $p < .005$, $d = 0.61$, source identification, $z = 2.86$, $p < .005$, $d = 0.63$, and multiple choice, $z = 2.12$, $p < .05$, $d = 0.51$. In misled categories, memory for misleading items was enhanced in comparison with control items: Level of confidence, $z = 4.73$, $p < .001$, $d = 0.82$, and multiple-choice task, $z = 3.95$, $p < .001$, $d = 0.77$. That is, a misinformation effect occurred in all three memory tasks. Nevertheless, in misleading categories, memory for misleading items was lower than for crime-relevant items: Level of confidence, $z = -1.85$, $p < .05$, $d = 0.46$, source identification, $z = 3.13$, $p < .005$, $d = 0.67$, and multiple-choice task, $z = 3.25$, $p < .001$, $d = 0.66$.

Physiological CIT Effects

Means and standard deviations of physiological responses to crime-relevant, crime-irrelevant, and misleading items of control and misled categories are presented in Table 2 and are based on raw scores of each data channel.

In the following analyses, differences in physiological responses to crime-relevant and crime-irrelevant items were

TABLE 1—*Means and standard deviations (SD) of each memory test. Responses to crime-relevant, crime-irrelevant, and misleading items are listed separately for categories with and without misleading information.*

	Control Categories				Misled categories					
	Crime-Relevant Items		Crime-Irrelevant Items		Crime-Relevant Items		Misleading Items		Irrelevant Items	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Level of confidence*	4.86	1.01	1.93	0.65	4.29	1.22	3.58	1.38	2.01	0.79
Source identification†	55.13	23.77	87.98	10.96	42.31	25.75	26.92	27.14	89.96	10.85
Multiple choice‡	61.93	23.8	38.07	23.8	52.27	27.92	26.14	23.47	21.59	21.96

*Scale from 1 (“definitely not seen”) to 7 (“definitely seen”).
†Percentage of correctly attributed items.
‡Percentage of items picked as “video items”.

TABLE 2—*Means and standard deviations (SD) of within-subject standardized z-scores for each data channel. Responses to crime-relevant, crime-irrelevant, and misleading items are listed separately for each category without and with misleading information.*

	Control Categories				Misled Categories					
	Crime-relevant Items		Crime-irrelevant Items		Crime-relevant Items		Misleading Items		Irrelevant Items	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SCR	0.21	0.45	−0.16	0.21	0.05	0.44	0.22	0.46	−0.02	0.38
pHR	−0.19	0.51	0.02	0.19	0.09	0.51	−0.14	0.45	0.09	0.28
RLL	−0.04	0.53	0.02	0.20	−0.07	0.44	−0.17	0.40	0.08	0.29
FPWL	−0.18	0.38	0.04	0.33	0.00	0.41	−0.11	0.37	0.09	0.38

calculated and tested against zero. According to previous studies, for a significant CIT effect, one expects enhanced SCRs, more pronounced HR decelerations, smaller RLL, and smaller FPWL after crime-relevant than after irrelevant items.

CIT Effect in Control Categories—In control categories, the difference between crime-relevant and crime-irrelevant items was tested for each physiological measure and the combined measure. For crime-relevant compared to crime-irrelevant items, SCR was greater, $t\ (37) = 4.46$, $p < .001$, $d = 1.02$; HR deceleration was more pronounced, $t\ (38) = -2.21$, $p < .05$, $d = -0.55$; RLL was not smaller, $t\ (38) = -0.67$, $p > .05$, $d = -0.15$; FPWL was smaller, $t\ (38) = -3.52$, $p < .001$, $d = -0.62$; the combined measure (SCR, pHR, and RLL) was greater, $t\ (37) = 3.70$, $p < .001$, $d = 0.92$.

CIT Effect in Misled Categories—Response differences between crime-relevant and crime-irrelevant items in misled categories were tested. No CIT effects were found for crime-relevant compared to crime-irrelevant items: SCR was not greater, $t\ (37) = 0.70$, $p > .05$, $d = 0.17$; HR deceleration was not more pronounced, $t\ (38) = -0.08$, $p > .05$, $d = 0.00$; RLL was not smaller, $t\ (38) = -1.64$, $p > .05$, $d = -0.40$; FPWL was not smaller, $t\ (38) = -1.28$, $p > .05$, $d = -0.23$; the combined measure was not greater, $t\ (37) = 1.32$, $p > .05$, $d = 0.34$.

