

## WHAT'S ON YOUR MIND? DETECTING CONCEALED INFORMATION

Bruno Verschuere<sup>1,2, 3\*</sup>, & Ewout H. Meijer<sup>2</sup>

1. Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

2. Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD  
Maastricht, The Netherlands

3. Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent – Belgium

\* Corresponding author      Dr Bruno Verschuere

Department of Clinical Psychology

University of Amsterdam

Weesperplein 4

1018 XA Amsterdam

The Netherlands

Phone: ++310205256799

Email: [b.j.verschuere@uva.nl](mailto:b.j.verschuere@uva.nl)

RUNNING HEAD: DETECTING CONCEALED INFORMATION

KEYWORDS: DECEPTION, GUILTY KNOWLEDGE, CONCEALED INFORMATION,  
MEMORY DETECTION

## **Abstract**

Lie detectors, can be applied in a wide variety of settings. But this advantage comes with a considerable cost: False positives. The applicability of the Concealed Information Test (CIT) is more limited, yet when it can be applied, the risk of false accusations can be limited to an arbitrary low level. The CIT assesses the recognition of critical information that is known only by the examiners and the culprit, e.g., the face of an accomplice. Large effects are obtained with the CIT, whether combined with peripheral, brain, or motor responses. We see three important challenges for the CIT. First, the false negative rate of the CIT can be substantial, particularly under realistic circumstances. A possible solution seems to restrict the CIT to highly salient details. Second, there exist effective faking strategies. Future research will tell whether faking can be detected or even prevented (e.g., using covert measures). Third, recognition of critical crime details does not necessarily result from criminal activity. It is therefore important to properly embed the CIT in the investigative process, while taking care when drawing conclusions from the test outcome (recognition, not guilt).

### **Key references (max 5)**

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology, 88*, 131-151.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*, 385-388.

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

## **Lie detection versus memory detection**

‘Did you do it?’. Just imagine the sheer possibilities if techniques were available that could determine whether a defendant’s answer to that question would be true or false. Several companies, researchers, and organizations claim to have developed such techniques, for instance using activity of the autonomic nervous system (Control Question polygraph test; e.g., Honts, 2005), facial temperature (Pavlidis, Eberhardt, & Levine, 2002), non-verbal behavior (Meyer, 2011) and Voice (Stress) Analysis (Lieberman, 2012). Yet, these lie detection techniques have been criticized for several reasons (Vrij, 2008). The greatest concern is probably that these techniques rely on the stress approach: liars are expected to show more signs of stress than truth tellers. Yet, truth tellers may also display the very signs of stress that are used to infer deception. When this occurs, the truth teller is falsely accused of being deceptive, the so called false positive or Othello-error (Ekman, 1985).

Consider for instance the Control Question polygraph Test. This test alternates *relevant* questions regarding the reason of the interrogation (‘Did you do it?’-like questions such as ‘On April, 23, did you shoot Sarah?’) with so called *control* questions (e.g., ‘During the first 25 years of your life, did you ever tell an important lie?’) while measuring skin conductance, blood pressure, and respiration. As lying is assumed to be characterized by increased stress, stronger physiological responding to the relevant questions relative to the control questions causes a deceptive test outcome. This reasoning is, however, heavily contested (Ben-Shakhar, 2008; Ben-Shakhar & Furedy, 1990; Iacono, 2006; Lykken, 1974; Meijer & Verschuere, 2010; National Research Council, 2003; Patrick & Iacono, 1991). After all, it is transparent to both the guilty

and the innocent examinee that the relevant questions are most important. Truth tellers may therefore also perceive the relevant questions as more stressful, and consequently show stronger physiological reactivity to those than to the control questions, resulting in a false positive test result. So even though lie detectors have the advantage of having very broad applicability, they also share a substantial risk; namely that of falsely accusing truth tellers.

In contrast to lie detectors, the Concealed Information Test (CIT; previously referred to as the Guilty Knowledge Test see Lykken, 1959) possesses exactly the opposite characteristics;. It's applicability is much more limited, yet the false positive rate can be minimized, and a priori set at arbitrary low level. Already in 1908, Münsterberg reasoned that the legal system could benefit from a method that could determine "*whether a suspected person knows anything about a certain place or man or thing*" (Münsterberg, 1908; p 132). This is precisely what the CIT aims to do.

