

Running head: ENCODING AND QUESTIONING IN A CIT

**Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format**

Wolfgang Ambach \*, Sebastian Dummel \*, Theresa Lürer\*, Dieter Vaitl \*/\*\*

\* Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany

\*\* Bender Institute of Neuroimaging, University of Giessen, Germany

Total pages: 36 (original submission)

Figures: 2

Tables: 1

Correspondence: Wolfgang Ambach  
Institut für Grenzgebiete der Psychologie und Psychohygiene,  
Freiburg, Germany  
Wilhelmstr. 3A  
D-79098 Freiburg  
Germany  
Telephone: +49-(0)761-20721-25  
Fax: +49-(0)761-20721-99  
E-mail: ambach@igpp.de

**Abstract**

Physiological responses in the Concealed Information Test (CIT) are known to depend on the depth of encoding critical items; CIT questions commonly refer to knowledge about critical items. It is unclear to what extent (1) different modes of item handling in a mock-crime, and (2) alternative questioning formats, e.g. asking about participants' particular actions with the critical items, influence the physiological responses.

In the presented mock-crime study with fifty-three participants, two questioning formats, i.e. "Did you see ...?" (*viewing* questioning) and "Did you steal ...?" (*stealing* questioning), were compared between subjects. The mode of encoding, stealing vs. merely viewing the critical objects, was varied within subject. Skin conductance, electrocardiogram, respiration, and finger pulse were recorded.

For both questioning formats and each physiological measure, physiological responses to stolen as well as merely viewed objects differed from those to irrelevant objects. Considering *viewing* questioning, responses to stolen and merely viewed objects did not differ, with the exception of greater phasic decreases of heart rate for stolen objects. Considering *stealing* questioning, responses to stolen objects exceeded those to merely viewed objects with each physiological measure.

The statistically proven interaction between mode of encoding a particular object and questioning format sheds light on the factors influencing the physiological responses in a CIT. The level of subjective significance of a particular item might emerge interactively from the mode of item handling and the questioning format.

**Key Words**

Concealed Information Test; Deception; Orienting Reflex; Mock Crime.

## Introduction

The *Concealed Information Test* (CIT) combines a systematic interrogation with a simultaneous measurement of several physiological data channels in order to detect guilty knowledge. The core assumption of the CIT is that a guilty subject's physiological responses are different for crime-related information compared to crime-irrelevant information (Lykken, 1959). The CIT consists of several multiple-choice questions each referring to another detail of the crime under investigation. Typically, there are four to five answer alternatives to each question, but only one alternative refers to the critical detail. For example, if an envelope was stolen out of an office, a typical CIT question could be "An office requisite has been stolen. Is this the stolen object?"; this question is combined with a sequence of five pictures representing the respective answer alternatives, e.g. a picture of (a) a pencil sharpener, (b) an envelope, (c) a highlighter, (d) a stapler and (e) a Scotch<sup>®</sup> Tape. In this example, the picture of the envelope (b) is the critical item; the other items are referred to as 'irrelevant'. It is assumed that only subjects possessing crime-related knowledge ('guilty' subjects) will recognize the correct item and show a different physiological response to it. Subjects without such knowledge ('innocents') will not be able to discriminate between the alternatives and therefore will not show a systematic response pattern. Numerous laboratory studies have shown that the CIT is a highly valid test for differentiating between guilty and innocent subjects (for review see Ben-Shakhar & Elaad, 2003).

### *Stimulus significance in the CIT*

Physiological, cognitive, and behavioural responses to a given stimulus (e.g. the picture of an object) are referred to as the *orienting reflex* (OR) (Sokolov, 1963). Two factors contribute independently to the strength of the OR: stimulus novelty and stimulus significance (Ben-Shakhar, 1994). Concerning the CIT, stimulus significance is mainly supposed to account for the differential responding to critical vs. irrelevant items. Critical items are thought to have a special significance or an added 'signal

value', whereas irrelevant items do not have this special meaning (Lykken, 1974). Experimental studies on the CIT have shown that a variety of procedures can be used to make items significant (Ben-Shakhar & Eiaad, 2003), e.g. by asking a subject to choose a card out of a deck of cards (e.g. Ben-Shakhar, 1994) or by asking a subject to memorize a list of words or pictures (e.g. Verschuere et al., 2004). In some studies, participants were first asked to commit a mock crime, in which they encountered all the relevant details of the crime that later served as critical items in the CIT (Bradley & Rettinger, 1992; Bradley & Warfield, 1984; Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Gamer, Gödert, Keth, Rill, & Vossel, 2008; Jokinen, Santtila, Ravaja, & Puttonen, 2006). All these studies have in common that the responses to critical items, which are supposed to be particularly significant to the guilty subject, are compared with the responses to irrelevant items without particular significance. However, there is evidence that critical items representing different aspects of a crime differently contribute to the detection of guilt (Carmel et al., 2003; Jokinen et al., 2006). This implies that critical items, which vary in their level of significance for the guilty subject, differently modulate the OR.

#### *Manipulating significance within subjects*

In studies using a mock-crime scenario, some researchers divided the critical items into various subtypes with different significance. Carmel et al. (2003) distinguished between items "peripheral" and "central" to a crime; they found the detection accuracy of the CIT to be higher when only "central" items were analyzed, compared to an analysis including all critical items. Because there was no experimental manipulation of significance of the critical items in this study and because "peripheral" items were less often remembered, no definite conclusions can be drawn with respect to the role of item significance. Jokinen et al. (2006) distinguished between "salient" and "non-salient" critical items; the former "were on information about guilty actions", whereas the latter "handled guilty knowledge" (Jokinen et al., 2006, p. 177). Guilty subjects showed larger skin conductance responses for "salient" than "non-salient" critical items. This implies that guilty action enhances a guilty

subject's OR more than guilty knowledge. Despite the fact that item "salience" in this study can be seen as related to item 'significance' (cf. Sokolov, 1963), in this study also, there was no experimental manipulation of significance; the differences in electrodermal responding could be ascribed to an insufficient encoding during the mock crime or differences in item memorability. Following the suggestion of Jokinen et al. (2006), a differentiation of critical items into highly and lowly significant items, according to their relation to actions of the self, seems promising. To avoid confusion concerning different terminologies, we will stick to the term and concept of 'significance'; previously, the term 'significance' has been used with varied meanings (see e.g. Barry, 1982; Ben-Shakhar et al., 1982; Bradley, 2009; Gati et al., 1996); here, we refer to the special importance and meaning a subject attributes to an item. In a few CIT studies, stimulus significance of the critical items was manipulated within subject (Rosenfeld et al., 2006; Verschuere et al., 2004). Rosenfeld et al. (2006) reported a better ERP-based detection of concealed information with autobiographical than incidentally acquired information. Verschuere et al. (2004) did not record psychophysiological data; their question of interest was whether items differing in their level of significance (*guilty knowledge vs. mere knowledge vs. neutral information*) differently affected attention during a probe classification task, which would be reflected in the participants' RTs. In three experiments, Verschuere et al. (2004) consistently found the longest RTs with highly significant items; the OR was regarded as enhanced in these trials. A significant difference between highly and lowly significant items was found in the first but not in the other two experiments. Verschuere et al. (2004) argued that the inconsistency in their findings was primarily due to an insufficient experimental manipulation; a stronger differentiation of item types was desired.