CIT Effect in Mislead Categories Compared to Control Categories—It was tested whether these effects were influenced by misleading information. In misled compared to control categories, the difference between crime-relevant and crime-irrelevant items in physiological responses was smaller for SCR, $t\ (37) = 2.39$, $p < .05$, $d = 0.51$, and for the combined measure, $t\ (37) = 1.91$, $p < .05$, $d = 0.38$. There was no such difference in pHR, $t\ (38) = -1.41$, $p > .05$, $d = -0.31$, RLL, $t\ (38) = -0.73$, $p > .05$, $d = -0.16$, or FPWL, $t\ (38) = -1.28$, $p > .05$, $d = -0.48$.

CIT-like Effect for Misleading Items in Misled Categories—Finally, it was tested whether misleading items showed a CIT-like effect (i.e., differences in physiological responses to misleading compared to irrelevant items were tested). Such response differences were expected in the same direction as was expected for the usual CIT effect.

A significant response difference between misleading and irrelevant items in misled categories was confirmed for each of the physiological measures and the combined measure: For misleading compared to irrelevant items, SCR was greater, $t\ (37) = 2.16$, $p < .05$, $d = 0.57$; HR deceleration was more pronounced, $t\ (38) = -2.97$, $p < .005$, $d = -0.61$; RLL was smaller, $t\ (38) = -3.25$, $p < .005$, $d = -0.67$, FPWL was smaller, $t\ (38) = -3.14$, $p < .005$, $d = -0.53$; the combined measure was greater, $t\ (37) = 4.14$, $p < .001$, $d = 1.06$.

Figure 1 illustrates CIT and CIT-like effects for each physiological measure and the combined measure: z-score differences and significance levels are displayed for crime-relevant versus crime-irrelevant items in control categories, for crime-relevant versus crime-irrelevant items in misled categories, and for misleading versus irrelevant items in misled categories.

ROC Analyses—Table 3 shows the results from the binary logistic regression analysis for the different measures and their combinations. Besides results for a fixed-decision criterion of 0.5, ROC areas were computed for a full-range variation of the decision criterion. ROC area differences between control and misled categories are visible, but confidence intervals are widely overlapping between category types. The ROC areas above chance level were proven only for FPWL and the latter two combinations in control categories.

Discussion

The CIT heavily relies on the memory of crime-relevant information. In practice, the CIT is rarely applied immediately after the crime happened. During the time interval between crime and interrogation, misleading information can influence the memory of the crime. Therefore, the question arises whether the CIT is still valid when misleading information influenced crime-relevant memory. For this purpose, a misinformation paradigm was used to distort the memory of specific crime details prior to a CIT.

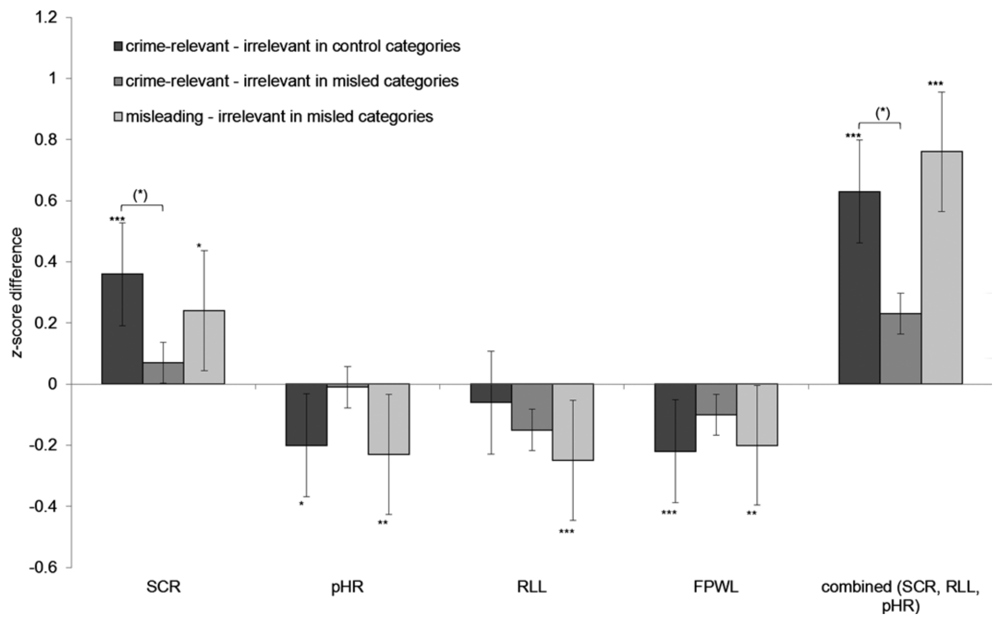

FIG. 1—*CIT and CIT-like effects for each physiological measure and the combined measure: z-score differences and significance levels for crime-relevant versus crime-irrelevant items in control categories, for crime-relevant versus crime-irrelevant items in misled categories, and for misleading versus irrelevant items in misled categories. Error bars represent the standard error of the mean, * $p < .05$, ** $p < .005$, *** $p < .001$.*