### **The Concealed Information Test**

Imagine the following scenario: There has been a hold-up, and the police has reason to believe that one of their usual suspects is involved. The suspect has no verifiable alibi, and is brought to the police station for further investigation. The suspect denies involvement in the holdup, or to even know anything about it. The police officer therefore proposes to take a CIT to determine whether the suspect really has no knowledge regarding the holdup. The suspect is connected to a machine, measuring skin conductance, breathing, and heart rate. Then, the question: 'Do you know what kind of masks the perpetrators were wearing? Is presented, along with several alternatives: 'Were they wearing masks of ....Soldiers? ....Disney figures? ...Clowns? ...Monkeys? ....Politicians? ....Aliens?'. The suspects answer 'no' to each of

the alternatives. Then, the next question starts, asking about what the perpetrators shouted to the security personnel, again presenting several alternatives. After a few questions, the test is over. The suspect is disconnected from the polygraph, and the physiological responses to the different alternatives of each question are compared. As all alternatives are equally plausible, a key feature of the CIT is that the innocent suspect will show similar physiological responses to all alternatives (see Figure I). Systematic stronger physiological responding to the correct alternatives reveals that the suspect knows more about the crime than previously admitted. Note that the interrogation itself may be stress-inducing, and increase autonomic arousal, but is unlikely to explain consistent stronger physiological responding to the correct alternatives.

INSERT FIGURE I AROUND HERE

## **Measures**

Initially, the CIT relied on a single physiological index, namely skin conductance response (SCR; Lykken, 1959). While it remains to most sensitive autonomic nervous system (ANS) measure to date, other ANS measures are typically co-recorded. These include both respiratory (tidal volume and pace of breathing, typically combined in a measure labeled ‘respiratory line length’) as well as cardiovascular measures (heart rate, and normalized pulse volume).

From the 1980ies, researchers also started to explore brain measures (Farwell & Donchin, 1991; Rosenfeld et al., 1988). Most brain studies recorded an electroencephalogram (EEG), and found that recognizing concealed information is accompanied with a positive deflection in the EEG that may start as early as 300ms after stimulus onset. The direction and latency of this event-related potential (ERP) gave it its name: P300 (P standing for positive). A larger P300 to

the correct alternatives than to the incorrect alternatives indicates recognition. Compared to skin conductance amplitude, ERP waves have a lower signal to noise ratio. This brought about a number of changes in administering the ERP-based CIT. Because the ERP-based CIT consists of the rapid flashing of hundreds of stimuli, ERP researchers searched for a means to assure that the examinee is paying attention to the stimuli. In the ERP-based CIT, the examinee first memorizes a set of target items, and has to indicate during the CIT whether or not the stimulus is a target.

#### BOX1 ABOUT HERE

More recently, neuroscientists have used brain-imaging technology such as functional magnetic resonance imaging (fMRI) to map which areas of the brain are most active when the examinee tries to conceal certain information. A meta-analysis showed that the (bilateral) inferior frontal gyrus, the (right) middle frontal gyrus, and the (right) temporoparietal junction were found to be specifically reactive to concealed information (Gamer, 2011). Activity in the temporoparietal junction has been associated with the detection of relevant and/or infrequent events; activity in the inferior frontal gyrus is often related to active response inhibition.

While the behavioral task in the ERP- and fMRI-based CIT merely served to assure processing of the items, it appeared that the behavioral responding may itself provide an index of concealed information. Behavioral responding to the critical items appears more error prone, and particularly slower than that to the control items. Subsequently researchers found that reaction times (RTs) provides a valid index also without the concurrent recording of ERPs (Seymour, Seifert, Shafto, & Mosmann, 2000; Verschuere, Crombez, Degrootte, & Rosseel, 2009).