### *CIT questioning formats*

The differentiation of questioning formats referring to acquired knowledge vs. performed actions was first made in studies including not only *guilty* and uninformed *innocent* subjects but also *informed innocents* (Bradley & Rettinger, 1992; Bradley &

Warfield, 1984; Gamer et al., 2008). Guilty subjects actually committed a crime and acquired knowledge related to their own actions, whereas *informed innocents* did not commit a crime but possess knowledge about crime-related details (e.g. from a newspaper article).

In the studies including a group of *informed innocents*, a modified version of the CIT was used, the so-called *Guilty Actions Test* (GAT, Bradley & Rettinger, 1992; Bradley & Warfield, 1984; Gamer et al., 2008). GAT questions ask for crime-related actions (e.g. "Did you steal this office requisite?"), whereas Guilty Knowledge Test (GKT) questions ask for crime-related knowledge (e.g. "Is this the stolen office requisite?"). Thus, when critical items are presented in a GAT, only guilty subjects answer deceptively whereas informed innocents tell the truth. The GAT was supposed to protect informed innocents from being falsely classified as guilty (Bradley & Warfield, 1984); in several studies, the GAT differentiated between guilty subjects and informed innocents (Bradley & Rettinger, 1992; Bradley & Warfield, 1984). Guilty subjects showed larger physiological response differences between critical and *irrelevant* items than *informed innocents*; both groups differed significantly from *innocents unaware*. However, in a study by Gamer et al. (2008), guilty subjects and *informed innocents* showed similar response patterns and were as likely to be classified as guilty.

Both tests were directly compared (Bradley, MacLaren, & Carle, 1996; Gamer, 2010); yet, these studies yielded inconsistent results. Bradley et al. (1996) found an interaction between test type (GAT vs. GKT) and state of guilt (*guilty* vs. *informed innocent*) indicating that only the GAT but not the GKT allowed for a differentiation between *guilty* participants and *informed innocents*. Using a similar study design, Gamer (2010) could not replicate these findings: Independent of the test type used, *informed innocents* and *guilty* subjects showed similar response differences between critical and *irrelevant* items.

In sum, the importance of the questioning format in the CIT is still rather insufficiently understood.

*Aim of the present study*

Instead of comparing validity estimates between CIT variants in the line of the preceding studies, and instead of directly adopting GAT or GKT questioning formats, we aimed at investigating the effects of two different types of item handling, two different questioning formats, and their possible interactions on physiological responding in a CIT. We wanted to investigate whether the question type "Did you steal ...?" (referred to as *stealing* questioning) influences a guilty subject's physiological response pattern to critical items more strongly than the question type "Did you see ...?" (referred to as *viewing* questioning). Additionally, we were interested in the different levels of stimulus significance induced by either performing a criminal action with the critical items (i.e. stealing them) or merely acquiring knowledge by viewing these items. In order to explore the importance and possible interactions of questioning format and mode of handling critical items, we varied both factors independently. We were particularly interested in whether subjects asked *stealing* questions showed larger response differences between highly and lowly significant items than guilty subjects asked *viewing* questions.

1. Physiological response differences between critical and irrelevant items were expected with both questioning formats and both types of critical items, i.e. highly significant (stolen) and lowly significant (merely viewed) items.
2. For both questioning formats, stronger response differences were expected between highly significant and irrelevant items than between lowly significant and irrelevant items.
3. We expected larger physiological response differences between highly significant critical items and irrelevant items with *stealing* questioning than with *viewing* questioning.

## Materials and Methods

### *Subjects*

Fifty-three healthy students (16 m, 37 f; mean age  $23 \pm 2.79$  years) voluntarily participated in the study. They were paid 9 Euros, with an additional incentive of 3 Euros. Informed consent was obtained from all participants. Three subjects were discarded from evaluation because of technical problems or insufficient compliance with the instructions.

### *Design and Procedure*

The experiment was divided into two parts (mock-crime in an 'administration room' and detection procedure in the 'laboratory'), each guided by a different experimenter. Significance of the mock-crime-objects (critical items) was manipulated within-subject: Items were either stolen (high significance) or merely viewed (low significance) by the subjects. In the second part of the experiment, subjects were randomly assigned to two experimental conditions: half of them were asked *viewing* questions ("Did you see this object ...?"), the other half *stealing* questions ("Did you steal this object...?"). From each group, 25 valid data sets were obtained. Afterwards, two memory tests were performed.

### *Mock-crime scenario*

In an office of the institute, subjects received a 'control sheet' from the first experimenter. They had to check the presence of 18 objects in the administration room. The choice of the 18 objects, two from each category, was randomized and balanced across subjects. The object categories, each comprising five objects, were: key pendants, kitchen objects, boxes, office materials, cosmetics, wooden toy fruits, drink packages, playing cards, and plastic flowers.

Back with the experimenter, participants received a rolled-up document with the instruction to perform a 'special task'. They had to go into the administration room again; there, they had to steal one specified object of each pair (*stolen* items,

counterbalanced across subjects). The other nine objects left in the administration room, i.e. the *seen* items, were not mentioned again in the instruction. Subjects had to put the nine *stolen* items into a suitcase, which they kept closely to themselves throughout the remaining experiment. An amount of 3 Euros was hidden in one of the stolen objects (a box); this later served as incentive to 'remain undetected'.

### *CIT questioning formats*

In the laboratory, the 'physiological investigation' took place with a second experimenter; recording devices were attached.