TABLE 3—*Subject classification statistics. ROC areas (with 95% confidence intervals) and correct classification rates (with a fixed criterion of 0.5; false positives and false negatives are noted in addition) that were obtained with the single measures and particular combinations (EDA + pHR + RLL with weights determined by regression; EDA + pHR + RLL + FPWL with weights determined by regression, EDA + pHR + RLL with equal weights).*

		With Hold-two-out Shrinkage Correction						Without Shrinkage Correction						
	Included Measures	Area	CI Low	CI High	CC	fpos	fneg	Area	CC	Beta Const	Beta 1st	Beta 2nd	Beta 3rd	Beta 4th
Control categories	SCR	0.6	0.467	0.733	0.618	0.211	0.171	0.62	0.618	−0.21	0.849			
	pHR	0.569	0.432	0.693	0.603	0.205	0.192	0.593	0.603	−0.059	−0.711			
	RLL	0.564	0.437	0.692	0.564	0.244	0.192	0.594	0.551	0.005	−0.418			
	FPWL	0.638	0.514	0.76	0.603	0.205	0.192	0.647	0.603	0.339	1.086			
	SCR + pHR + RLL (regression)	0.575	0.435	0.702	0.474	0.237	0.289	0.636	0.513	−0.281	0.912	−0.778	−0.326	
	SCR + pHR + RLL + FPWL (regression)	0.653	0.524	0.772	0.592	0.211	0.197	0.727	0.632	0.129	0.951	−0.961	−0.386	1.385
	SCR + pHR + RLL (equal weights)	0.635	0.502	0.752	0.592	0.197	0.211	0.649	0.579	−0.206	1.125			
Misled categories	SCR	0.454	0.323	0.59	0.553	0.211	0.237	0.521	0.553	−0.007	0.179			
	pHR	0.539	0.414	0.672	0.538	0.231	0.231	0.566	0.538	0.026	−0.409			
	RLL	0.495	0.366	0.627	0.538	0.244	0.218	0.542	0.538	−0.001	−0.211			
	FPWL	0.141	0.072	0.227	0.231	0.423	0.346	0.512	0.513	0.001	0.006			
	SCR + pHR + RLL (regression)	0.432	0.303	0.567	0.487	0.237	0.276	0.588	0.618	0.009	0.205	−0.327	−0.202	
	SCR + pHR + RLL + FPWL (regression)	0.419	0.291	0.551	0.5	0.237	0.263	0.595	0.618	0.032	0.231	−0.35	−0.152	0.223
	SCR + pHR + RLL (equal weights)	0.545	0.416	0.676	0.592	0.224	0.184	0.576	0.592	0.002	0.43			

Data are presented separately for control and misled categories. Beta weights of the constant and each of the included measures were based on an additional analysis without shrinkage correction.

SCR, electrodermal activity; pHR, phasic heart rate; RLL, respiratory line length; FPWL, finger pulse waveform length; area, area under ROC curve; CI low, lower boundary of 95% confidence interval; CI high, higher boundary of 95% confidence interval.

CC, correct classification rate with 0.5 decision criterion; fpos, rate of false positives (of all subjects) with 0.5 decision criterion; fneg, rate of false negatives (of all subjects) with 0.5 decision criterion; beta const, regression analysis: weight of constant; beta 1st/..., regression analysis: weight of 1st/2nd/... measure.

Misinformation Effect

For the categories in which misleading information was introduced, a misinformation effect emerged as postulated. That is, in misled compared to control categories, crime-relevant information was associated with a lower level of confidence and less accurate source identification; analogously, crime-relevant items were picked less often as video items in the multiple-choice task (i.e., 52% correctly remembered crime-relevant items in misled categories, 62% in control categories). Furthermore, in misled categories, misleading items were associated with higher level of confidence and were more often falsely attributed as “video” items in the multiple-choice task than control items. All three different memory tasks showed similar behavioral effects rather than providing incremental information.