#### **Validity**

Since its introduction at the end of the 1950ies by David Lykken (1959), hundreds of studies on the CIT have been conducted. The vast majority of these studies took place in university

laboratories with undergraduate students, who are instructed to hide certain information. The type of information they are instructed to hide varies, but can generally be reduced to information about a playing card, about code words, about autobiographical details, or about a mock crime, with the latter two being the most popular paradigms. In the card test paradigm, participants are asked to select one out of several cards, and are subsequently interrogated on the card they chose: 'Did you chose number...4? ...8? ...6? ...3? ...2?'. In the code word paradigm, participants are instructed to memorize certain information (e.g., 'Operation Black Hawk'), and later have to hide which information they memorized: 'What was the name of the operation? Was it...Operation Barbarossa? ...Operation White Dusk? ...Operation Black Hawk? ...Operation Zen To?...Operation Two Fathers?'. In the autobiographical test, participants are asked *in tempore non suspecto* to provide autobiographical information such as their name, birthdate, phone number, and social security number, and asked to hide that information during the subsequent CIT (e.g., Is your birthday....August, 15?...April, 2?...July, 29?...December, 3?...May, 6??'). Finally, in the mock crime paradigm, participants assigned to the 'guilty' condition are instructed to commit a mock crime (e.g., theft of an Ipod), whereas those assigned to the 'control' condition complete an activity unrelated to the crime (e.g., go for a walk). During the CIT, participants from both conditions are asked about intimate details of the crime, such as 'What was stolen? Was it....a laptop?...an Ipod?...a credit card?...200euro?...a smartphone?'.

One of the advantages of the CIT is that the false positive rate can be set at an arbitrary low level. To avoid false positives, the CIT test items should be homogeneous. Importantly, this can be checked a priori. After item selection, a paper and pencil forced choice CIT can be administered in a number of innocents, in order to assure that the critical items are not select



above chance (Doob & Kirschenbaum, 1973). Furthermore, it has been recommended to present the examinee all questions and alternatives prior to conducting the test to rule out idiosyncratic reasons for heterogeneity between the items that could bias the test (Nakayama, 2002). Although concern was raised that previewing the CIT items may decrease the validity of the CIT (National Research Council, 2003) there is no empirical support for this concern (Verschuere & Crombez, 2008), and it may be legally required to preview the CIT items with the examinee.

Reviewing 15 mock crime studies, Eyal (1998) found detection rates of 81% for guilty examinees and 96% for the innocent. Similar accuracy rates were reported by Ben-Shakhar and Furedy (1990), who reviewed 10 mock crime studies and found detection rates of 84% and 94%, respectively. MacLaren (2001) reported successful detection of 76% of participants with concealed knowledge and 83% of those without concealed knowledge. In sum, whereas the CIT has a substantial false negative rate (16-24%), the false positive rate of the CIT is modest (4-17%). A meta-analysis on 169 conditions of 80 studies using skin conductance amplitude showed that large effects were obtained in all four paradigms. Cohen's  $d$  (see Box 1) for the card, code, autobiographical, and mock crime paradigm was 1.35, 1.16, 1.58, and 2.09, respectively. Across paradigms, Cohen's  $d$  was 1.55. A more recent meta-analysis (Meijer, Klein Selle, Elber, & Ben-Shakhar, submitted) investigated the accuracy of not only SCR, but also of P300, respiration, and heart rate, only for the autobiographical and mock crime paradigm. P300 showed an effect size of 1.56 for the mock crime paradigm and 2.17 for the personal item paradigm. No difference between the two paradigms was found for SCR, respiration and heart rate, which showed effect sizes of 1.55, 1.11, and 0.89 respectively.