The CIT consisted of nine blocks referring to the nine item categories (e.g. key pendants, cosmetics). Each block comprised one question with five answer alternatives: the *stolen* and *seen* item of each category and three corresponding *irrelevant* items all unknown to the subjects. The text of each question appeared on the screen five times in sequence, with a different picture of one of the five answer alternatives below the question. The first item presented for each question served as buffer item; the according trials were discarded from analysis. Preceding each block, two *neutral* items were presented as distractors. The according questions referred to everyday objects, which had to be identified (e.g. "Is this a scarf?"). These two questions had to be answered correctly, one with 'yes' and the other with 'no' (in a pseudorandomized sequence), to prevent subjects from answering automatically with 'no'. Responses to these neutral questions were not evaluated. Together with the two neutral questions preceding each category, this resulted in a total of 63 item presentations. The main run was preceded by a training run consisting of two blocks, each with five neutral items. Questions and item pictures were presented foveally on a 19" monitor at a distance of 90 cm for 10 seconds, followed by a blank screen for equally distributed 5.0–7.5 second intervals. Picture size was 6.0° by 8.0° of visual angle. Four seconds after a question was asked, two indication fields containing question marks appeared on either side of the item picture; this prompted the subjects to answer. Then, answers had to be given as quickly as possible by pressing one of the two response keys and by vocally responding with 'yes' or 'no'.

Key assignment was balanced across subjects. Following the answer, the given 'yes' or 'no' replaced the question marks and remained visible on the screen as long as the item question was presented.

Subjects asked *viewing* questions were told to hide their knowledge about all the objects that had been in the administration room – thus, to deny all knowledge about *stolen* items as well as *seen* items. Questions were e.g. "Did you *see* this cosmetic in the administration room?" (note, this is different from the typical, non-involving GKT wording, but we thereby minimized the difference between both question texts).

Subjects asked *stealing* questions had to conceal their knowledge about the *stolen* objects, but they had to answer truthfully with 'no' concerning the *seen* (but not stolen) items (note, this is different from the typical GAT instruction). Questions were e.g. "Did you *steal* this cosmetic from the administration room?".

After subjects were disconnected from the leads, they underwent two memory tests: First, all five pictures of each category were presented on the screen simultaneously, one item category after the other; subjects were asked to identify the item they had *stolen* within each category. In the second test, they had to identify the item they had *seen* but not stolen in the mock crime.

### *Physiological measures*

The physiological recordings took place in a dimly lit, electrically and acoustically shielded experimental chamber (*Industrial Acoustics GmbH, Niederkrüchten, Germany*). Subjects sat in an upright position so that they could comfortably see the monitor and reach the keyboard. Temperature in the cabin was set at about 21 °C at the beginning of the first run, with an increase of maximum 2 °C throughout the course of the experiment.

Skin conductance, respiratory activity, electrocardiogram (ECG), and finger plethysmogram were registered. Physiological measures were A/D-converted and logged by the *Physiological Data System I 410-BCS* manufactured by *J&J engineering (Poulsbo, Washington)*. The A/D-converting resolution was 14 bit, allowing skin conductance to be measured with a resolution of 0.01 µS. All data were

sampled with 510 Hz. Triggers indicating question onsets were registered with the same sampling frequency.

For skin-conductance recordings, 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*), and a constant voltage of 0.5 volts were used. The electrodes were fixed at thenar and hypothenar sites of the nondominant hand. For registration of respiratory activity, two PS-2 biofeedback respiration sensor belts (*KarmaMatters, Berkeley, California*) with a built-in length-dependent electrical resistance were used. They were fixed at the upper thorax and the abdomen. ECG was measured with *Hellige* electrodes (diameter 1.3 cm) according to Einthoven II. Finger pulse signal was transmitted by an infrared system in a cuff around the middle finger of the nondominant hand.

#### *Behavioral measures*

Subjects responded with 'yes' or 'no' by key presses as well as verbal expressions. Key presses indicating 'yes' or 'no' answers were time-logged, synchronized with the physiological measures, and stored on the stimulus-presenting computer. Importantly, answers were delayed by four seconds in this study; after this delay, most stimulus processing and answer preparation can be assumed to be completed; in addition, it is rather easy to perform strategic manipulations by voluntarily controlling reaction speed after the delay. Therefore, behavioral data were not further analyzed.

#### *Data reduction and statistical analysis*

Skin conductance data from six subjects (four with *viewing* questioning, two with *stealing* questioning) had to be discarded from analysis because of electrodermal hypo- or non-responding (more than 85% non-responses). Skin conductance response was defined as any increase in conductance initiated within a time window from 1.0 to 5.0 seconds after trial onset. The amplitude of the response was automatically evaluated as the difference between response onset and the

subsequent maximum value in the set time window (Furedy, Posner, & Vincent, 1991).

Respiratory data were low-pass filtered (10 dB at 2.8 Hz); respiration line length (RLL) was automatically computed over a time interval of 10 seconds after trial onset. The RLL measure integrates information about frequency and depth of respiration. The method was derived from Timm (1982) and modified by Kircher and Raskin (2003).

ECG data obtained from one subject (with *viewing* questioning) had to be excluded from analysis because of technical failure. After notch filtering at 50 Hz, R-wave peaks were automatically detected and visually controlled. The R-R intervals were transformed into heart rate (HR) and real-time scaled (Velden & Wölk, 1987). The HR during the last second before trial onset served as pre-stimulus baseline. The phasic heart rate (pHR) was calculated by subtracting this baseline value from each second-per-second poststimulus value. For extracting the trial-wise information of the phasic HR, the mean change in HR within 15 seconds after trial onset, compared to the prestimulus baseline, was calculated (Bradley & Janisse, 1981; Verschuere, Crombez, Koster, Van Bockstaele, & De Clercq, 2007).

Finger pulse waveform length (FPWL) data from seven subjects (three with *viewing* questioning, four with *stealing* questioning) had to be discarded from analysis because of insufficient signal quality. The FPWL within the first 15 seconds after trial onset was calculated from the finger pulse waveform and then subjected to further analyses (Elaad & Ben-Shakhar, 2006). It comprises information about both HR and pulse amplitude.

The delay between the prompt to answer indicated by the appearance of the question marks on the screen and the pressing of the key was calculated as RT.

A within-subject standardization of measured values has been proposed by Lykken and Venables (1971). Here, according to Ben-Shakhar (1985), Gamer, Rill, Vossel, & Gödert (2006), and Gronau, Ben-Shakhar, & Cohen (2005), the physiological measures are z-transformed for each subject and for each data channel. All *seen*,

*stolen*, and *irrelevant* trials (but not *neutral* trials and the first trials of each stimulus category) were used to calculate individual means and standard deviations. The z-transformed values were used in subsequent statistical analyses.