CIT Effect in Control Categories

It was assumed that in control categories, a CIT effect emerges for all peripheral measures. A CIT effect was found for SCR, pHR, FPWL, and the combined measure. In line with previous research, SCR and combined measure values were greater after crime-relevant than after crime-irrelevant items, whereas FPWL and pHR were lower after crime-relevant than after crime-irrelevant items. Effect sizes of SCR, pHR, and FPWL measures were smaller than those typically observed (see 1). Contrary to our expectation, no effects in RLL were found. The missing effects in RLL, as well as the rather small effect sizes in SCR, pHR, and FPWL, may be attributable to the particular procedure used in this study, which will be discussed in more depth later on.

The CIT is Susceptible to Misleading Information

It was assumed that CIT effects in misled categories are smaller compared to those in control categories. This was found for SCR and the combined measure but not for other measures. This outcome may be attributable to the small or even missing CIT effects in FPWL, pHR, and RLL in control categories. Remarkably, no CIT effects were found in misled categories. Instead, differences between misleading and irrelevant items were found in all physiological measures. Thus, in misled categories, misleading information was associated with a CIT-like effect, whereas the actual crime-relevant information was not. Visually, these CIT-like effects appeared larger than the CIT effects observed in control categories, even though the behavioral data indicate participants did remember more crime-relevant than misleading items. This visual discrepancy might be explained by the fact that misleading information was encoded directly before conducting the CIT, whereas crime-relevant information was encoded 1 week earlier.

The present results indicate that the CIT is susceptible to misleading information. After misleading information was provided, none of the physiological measures exhibited a CIT effect. Instead, enhanced physiological responses to misleading information entailed a CIT-like effect. The present results, then, are indicative of a conditioned external validity of the CIT whenever memory is subjected to alterations between crime and interrogation.

Results from the binary logistic regression analysis provide an additional view: The accuracy of classifying participants as “guilty” or “innocent” was estimated after inclusion of simulated data for hypothetical innocent participants. Taking into account only the control categories (i.e., regarding the CIT without misleading information), ROC areas came out lower than in most other CIT studies: Areas did not exceed a value of 0.653 (which was achieved by inclusion of all four measures) and significantly exceeded the random level of 0.5 only for FPWL; for the equal weights combination of EDA, pHR, and RLL; and for full inclusion of all four measures but not for the single measures EDA, pHR, and RLL. Taking into account only the misled categories, none of the areas significantly exceeded the chance level. The pattern reflected in the ROC analysis mostly fits the pattern of effect sizes, but, because of the additional random influence introduced with simulating data for hypothetical innocents, it is not perfectly the same. In sum, classification results provide a different perspective, but they underline what was concluded from the observed effect sizes and the according inferential statistics: CIT accuracy was severely compromised when misleading information was introduced to the participants preceding the CIT.

The results of the present study challenge the validity of the CIT: In meta-analyses, SCR is typically classified as the most valid autonomic parameter (1). Remarkably, CIT effects in SCR and the combined measure disappeared in our study when memory was distorted by misleading information. These observations are in line with results from a study conducted by Gronau and colleagues who induced retroactive interference prior to a CIT: Detection accuracy determined by SCR and a measure combining SCR and respiration was decreased when memory was distorted (34). However, while Gronau and colleagues found detection accuracy above chance level even after retroactive interference, CIT effects vanished entirely in the present study. This outcome difference between the two studies may be due to the different procedures used to encode crime-relevant information (i.e., deeper encoding by a mock crime vs. by a video). In addition, distorting memory via the rather explicit retroactive interference might be less efficient than the more subtle misinformation procedure. Moreover, we included the misleading items in our CIT, whereas in the CIT used by Gronau and colleagues, no items resembling the retroactive interference task were included. It is conceivable that the appearance of misleading items in the CIT might have influenced the CIT effect in our study. To rule out this possibility in future studies, we propose to include an additional CIT condition that does not comprise the misleading items. This change will help to decide whether the CIT effect in our study was transferred into a CIT-like effect (contrasting responses to misleading vs. irrelevant items) or only diminished.

In our study, we found new, CIT-like, effects. All physiological measures showed enhanced responding to misleading compared to irrelevant information. To the authors’ knowledge, this is the first study reporting effects similar to a CIT effect for falsely memorized information.

Limitations and Suggestions for Future Studies

This study was one of the first attempts to combine a misinformation paradigm with a CIT. Thus, some changes in the typical CIT procedure were necessary to achieve a reasonable compromise between an encoding of crime-relevant information sufficient to produce CIT effects while rendering memory vulnerable to postevent misinformation. Misinformation was shown to have no effect on information that has been encoded perfectly (37). In this study, the recognition rates for crime-relevant items were low and thus resulted in small (SCR, pHR, and FPWL) or even missing (RLL) CIT effects in control categories. The relatively small ROC areas obtained from the binary logistic regression analysis for control categories reflect the same and are most likely are due to the same factors. We now discuss methodological issues and name suggestions for future studies that might improve these issues.