BOX2 ABOUT HERE

Laboratory research points to the great potential of the CIT. However the examinees (criminal suspects vs. undergraduate students), the examiners (police vs. research assistant), the interrogation situation (police interview vs. laboratory experiment), the nature of the questions (regarding an actual criminal act vs. card number5), and the stakes (winning a small financial incentive vs. prison time) all differ greatly from those in real-life criminal investigations. As a result, the extent to which these findings generalize to real world applications remains unknown. On the one hand, as laboratory studies entail strict experimental control, these estimates may overstate the CIT's validity. On the other hand, real life circumstances such as the threat of prison time may actually increase the signal value of the relevant items, and the results from laboratory studies may therefore underestimate the CIT's validity. There is virtually no research that examines how the CIT performs in real life. This is all the more surprising given that, in Japan, the CIT is systematically applied with several thousands of tests being conducted per year, and the outcome of the test has been accepted as evidence in court (Osugi, 2011). There are only two published field studies, in which investigators examined the accuracy of the CIT as applied by the Israeli police in real-life criminal investigations. These studies indeed indicate that criminals escape detection at a higher rate than what was observed in laboratory settings, but this might be contributed to the low number of questions used in the tests. Yet, specificity remained very high, with only very few innocents being falsely accused of possessing crime knowledge. Further field testing is essential to demonstrate the validity of the CIT in real-life settings. The key for such research lies in Japan, where the CIT is applied on a daily basis in real-life criminal investigations.

## **Theory**

Laboratory research confirmed that guilty examinees show enhanced responses to the critical items in the CIT. But why? Is it fear of getting caught? Stress? Guilt? Yet, how likely is it that such mechanisms are at play in a laboratory set-up? Take for instance the card test where someone chooses one of several playing cards and appears to react more strongly to the chosen card than to the other cards. Even with such neutral stimuli, and even when there is nothing at stake, the CIT still has validity (Ben-Shakhar & Eyal, 2003). Given that differential responding in the CIT is still obtained under minimal conditions, mechanisms such as fear and stress are unlikely to drive the CIT. Rather than stressing fear or guilt, Lykken (1974) argued that Orienting theory can help to explain the CIT: "*... for the guilty subject only, the 'correct' alternative will have a special significance...which will tend to produce a stronger orienting reflex than that subject will show to other alternatives (p. 728)*".

The orienting reflex (OR) Lykken refers to was first described by Pavlov (1927). It is a conglomerate of behavioral and physiological responses to changes in the environment (Sokolov, 1963). Such changes elicit an investigative reaction, and the OR is therefore also known as the 'What is it reflex?'. The biologic relevance of the reflex is clear: The (potentially important) novel stimulus draws attention, allowing the organism to investigate it and determine how to respond to it. Laboratory research has shown that novel stimuli (e.g., a deviant tone among a series of standard tones) can indeed elicit an OR (Turpin, 1986). It was also found that the OR is more reliably elicited and greater for stimuli that are significant (e.g., task-relevant or biologically relevant). Thus, stimuli that are novel and/or significant elicit an orienting reflex. In the quote above, Lykken only refers to significance to explain physiological responding in the CIT. Novelty is also likely to be at play in the CIT. Note that the critical items are presented less often than the control items – typically in a 1 to 4 ratio. So, given item homogeneity, the critical

stimulus represent a change in the environment only for the guilty examinee. The idea that novelty also contributes to the CIT is supported by the observation that the greater the number of control stimuli (hence, the greater the change in the environment), the greater the responding to the critical items (Ben-Shakhar & Gati, 2003).

Several testable predictions can be derived from the orienting theory. For instance, the physiological responses that accompany the OR are well described: An increase in skin conductance, a decline in heart rate, and suppression of respiration (Lynn, 1966). That critical stimuli in the CIT closely mimic this pattern (Verschuere, Crombez, De Clercq, & Koster, 2004; Verschuere, Crombez, Smolders, & De Clercq, 2009), supports the Orienting theory of the CIT. The ERP and fMRI findings can also be interpreted within the Orienting theory, P300 and OR amplitudes, for example, share several antecedent factors. Just like ORs, P300 amplitude is inversely related to the subjective frequency of the eliciting stimulus (Donchin & Coles, 1988), it increases with the significance of the eliciting stimulus (Johnston, Miller & Burleson, 1986; Yeung & Sanfey, 2004; Castro & Diaz, 2001; Berlad & Pratt, 1995), and with the degree to which the critical stimulus deviates from the standard stimuli (Ford, Roth & Kopell, 1976). Both the P300 and ANS measures seem to reflect attentional processes related to the mobilization for action following motivationally significant stimuli (Nieuwenhuis, de Geus, & Aston-Jones, 2011).