Statistical analyses were performed with *SYSTAT*, Version 13 (*SYSTAT Software, Inc.*, Monte Carlo).

For each measure, a 2\*2 ANOVA was calculated with the between-subject factor Group (*viewing* vs. *stealing* questioning) and the within-subject factor Item Type (*stolen* vs. *seen* items). Significance level for the assessment of main and interaction effects was set to 0.05; epsilon ( $\epsilon$ ) is calculated as ANOVA effect size measure. Follow-up t-tests for matched samples (one-tailed, significance level 0.05) were carried out when interactions were found; Cohen's d was calculated as estimate of effect size (Cohen, 1988; Rosnow & Rosenthal, 1996).

## Results

### *Memory test*

In the memory test, 96.4% of the *stolen* and 94.4% of the *seen* items were identified correctly (*stolen* items: 95.3% in the *knowledge* and 97.4% in the *stealing* questioning group; *seen* items: 93.8% in the *viewing* and 95.1% in the *stealing* questioning group. A 2 x 2 repeated-measures ANOVA neither revealed a main effect for Item Type (*stolen* vs. *seen* items;  $F_{1,48} = 3.40$ ;  $p = 0.07$ ) nor for Group (*viewing* vs. *stealing* questioning;  $F_{1,48} = 0.73$ ;  $p > 0.1$ ); no Item Type by Group interaction was found ( $F_{1,48} = 0.624$ ;  $p > 0.1$ ). Categories with false identification of either the *stolen* or the *seen* item were discarded from evaluation.

### *Overview of psychophysiological measures*

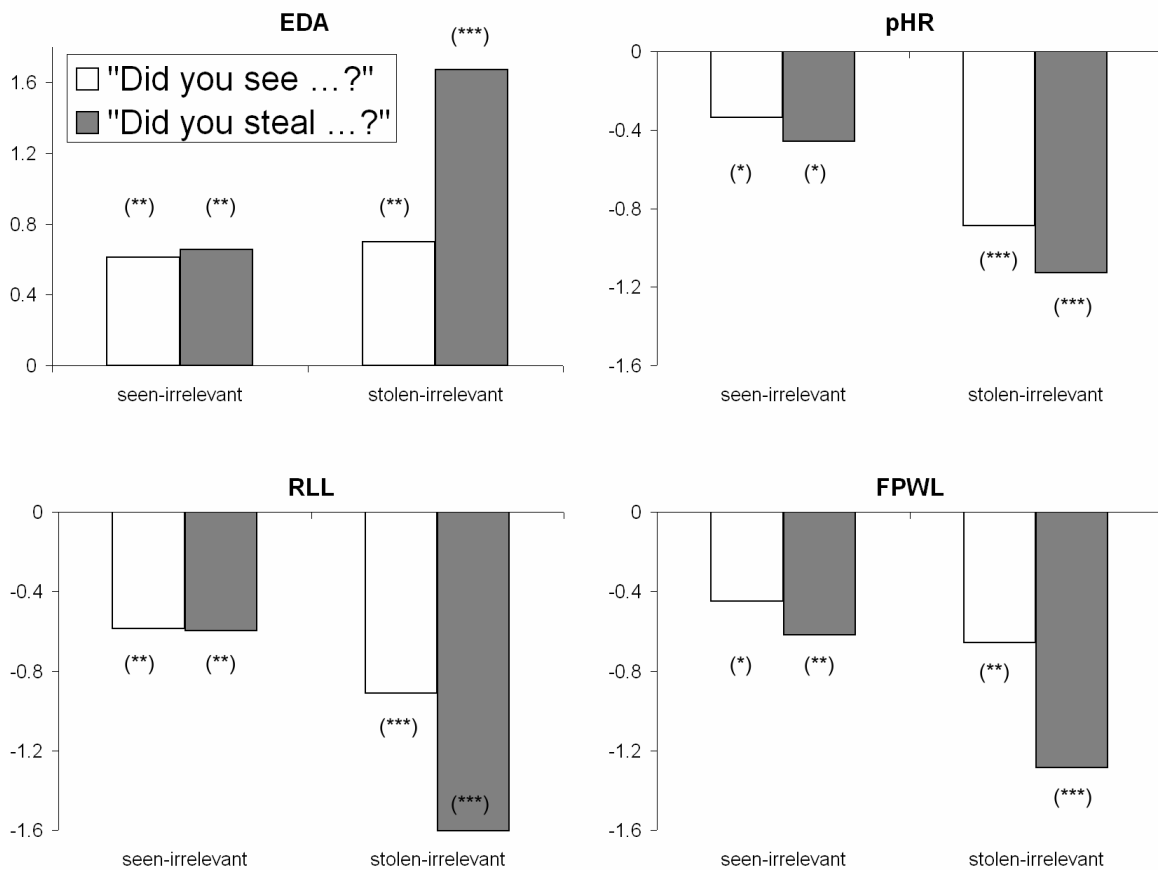
Preceding data standardization and test statistics, descriptive statistics based on raw scores are presented. Table 1 summarizes means and standard errors of means of raw scores for each data channel separately for both questioning formats.

**Table 1.** Means and standard errors of means (SEM) of raw scores for each data channel. Responses to *stolen*, *seen*, and *irrelevant* items are listed separately for *viewing* and *stealing* questioning.

	<i>viewing</i> questioning						<i>stealing</i> questioning					
	<i>stolen</i> items		<i>seen</i> items		<i>irrelevant items</i>		<i>stolen</i> items		<i>seen</i> items		<i>irrelevant items</i>	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
EDA [nS]	138	41	135	41	85	24	269	51	132	25	94	18
pHR [1/min]	-0.54	0.6	0.50	0.42	1.44	0.34	-0.87	0.37	0.51	0.39	1.46	0.33
RLL [arb. units]	3951	300	4044	313	4196	311	3705	365	4139	415	4325	435
FPWL [arb. units]	169	21	169	20	175	20	125	22	136	24	142	25
RT [ms]	963	72	874	55	801	51	656	38	685	44	655	34

Figure 1 illustrates the differential responses to *stolen* vs. *irrelevant* items, as well as *seen* vs. *irrelevant* items, for both questioning formats. Estimated effect sizes (Cohen's *d*) based on z-standardized data are depicted for each of the physiological measures.