First, instead of committing a mock crime, participants watched a video that showed a crime scenario. Therefore, participants were treated as informed witnesses of the crime shown in the video rather than as guilty suspects, which has been shown to produce fewer CIT effects (see 1,9). Moreover, we did not name the crime-relevant to-be-remembered items in advance, as is often done in typical mock crime scenarios. When items are neither learned nor rehearsed, initial encoding also decreases (see 40,41). Therefore, in our study, crime-relevant information might not have been encoded as deeply as it would have been in a mock crime, resulting in small CIT effects as well as the missing CIT effects in RLL in control categories. In future studies, a deeper encoding of crime-relevant information might be achieved by using a classical mock crime procedure. However, it should be kept in mind that a deeper encoding of crime-relevant information may also be associated with less successful distortion of memory.

Second, we used a mixture of central (e.g., the box or the envelope) and peripheral (e.g., the fruit or the picture) crime-relevant items in the video. This was due to the twofold background of this study: Previous studies using the CIT indicate greater and more valid CIT effects for central compared to peripheral items (e.g., 1,38–41), whereas previous misinformation studies suggest using peripheral items to increase the effect of misinformation (e.g., 58,59). Nevertheless, studies have demonstrated that even if CIT effects are greater for central compared to peripheral items, recognition rates did not necessarily differ (38,39; but see 41). Moreover, the items used per category were very similar for some categories in this study. In applied contexts as well as in other studies, unique item alternatives were used (especially for central items, such as the murder weapon and a stolen good) (3). However, a recent study using the same stimulus set but an immediate CIT-like recognition test found high recognition rates (82% compared to 62% in this study; 60). Summarizing our suggestions for selecting stimuli in future studies, CIT effects might increase when items that appear more centrally in the story and that are unique within their category are used.

Third, the encoding of crime-relevant information and the application of the CIT were separated by a 1-week retention interval. This retention interval was proposed by other studies using misinformation (e.g., 60–62). In previous studies, similar time intervals also reduced CIT effect sizes, but effects still occurred (especially for central items; 19,38–41). In the present study, the combination of a 1-week retention interval with mixed central and peripheral items might have contributed to the reduced effects in SCR, pHR, and FPWL, as well as for the lack of effects in RLL. As the small recognition rates in the threefold memory task indicate, both methodological issues might have weakened the encoding of crime-relevant information; therefore, a floor effect might have occurred which rendered the response differences between the control and misled categories insignificant. Nonetheless, it has to be considered that in practice, even longer retention intervals between the crime and the CIT interrogation have to be faced. To further improve our design, we propose that future studies may shorten the retention intervals or combine them with distractor tasks (see also 58,60,61).

Summarizing, we achieved reasonable compensation between an encoding of crime-relevant information sufficient to produce (small) CIT effects while remaining vulnerable to postevent misinformation. Using this procedure, a misinformation effect, as well as CIT effects in three of the four peripheral measures and a combined measure, occurred. In the end, our design was adequate for eliciting misinformation, as well as CIT effects, but the CIT effects might have been stronger if we had used an experimental method that emphasized the CIT effects rather than distortion of the memory. Future studies may want to use mock crime procedures, distractor tasks, or shorter retention intervals and focus more on central crime-relevant information. Optimizing the depth of encoding in memory distortion studies using the CIT remains a balancing act.

Applied Perspective

In practice, two types of CITs are routinely applied in Japan: The "known-solution CIT" and the "searching CIT" (5,6). The known-solution CIT assesses whether a suspect has knowledge about a specific detail of a crime that the police has already identified. The searching CIT also assesses whether a suspect has knowledge about a crime, but contrary to the known-solution CIT, police did not identify the questioned detail beforehand. That is, the details asked for are plausible answer options that the interrogators suggested to be true (5,6). The known-solution CIT is the most common CIT procedure (1,5) and was applied in this study. Missing CIT effects when memory is distorted, as well as CIT-like effects associated with misleading information, may considerably limit the conclusiveness of both types of CIT under specific conditions, in which memory of the crime is distorted.