**Figure 1.** Effect sizes for the differential responses to *stolen* vs. *irrelevant* items and to *seen* vs. *irrelevant* items: For *viewing* questioning ("Did you see ...?") and *stealing* questioning ("Did you steal ...?"), Cohen's *d* is depicted for electrodermal activity (EDA), phasic heart rate (pHR), respiration line length (RLL), and finger pulse waveform length (FPWL). Asterisks indicate the level of significance of the effects as revealed by one-tailed t-tests (*seen*-vs.-*irrelevant* or *stolen*-vs.-*irrelevant*, respectively; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

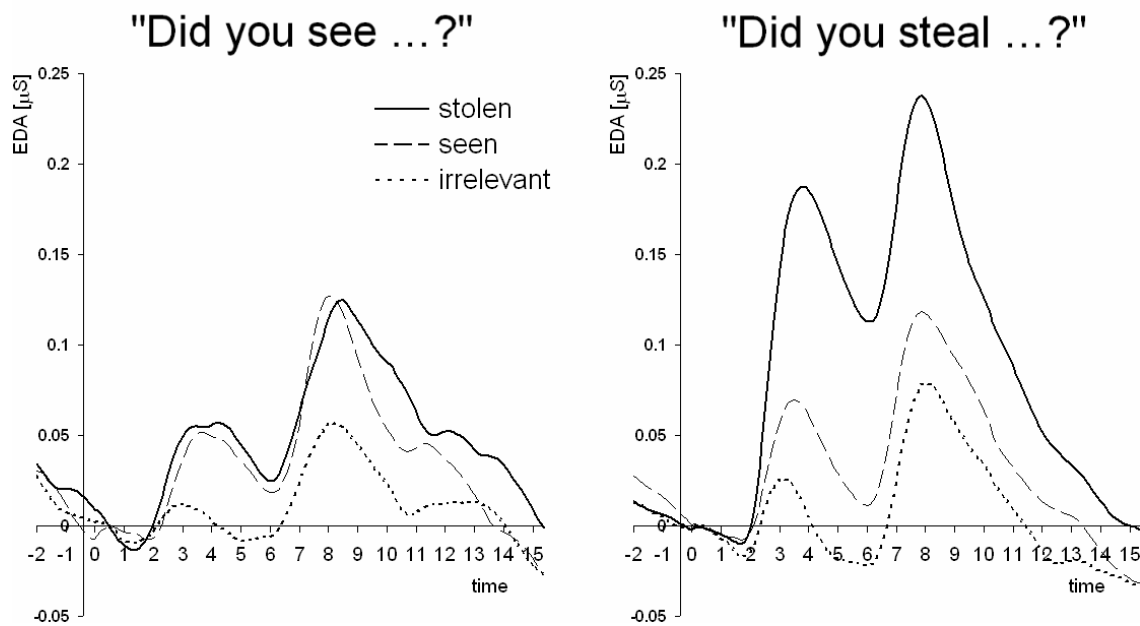


### Skin conductance

Figure 2 shows the averaged intra-trial course of skin conductance depicting grand means for trials with *stolen*, *seen*, and *irrelevant* items separately for both questioning formats. Grand means show two EDA response components with an onset and peak asynchrony of four seconds, which is in accordance with the four-second delay of answers<sup>1</sup>. (Footnote <sup>1</sup>: In addition to the results reported here, a decomposition analysis according to Lim, Rennie, Barry, et al. (1997) was performed for the electrodermal responses; the decomposition algorithm was adopted from Ambach et al. (2008). The results for the decomposed first electrodermal response component,

which followed item presentation, were almost identical to the EDA results reported here. The results for the second electrodermal response component, which followed the answer, were less sensitive to item type and test type. They did not yield incremental information over the first component. For clarity, all reported EDA results refer to a trough-to-peak scoring of responses to item presentation.) Notably, responses to *seen* and *irrelevant* items appear barely influenced by the questioning format; response amplitudes to *stolen* items, however, exceeded those to *seen* items by far with *stealing* questioning, but no analogous difference was visible with *viewing* questioning.

**Figure 2.** Grand means of skin conductance responses to *stolen*, *seen*, and *irrelevant* items with *viewing* questioning ("Did you see ...?") and *stealing* questioning ("Did you steal ...?").



The (2 x 2) ANOVA for EDA revealed a main effect for Item Type ( $F_{1,42} = 23.66$ ;  $p < 0.001$ ;  $\epsilon = 0.75$ ), a main effect for Group ( $F_{1,42} = 8.39$ ;  $p < 0.01$ ;  $\epsilon = 0.45$ ), and an Item Type by Group interaction ( $F_{1,42} = 22.50$ ;  $p < 0.001$ ;  $\epsilon = 0.73$ ). EDA responses were greater to *stolen* than to *seen* items with *stealing* ( $t_{22} = 7.30$ ;  $p < 0.001$ ;  $d = 1.52$ ) but not with *viewing* questioning ( $t_{20} = -1.17$ ;  $p > 0.1$ ).

### Respiration

The (2 x 2) ANOVA for RLL revealed a main effect for Item Type ( $F_{1,48} = 17.19$ ;  $p < 0.001$ ;  $\epsilon = 0.60$ ), a main effect for Group ( $F_{1,48} = 4.40$ ;  $p < 0.05$ ;  $\epsilon = 0.30$ ), and an Item Type by Group interaction ( $F_{1,48} = 10.89$ ;  $p < 0.01$ ;  $\epsilon = 0.48$ ). RLL was smaller for *stolen* than for *seen* items with *stealing* but not with *viewing* questioning ( $t_{24} = -4.91$ ;  $p < 0.001$ ;  $d = -0.98$  and  $t_{24} = -1.31$ ;  $p > 0.1$ , respectively).

#### *Heart rate*

The (2 x 2) ANOVA for pHR revealed a main effect for Item Type ( $F_{1,47} = 15.94$ ;  $p < 0.001$ ;  $\epsilon = 0.58$ ) indicating greater HR decelerations for *stolen* than *seen* items; no main effect for Group ( $F_{1,47} = 0.49$ ;  $p > 0.1$ ) and no Item Type by Group interaction ( $F_{1,47} = 1.08$ ;  $p > 0.1$ ) were found.

#### *Finger pulse*

The (2 x 2) ANOVA for FPWL revealed a main effect for Item Type ( $F_{1,41} = 7.93$ ;  $p < 0.01$ ;  $\epsilon = 0.44$ ) and an Item Type by Group interaction ( $F_{1,41} = 6.95$ ;  $p < 0.05$ ;  $\epsilon = 0.41$ ), but no main effect for Group ( $F_{1,41} = 1.60$ ;  $p > 0.1$ ). FPWL was smaller for *stolen* than for *seen* items with *stealing* but not with *viewing* questioning ( $t_{20} = -3.98$ ;  $p < 0.001$ ;  $d = -0.87$  and  $t_{21} = -0.63$ ;  $p > 0.1$ , respectively).