In a known-solution CIT, reactions to one crime-relevant and several crime-irrelevant items are compared. Guilty suspects may appear innocent due to enhanced physiological reactions to misleading instead of crime-relevant information. If the memory of a guilty suspect is distorted, subjective significance of the crime-relevant information may have been diminished. In this case, no enhanced physiological reactions after crime-relevant information compared to crime-irrelevant information would occur. Consequently, this might lead to more suspects being declared innocent by mistake, resulting in a higher rate of false-negative results.

In a searching CIT, the validity is also decreased. The correct answers to the CIT questions are not known beforehand. Plausible answer options are presented to identify important crime-relevant information by analyzing the physiological reactions of a suspect (5,6). If the suspect reacts more strongly to misleading than to crime-relevant information, the interrogator is led astray by inadequate information about the crime.

Conclusion

Combining the misinformation paradigm of Loftus, Miller, and Burns with Lykken's CIT revealed that the CIT is susceptible to misinformation (8,63). Taken together with other studies, this indicates that the external validity of the CIT is limited to conditions under which the memory of crime-relevant information is neither distorted (34) nor suppressed by the suspect (26,27). In the present study, physiological responses to crime-relevant information were not only reduced by misinformation; strongest responses were even shifted from crime-related to the given misleading information. Hence, further studies using mock crime procedures are needed to improve the understanding of the impact misinformation may have on the validity of the CIT. In practice, when memory is malleable, false-negative results in known-solution CITs, as well as wrong information gathered in searching CITs, have to be taken into account.

References

1. Meijer EH, Klein Selle N, Elber L, Ben-Shakhar G. Memory detection with the Concealed Information Test: a meta-analysis of skin conductance, heart rate, and P300 data. Psychophysiology 2014;51(9):879–904.
2. Hira S, Furumitsu I. Polygraphic examinations in Japan: application of the Guilty Knowledge Test in forensic investigations. Int J Police Sci Manag 2002;4(1):16–27.
3. Matsuda I, Nittono H, Allen JJB. The current and future status of the Concealed Information Test for field use. Front Psychol 2012;3:532. https://doi.org/10.3389/fpsyg.2012.00532. Epub 2012 Nov 27.
4. Nakayama M. Practical use of the concealed information test for criminal investigation in Japan. In: Kleine M, editor. Handbook of polygraph testing. San Diego, CA: Academic Press, 2002;49–86.
5. Ogawa T, Matsuda I, Tsuneoka M, Verschuere B. The Concealed Information Test in the laboratory versus Japanese field practice: bridging the scientist-practitioner gap. Arch Forensic Psychol 2015;1(2):16–27.
6. Osugi A. Daily application of the Concealed Information Test: Japan. In: Verschuere B, Ben-Shakhar G, Meijer E, editors. Memory detection: theory and application of the Concealed Information Test. Cambridge, U.K.: University Press, 2011;253–75.
7. Zaitsu W. External validity of Concealed Information Test experiment: comparison of respiration, skin conductance, and heart rate between experimental and field card tests. Psychophysiology 2016;53(7):1100–7.
8. Lykken DT. The GSR in the detection of guilt. J Appl Psychol 1959;43 (6):385–8.
9. Ben-Shakhar G, Elaad E. The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. J Appl Psychol 2003;88(1):131–51.
10. MacLaren VV. A quantitative review of the Guilty Knowledge Test. J Appl Psychol 2001;86(4):674–83.
11. Gamer M. Detecting concealed information using autonomic measures. In: Verschuere B, Ben-Shakhar G, Meijer E, editors. Memory detection: theory and application of the Concealed Information Test. Cambridge, U.K.: University Press, 2011;27–45.
12. Sokolov EN. Higher nervous functions: the orienting reflex. Annu Rev Psychol 1963;25:545–80.
13. Ben-Shakhar G, Tamar A, Poznansky-Levy A, Asherowitz R, Lieblich I. Stimulus novelty and significance as determinants of electrodermal responsivity: the serial position effect. Psychophysiology 1989;26(1):29–38.
14. Ambach W, Dummel S, Lüer T, Vaitl D. Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. Int J Psychophysiol 2011;81(3):275–82.
15. Verschuere B, Crombez G, de Clercq A, Koster EHW. Autonomic and behavioral responding to concealed information: differentiating orienting and defensive responses. Psychophysiology 2004;41(3):461–6.
16. Klein Selle N, Verschuere B, Kindt M, Meijer E, Nahari T, Ben-Shakhar G. Memory detection: the effects of emotional stimuli. Biol Psychol 2017;129(1):25–35.
17. Ambach W, Stark R, Peper M, Vaitl D. Separating deceptive and orienting components in a Concealed Information Test. Int J Psychophysiol 2008;70(2):95–104.
18. Ambach W, Assmann B, Krieg B, Vaitl D. Face and voice as social stimuli enhance differential physiological responding in a Concealed Information Test. Front Psychol 2012;3:510. https://doi.org/10.3389/fpsyg.2012.00510. Epub 2012 Nov 19.
19. klein Selle N, Verschuere B, Kindt M, Meijer EH, Ben-Shakhar G. Orienting versus inhibition in the Concealed Information Test: different cognitive processes drive different physiological measures. Psychophysiology 2012;53(4):579–90.
20. Ben-Shakhar G. Countermeasures. In: Verschuere B, Ben-Shakhar G, Meijer E, editors. Memory detection: theory and application of the Concealed Information Test. Cambridge, U.K.: University Press, 2011;200–14.
21. Ben-Shakhar G. Current research and potential applications of the concealed information test: an overview. Front Psychol 2012;3:342. https://doi.org/10.3389/fpsyg.2012.00342. Epub 2012 Sep 12.
22. Peth J, Suchotzki K, Gamer M. Influence of countermeasures in the validity of the Concealed Information Test. Psychophysiology 2016;53 (9):1429–40.
23. Elaad E, Ben-Shakhar G. Countering countermeasures in the Concealed Information Test using covert respiration measures. Appl Psychophysiol Biofeedback 2009;34(3):197–208.
24. Honts CR, Devitt ML, Winbush M, Kircher JC. Mental and physical countermeasures reduce the accuracy of the Concealed Knowledge Test. Psychophysiology 1996;33(1):84–92.
25. Ben-Shakhar G, Dolev K. Psychophysiological detection through the guilty knowledge technique: effect of mental countermeasures. J Appl Psychol 1996;81(3):273–81.
26. Bergström ZM, Anderson MC, Buda M, Simons JS, Richardson-Klavehn A. Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. Biol Psychol 2013;94(1):1–11.
27. Hu X, Bergström ZM, Bodenhausen GV, Rosenfeld JP. Suppressing unwanted autobiographical memories reduces their automatic influences: evidence from electrophysiology and an implicit autobiographical memory test. Psychol Sci 2015;26(7):1098–106.
28. Allen JJ, Mertens R. Limitations to the detection of deception: true and false recollections are poorly distinguished using an event-related procedure. Soc Neurosci 2009;4(6):473–90.
29. Ward A, Rosenfeld JP. Attempts to suppress episodic memories fail but do produce demand: evidence from the P300-based complex trial protocol and an Implicit Memory Test. Appl Psychophysiol Biofeedback 2017;42(1):13–26.
30. Rosenfeld JP, Ward A, Drapekin J, Labkovsky E, Tullman S. Instructions to suppress semantic memory enhances or has no effect on P300 in a concealed information test (CIT). Int J Psychophysiol 2017;113:29–39.
31. Baioui A, Ambach W, Walter B, Vaitl D. Psychophysiology of false memories in a Deese-Roediger-McDermott paradigm with visual scences. PLoS ONE 2012;7(1):e30416. https://doi.org/10.1371/journal.pone.0030416. Epub 2012 Jan 17.
32. Deese J. On the prediction of occurrence of particular verbal intrusions in immediate recall. J Exp Psychol Gen 1959;58(1):17–22.
33. Roediger HL, McDermott KB. Creating false memories: remembering words not presented in lists. J Exp Psychol Learn Mem Cogn 1995;21 (4):803–14.
34. Gronau N, Elber L, Satran S, Breska A, Ben-Shakhar G. Retroactive memory interference: A potential countermeasure technique against psychophysiological knowledge detection methods. Biol Psychol 2015;106 (3):68–78.
35. Loftus EF. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. Learn Mem 2005;12(4):361–6.
36. Bernstein DM, Loftus EF. How to tell if a particular memory is true or false. Perspect Psychol Sci 2009;4(4):370–4.
37. Tousignant JP, Hall D, Loftus EF. Discrepancy detection and vulnerability to misleading postevent information. Mem Cogn 1986;4(14):329–38.
38. Gamer M, Kosiol D, Vossel G. Strength of memory encoding affects physiological responses in the Guilty Actions Test. Biol Psychol 2010;83 (2):101–7.
39. Nahari G, Ben-Shakhar G. Psychophysiological und behavioral measures for detecting concealed information: the role of memory for crime details. Psychophysiology 2011;48(6):733–44.
40. Carmel D, Dayan E, Naveh A, Raveh O, Ben-Shakhar G. Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. J Exp Psychol Appl 2003;9 (4):261–9.
41. Peth J, Vossel G, Gamer M. Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. Psychophysiology 2012;49(3):381–90.
42. Takarangi MKT, Parker S, Garry M. Modernising the misinformation effect: development of a new stimulus set. Appl Cogn Psychol 2006;20 (5):583–90.
43. Furedy JJ, Posner RT, Vincent A. Electrodermal differentiation of deception: perceived accuracy and perceived memorial content. Int J Psychophysiol 1991;11(1):91–7.
44. Velden M, Wölk C. Depicting cardiac activity over real time: a proposal for standardization. Int J Psychophysiol 1987;1(2):173–5.
45. Bradley MT, Janisse MP. Accuracy demonstrations, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures. Psychophysiology 1981;18(3):307–15.
46. Verschuere B, Crombez G, Koster EHW, van Bockstaele B, de Clerqu A. Startling secrets: startle eye blink modulation by concealed crime information. Biol Psychol 2007;76(1):52–60.