## Discussion

Factors determining differential physiological responding in the CIT are insufficiently explored. The present study compared the relative influence of two different questioning formats on the physiological responses in a CIT. In order to explore the importance of thievish acting with the two different questioning formats additionally, the mode of handling the critical objects was varied within subject: critical objects were either stolen or merely viewed by the participants.

### *Findings across questioning formats*

With both questioning formats and each of the physiological measures, significant response differences were found between *stolen* and *irrelevant* items as well as between *seen* and *irrelevant* items. This finding is in line with earlier findings: Psychophysiological detection of knowledge does not require deceptive responding (e.g. Furedy & Ben-Shakhar, 1991), nor does it require criminal interaction with the critical objects, as studies including 'informed innocents' have shown (e.g. Gamer, 2010).

### *'Viewing' questioning: "Did you see ...?"*

Regarding the CIT with *viewing* questioning, overall effect sizes appear lower than in many other studies (see Ben-Shakhar & Elaad, 2003). This can be due to methodological reasons: Atypically for CIT studies, each CIT question was combined with two *critical* and three *irrelevant* items in the present study; a more usual ratio is one critical item to five *irrelevants*. The fact that two critical objects, one stolen and one merely seen, were associated with particular significance, presumably made these critical objects (and also the response they required) less prominent and might thereby have weakened differential responding. Physiological response differences between item types generally depend on the ratio of critical to *irrelevant* items presented (e.g. Lieblisch et al., 1970). In addition, details of the written instruction

might have negatively influenced motivational or intentional factors, which are known to modulate the differential responses (Furedy & Ben-Shakhar, 1991). Concerning the *viewing* questioning format, the different modes of handling the critical objects (*viewing* vs. *stealing*) influenced only pHR but not the other measures; EDA responses to *seen* vs. *stolen* items were almost identical. This implies that different levels of item significance arising from different item handling are not necessarily reflected in differential physiological responding when participants are asked *viewing* questions. While the mock-crime instructions assigned a different significance to stolen and merely viewed objects, the *viewing* questions (e.g. "Did you see this cosmetic in the administration room?") did not resume this differentiation: With both item types, it was the knowledge about the items, not one's active involvement, which was deceptively denied. One might assume that one component of 'significance' - caused by the particular mode of object handling - is less important for differential physiological responding when participants are asked *viewing* questions. Yet, the dependence of pHR on item handling still lacks explanation; deceptive answering cannot explain this observation because all answers to *seen* as well as to *stolen* items were deceptive with this questioning format.

*'Stealing' questioning: "Did you steal ...?"*

Regarding the CIT with *stealing* questioning, the physiological response differences to *seen* vs. *irrelevant* items were similar to those with *viewing* questioning. This seems remarkable because with this questioning format participants did not answer deceptively to *seen* items but honestly denied to have stolen these objects; thus, the *seen*-vs.-*irrelevant* response differences were not co-determined by deception, as was the case with *viewing* questioning. It remains open whether the influence of deceptive answering on physiological responding was weaker than in previous studies (see Bradley et al., 1996; Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991) or balanced out by additional factors (e.g. the level of motivation; see Ben-Shakhar & Elaad, 2003), which might have differed between questioning formats.

The most important finding of this study is the massively and selectively enhanced responding with *stealing* questioning to *stolen* items as compared to all other combinations of questioning and item type. More specifically, the responses to *stolen* items were enhanced over those to merely *seen* items with *stealing*, but not with *viewing* questioning (with the exception of pHR). Statistically, this finding is reflected in a significant interaction of questioning format and item type (*seen* vs. *stolen* items).

Although the present study was partly motivated by the diverging results from previous studies directly comparing GKT and GAT (Bradley et al., 1996; Gamer, 2010), it was not intended to participate in the debate on the relative validity of these test formats. Rather, the study goal was to obtain information on basic mechanisms of the CIT, which might be differently influential in GKT and GAT. Therefore, we wanted to compare 'knowledge about thievish acting' and 'knowledge about merely viewed objects' with maximum test power. Consequently, we varied stealing vs. mere viewing within subject, whereas the two above-mentioned studies compared groups of 'guilty' and 'informed innocent' participants. Hence, in the present study, all participants committed thievish actions, but they also acquired 'mere' knowledge (which is here meant to denote knowledge about seen but not stolen objects). In addition, our *viewing* questions had an active wording ("Did you see ...?"), which is more typical of the GAT than the GKT; this prohibits equating the *viewing* questions with typical GKT questions. Additionally, the typical GAT, but not the *stealing* questioning used in this study, requires guilty examinees to respond deceptively also to merely seen items. In sum, one cannot directly extrapolate from the present results for *stolen* and *seen* objects to the physiological responses of 'guilty' and 'informed innocent' participants in other studies. Specifically, the first-glance similarity between the questioning formats in this study and studies comparing GAT and GKT is not profound; implications with respect to those studies are inevitably limited due to fundamental design differences.

Gamer (2010) found neither electrodermal nor respiratory or heart rate response differences between GKT and GAT. He argued that the depth of encoding (and also

the level of motivation to remain undetected), if different between 'guilty' participants and 'informed innocents', might bias the comparison of GKT with GAT. In our study, the final memory test neither yielded significant differences between questioning formats nor between modes of item handling. The observed tendency towards greater memorability of *stolen* than *seen* items (across tests) might possibly indicate a generally deeper encoding of *stolen* than *seen* items. Participants had more intensive contact with the *stolen* than the *seen* items. This could principally have contributed to enhanced physiological effects with *stolen* than *seen* items. However, the equivalent responding to *stolen* and *seen* items with *viewing* questioning contradicts this speculation. Also, differences in memorability can barely account for effect differences between questioning formats.

#### *Possible explanations and confounds*

With the more knowledge-related *viewing* questioning, the mode of item handling did not influence physiological responses, but with the more action-related *stealing* questioning, it clearly did. This prominent finding has to/can be explained and discussed in the light of a variety of influencing factors. First, *stealing* questioning implied truthful answering to questions related to the merely viewed objects; therefore, the mode of item handling was confounded with deceptive answering in subjects asked *stealing* questions. Deceptive answering was found to enhance differential physiological responding when motivation to conceal knowledge is low (Ben-Shakhar and Eyal, 2003). Second, physiological responses might have been influenced by the semantic match or mismatch between the participant's action with an object ('see' or 'steal') and the question text ("Did you see ...?" or "Did you steal ...?"). It is well known (e.g. from oddball tasks) that physiological responding depends on the task-relevance of a stimulus. Whether a presented object is task-relevant or not, is in this study defined by the questioning format. It is conceivable that with *stealing* questioning, all *stolen* objects are task-relevant, whereas with *viewing* questioning, *seen* items are (at least) as task-relevant as *stolen* items. Thus, the match or mismatch between mode of item handling and questioning format might contribute to explaining the observed interaction. Third, the questioning format, per

se, might have impacted on physiological responsivity. Being asked questions about one's criminal actions, as compared to being asked about one's knowledge about viewed objects, might be felt as more accusing and demanding; this might enhance the motivation to remain undetected (see Gamer, 2010), which is known to enhance differential responding in a CIT (Ben-Shakhar and Elaad, 2003)<sup>2</sup>. (Footnote <sup>2</sup>: An exploratory analysis of the reaction times measured after the four-second delay, might support this idea. Mean reaction times were shorter with *stealing* than *viewing* questioning. As a tentative explanation, participants might have been more alert and prepared in the condition with *stealing* questioning, so that they perceived the visual prompt to answer more quickly and performed the answer-related motor activity faster.) Fourth, according to the 'oddball' effect, physiological responses to stimuli requiring a specific response increase with decreasing frequency of these stimuli (e.g. Johnson & Rosenfeld, 1992). Here, with *stealing* questioning, only one item per category required a deceptive response, whereas with *viewing* questioning, deceptive answering was required twice as frequently.

Taking together these possible influences on physiological responding, the present results might be conservatively discussed as being due to the combination of (a) enhanced physiological responses when a deceptive answer was given, and (b) enhanced physiological responses in the group undergoing the CIT with *stealing* questions (e.g. due to enhanced motivation or response conflict). It could however be argued, that the contribution of deceptive answering to differential responding would be unusually effective in this case. Following this explanation, deceptive answering would also have increased differential responding to the greatest extent in combination with *stealing* questioning, which supposedly enhanced the motivation to remain undetected. Ben-Shakhar and Elaad (2003), however, found deceptive answering most influential on physiological responding when this motivation was low. A more detailed explanation might additionally consider that (c) a criminal action (stealing an object) might enhance the subjective significance of an object more than merely viewing an object, and that (d) a semantic match might enhance, whereas a semantic mismatch might diminish the physiological responses elicited with a CIT question. This more complex view might be more appropriate to explain the finding

that physiological responses were determined interactively by encoding procedure and questioning format.

To further clarify the question of the relative validity of GAT vs. GKT, future studies should reduce confounds: Research needs to vary the factors 'item handling' (viewing vs. stealing) and 'questioning format' (GAT vs. GKT) independently between subjects; this would ensure that only one particular object per category is significant. In addition, deceptive responses to merely viewed items should be required in both test formats. From the applied perspective, an inclusion of a group of innocent subjects would also be desirable.

#### *Differences between measures*

The various measures were affected differently by the experimental manipulations. Phasic HR was enhanced for *stolen* as compared to *seen* items with both questioning formats; no interaction of item handling with questioning format was found. For the other measures, a *stolen-vs.-seen* difference was observed with *stealing* but not with *viewing* questioning. To exclude the possibility that error variance is responsible for the observed dissociation of physiological response patterns, a replication of the present results should clarify whether the multi-channel results reproducibly trace back to the orthogonal variation of item handling and questioning format.

For an outlook, the question as to what CIT sub-processes are reflected in the various measures should guide further research. The idea that electrodermal and cardiac measures are mediated by different pathways and depend on different mental sub-processes was formulated in the preliminary process theory (Barry, 1996). The implications of this theory for CIT conceptualization are still barely understood. It has recently been suggested that the fractionation of the various CIT measures could help to identify different CIT sub-processes (Ambach, Stark, & Vaitl, 2011). Response fractionation was also observed in a study which employed an action-related question format in a CIT with electrodermal and cardiac recording (Gamer et al., 2008, p. 61); the non-uniform physiological response patterns observed there suggested that "GAT results could not exclusively be interpreted by referring to the orienting response".

Differential sensitivity of the psychophysiological measures to the experimental manipulations of the present study, if replicable, might be useful to identify factors influencing the individual measures and, thereby, elaborating CIT theory.

The present results also suggest a reconsideration of the concept of object 'significance', usually understood as one-dimensionally determining the physiological responses in a CIT. It might be fruitful to regard object 'significance' not as one-dimensionally determining a CIT response, but rather to postulate sub-components. One general component of object 'significance' might be associated with knowledge per se, i.e. the episodic memory for the objects viewed previously. With respect to the present study, the active involvement of the 'ego' in the crucial knowledge might be a further component of 'significance'. The semantic match or mismatch between question text and episodic memory might also contribute to 'significance'.

### *Conclusions*

Psychophysiological responses in a CIT revealed a prominent interaction between mode of item handling (stealing vs. viewing) and CIT questioning format (*stealing* vs. *viewing* questioning). Taking possible confounds into account, it still seems likely that question format has considerable impact on physiological responding in a CIT.

Possible implications tentatively point towards (a) a dependence of differential responding on the semantic match between question content and episodic memory, (b) the need to relativize the one-dimensional view of the CIT-related construct of 'significance', and (c) the assumption that the various physiological measures reflect different CIT sub-processes. Besides focusing on matters of practical application, further studies should vary the present design in order to further and systematically disentangle the influences of depth of encoding, question wording, deceptive answering, and item-type frequency.