47. Timm HW. Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. J Appl Psychol 1982;67(4):391–400.
48. Kircher JC, Raskin D. The Computerized Polygraph System II (Software Version 4.01). Salt Lake City, UT: Scientific Assessment Technologies, 2003.
49. Ben-Shakhar G, Elaad E. Finger pulse waveform length in the detection of concealed information. Int J Psychophysiol 2006;61(2):226–34.
50. Lykken DT, Venables PH. Direct measurement of skin conductance: a proposal for standardization. Psychophysiology 1971;8(5):656–72.
51. Ben-Shakhar G. Standardization within individuals: a simple method to neutralize individual differences in skin conductance. Psychophysiology 1985;22(3):292–9.
52. Gamer M, Rill H, Vossel G, Gödert HW. Psychophysiological and vocal measures in the detection of guilty knowledge. Int J Psychophysiol 2010;60(1):76–87.
53. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, MI: Erlbaum, 1988.
54. Field A. Discovering statistics using SPSS. London, U.K.: SAGE, 2009.
55. Meijer EH, Smulders FTY, Johnston JE, Merckelbach HLGJ. Combining skin conductance and forced choice in the detection of concealed information. Psychophysiology 2007;44(5):814–22.
56. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J Math Psychol 1975;12(4):387–415.
57. Gronau N, Ben-Shakhar G, Asher C. Behavioral and physiological measures in the detection of concealed information. J Appl Psychol 2005;90 (1):147–58.
58. Wilford MM, Chan JCK, Tuhn SJ. Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. J Exp Psychol Appl 2014;20(1):81–93.
59. Hardt O, Einarsson EO, Nader K. A bridge over troubled water: reconsolidation as a link between cognitive and neuroscientific memory research traditions. Annu Rev Psychol 2010;61(1):141–67.
60. Volz K, Leonhart R, Stark R, Vaitl D, Ambach W. Psychophysiological correlates of the misinformation effect. Int J Psychophysiol 2017;117:1–9.
61. Pansky A, Tenenboim E, Bar SK. The misinformation effect revisited: Interactions between spontaneous memory processes and misleading suggestions. J Mem Lang 2011;64(3):270–87.
62. Paz-Alonso PM, Goodman GS. Trauma and memory: effects of post-event misinformation, retrieval order, and retention interval. Memory 2008;16(1):58–75.
63. Loftus EF, Miller DG, Burns HJ. Semantic integration of verbal information into a visual memory. J Exp Psychol Hum Learn Mem 1978;4 (1):19–31.

Additional information and reprint requests:
Katja Volz, M.Sc.
Institut für Grenzgebiete der Psychologie und Psychohygiene
Wilhelmstraße 3a
D-79098 Freiburg
Germany
E-mail: volz@igpp.de

Appendix 1 Item material (eight categories with five items each). Items were translated from German to English

Category	Items				
Name on a door plate	Paust	Farck	Krell	Gasch	Deetz
Picture of a landscape	Beach	Forest	Waterfall	Mountains	Flower field
Box	Blue tea caddy	Pink heart-shaped box	Wooden casket	Silver metaled box	Red-white checked box
Fruit	Apple	Lemon	Pear	Mandarine	Kiwi
Shape of a key pendant	Heart	Star	Fish	Flower	Triangle
Color of an envelope	Red	Blue	Green	Black	Brown
Drink	Carrot juice	Cappuccino	Energy drink	Beer	Peach iced tea
Playing card	Cat	Donkey	Snake	Tortoise	Ostrich