## References

- Ambach, W., Stark, R., Peper, M., & Vaitl, D. (2008). Separating deceptive and orienting components in a Concealed Information Test. *International Journal of Psychophysiology, 70*(2), 95-104. doi:10.1016/j.ijpsycho.2008.07.002
- Ambach, W., Stark, R., & Vaitl, D. (2011). An interfering n-back task facilitates the detection of concealed information with EDA but not with cardiopulmonary physiology. *International Journal of Psychophysiology, 80*, 217-226. doi:10.1016/j.ijpsycho.2011.03.010
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology, 12*(4), 387-415. doi:10.1016/0022-2496(75)90001-2
- Barry, R.J. (1982). Novelty and significance effects in the fractionation of phasic OR measures: a synthesis with traditional OR theory. *Psychophysiology, 19*(1), 28-35.
- Barry, R.J. (1996). Preliminary process theory: Towards an integrated account of the psychophysiology of cognitive processes. *Acta Neurobiologiae Experimentalis, 56*(1), 469-484.
- Ben-Shakhar, G., Lieblich, I., Kugelmass, S. (1982). Interactive effects of stimulus probability and significance on the skin conductance response. *Psychophysiology, 19*(1), 112-114.
- Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology, 22*(3), 292-299. doi: 10.1111/j.1469-8986.1985.tb01603.x
- Ben-Shakhar, G. (1994). The roles of stimulus novelty and significance in determining the electrodermal orienting response: Interactive versus additive approaches. *Psychophysiology, 31*(4), 402-411. doi: 10.1111/j.1469-8986.1994.tb02448.x
- Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology, 88*(1), 131-151.
- Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology, 46*, 1-11.
- Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18*(3), 307-315. doi: 10.1111/j.1469-8986.1981.tb03040.x
- Bradley, M.T., MacLaren, V.V., & Carle, S.B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *Journal of Applied Psychology, 81*(2), 153-160.
- Bradley, M. T., & Rettinger, J. (1992). Awareness of crime-relevant information and the Guilty Knowledge Test. *Journal of Applied Psychology, 77*(1), 55-59.
- Bradley, M.T., & Warfield, J.F. (1984). Innocence, Information, and the Guilty Knowledge Test in the Detection of Deception. *Psychophysiology, 21*(6), 683-689. doi: 10.1111/j.1469-8986.1984.tb00257.x
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the Guilty Knowledge Test from simulated experiments: the external validity of mock crime studies. *Journal of Experimental Psychology: Applied, 9*(4), 261-269. doi: 10.1037/1076-898X.9.4.261
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. San Diego, CA: McGraw-Hill.

- Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology*, *26*(4), 442-451. doi: 10.1111/j.1469-8986.1989.tb01950.x
- Elaad, E., & Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *International Journal of Psychophysiology*, *61*(2), 226-234. doi:10.1016/j.ijpsycho.2005.10.005
- Furedy, J. J., & Ben-Shakhar, G. (1991). The Roles of Deception, Intention to Deceive, and Motivation to Avoid Detection in the Psychophysiological Detection of Guilty Knowledge. *Psychophysiology*, *28*(2), 163-171. doi: 10.1111/j.1469-8986.1991.tb00407.x
- Furedy, J. J., Posner, R. T., & Vincent, A. (1991). Electrodermal differentiation of deception: Perceived accuracy and perceived memorial content manipulations. *International Journal of Psychophysiology*, *11*(1), 91-97. doi:10.1016/0167-8760(91)90376-9
- Gamer, M. (2010). Does the Guilty Actions Test allow for differentiating guilty participants from informed innocents? A re-examination. *International Journal of Psychophysiology*, *76*(1), 19-24. doi:10.1016/j.ijpsycho.2010.01.009
- Gamer, M., Gödert, H.W., Keth, A., Rill, H.G., & Vossel, G. (2008). Electrodermal and phasic heart rate responses in the Guilty Actions Test: comparing guilty examinees to informed and uninformed innocents. *International Journal of Psychophysiology*, *69*(1), 61-68. doi:10.1016/j.ijpsycho.2008.03.001
- Gamer, M., Rill, H. G., Vossel, G., & Gödert, H. W. (2006). Psychophysiological and Vocal Measures in the Detection of Guilty Knowledge. *International Journal of Psychophysiology*, *60*(1), 76-87. doi:10.1016/j.ijpsycho.2005.05.006
- Gati, I., Ben-Shakhar, G., Avni-Liberty, S. (1996). Stimulus novelty and significance in electrodermal orienting responses: The effects of adding versus deleting stimulus components. *Psychophysiology*, *33*, 637-643.
- Gronau, N., Ben-Shakhar, G., & Cohen, A. (2005). Behavioral and Physiological Measures in the Detection of Concealed Information. *Journal of Applied Psychology*, *90*(1), 147-158. doi: 10.1037/0021-9010.90.1.147
- Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*(3), 839-843.
- Johnson, M.M., & Rosenfeld, J.P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, *12*(3), 289-306. doi:10.1016/0167-8760(92)90067-L
- Jokinen, A., Santtila, P., Ravaja, N., & Puttonen, S. (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. *International Journal of Psychophysiology*, *62*(1), 175-184. doi:10.1016/j.ijpsycho.2006.04.004
- Kircher, J. C., & Raskin, D. C. (2003). *The computerized polygraph system II (Software Version 4.01)*. Salt Lake City, Utah (USA): Scientific Assessment Technologies.
- Lim, C.L., Rennie, C., Barry, R., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, *25*(2), 97-109. doi:10.1016/S0167-8760(96)00713-1
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385-388. doi:10.1037/h0046060
- Lykken, D.T. (1974). Psychology and the lie detector industry. *American Psychologist*, *29*(10), 725-739. doi:10.1037/h0037441

- Lykken, D. T., & Venables, P. H. (1971). Direct Measurement of Skin Conductance: A Proposal for Standardization. *Psychophysiology*, 8(5), 656-672. doi: 10.1111/j.1469-8986.1971.tb00501.x
- 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(5), 814-822. doi: 10.1111/j.1469-8986.2007.00543.x
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691-692. doi:10.1093/biomet/78.3.691
- Rosenfeld, J.P., Biroshak, J.R., & Furedy, J.J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, 60(3), 251-259. doi:10.1016/j.ijpsycho.2005.06.002
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1, 331-340. doi: 10.1037/1082-989X.1.4.331
- Seymour, T.L., Seifert, C.M., Shafto, M.G., & Mosmann, A.L. (2000). Using response time measures to assess "guilty knowledge". *Journal of Applied Psychology*, 85(1), 30-37. doi: 10.1037//0021-9010.85.1.30
- Sokolov, E. N. (1963). Higher nervous functions: the orienting reflex. *Annual Review of Physiology*, 25, 545-580. doi: 10.1146/annurev.ph.25.030163.002553
- Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology*, 67(4), 391-400. doi:10.1037/0021-9010.67.4.391
- Velden, M., & Wölk, C. (1987). Depicting cardiac activity over real time: A proposal for standardization. *Journal of Psychophysiology*, 1(2), 173-175.
- Verschuere, B., Crombez, G., & Koster, E.H.W. (2004). Orienting to guilty knowledge. *Cognition and Emotion*, 18(2), 265-279. doi: 10.1080/02699930341000095
- Verschuere, B., Crombez, G., Koster, E.H.W., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: startle eye blink modulation by concealed crime information. *Biological Psychology*, 76(1-2), 52-60. doi:10.1016/j.biopsycho.2007.06.001