Finally, it is important to note that critical stimuli elicit a decrease rather than an increase in heart rate, again making an explanation in terms of anxiety or stress less plausible. Although Orienting theory cannot explain all research findings (e.g., heart rate deceleration elicited by relevant items is more extended than one would expect from OR theory), and other processes may also play a role (e.g., response inhibition see Verschuere, Crombez, Koster, Van Bockstaele,

& De Clercq, 2007), it remains the most comprehensive and parsimonious theory to date to explain differential responding in the CIT.

### **Existing, renewed, and novel applications**

Despite its sound theoretical underpinning in Orienting theory and its large research base, the CIT is hardly applied in real life – with Japan being the notable exception. There are several reasons why the CIT is not embraced by practitioners (Krapohl, 2011), but two deserve explicit attention. First, practitioners are reluctant to change their current practice. Investigators using the Control Question polygraph Test, for instance, often fail to recognize its limitations nor a need for another test. Secondly, the applicability of the CIT is more limited than any other lie test. Lie detectors have the advantage that they can be applied in nearly every case, since they claim to provide answer to the crucial ‘Did you do it?’-question. The CIT, on the other hand, can only be applied when (1) the investigators can come up with sufficient critical details that are likely to be known by the culprit, (2) these details are unknown to and can be kept secret from innocent suspects until the examination, and (3) the suspect is willing to cooperate with the examination (Krapohl, 2011). These conditions restrict the applicability of the CIT (C. R. Honts, Raskin, & Kircher, 2002; Krapohl, McCloughan, & Senter, 2006; Podlesny & Raskin, 1978).

Recognizing that the CIT can only be validly applied under limited conditions does not negate the fact that when it can be applied, it may provide very convincing evidence. Some have compared the CIT to fingerprinting in this respect (Iacono, 2011): Fingerprints cannot be obtained from every crime scene, but when they are found, and there is no reasonable alternative explanation for their presence, they provide important forensic evidence. One way to think about possible applications is to start from the different laboratory paradigms that have been used: Card, code, autobiographical, and mock crime. Clearly, the translation from the mock crime paradigm

to real-life applications is most straightforward. In Japan, the CIT is used to detect recognition of crime details, including cases of theft, burglary, injury case, homicide, arson, kidnapping, and drug possession (Osugi, 2011). Common questions involve the stolen item and actions against the victim in case of a robbery; the location where the fuel was obtained in case of arson; and the modus operandi and the location of the crime provide possible questions for a variety of crimes.

When translating the code word and card test paradigms to real-life situations, one can think of assessing possible security breaches. Do unauthorized persons know the password used to hack a computer system? Similarly, membership of a terrorist organization can be determined by showing recognition of key details known only by members of the terrorist organization and the intelligence agencies. The autobiographical CIT can be applied in cases when there is reason to doubt the veracity of claims of amnesia (Huntjens, Verschuere, & McNally, 2012; Meijer, Smulders, Johnston, & Merckelbach, 2007). The suspect could then be questioned on the claimed amnesic information. The autobiographical CIT may also be of use when there is reason to query one's identity, as may arise in cases of espionage, but also in criminal cases when someone is suspected to be a wanted criminal living under another alias (see for instance the John Demjanjuk case; Wagenaar, 1988). In those cases, the suspect can be presented with intimate autobiographical information to help determine the identity of the suspect. Similarly, the CIT could be applied to immigration cases to verify an asylum seeker's story. For example, the country of origin could be determined by presenting question about details one would only know if s/he grew up there.

In the examples above, the CIT is used to infer knowledge of information known to the investigator. However, the CIT can also be employed when the correct option is *not* known, and the purpose of the investigation is to detect which of several options is the correct one. This

approach is often referred to as the Searching-CIT (S-CIT; Osugi, 2011) and can be used to discover, for example, the location of the body of a murder victim when the perpetrator is known (Nakayama, 2002), or uncover details related to a terrorism scenario (Meixner & Rosenfeld, 2011). Meijer, Smulders, and Merckelbach (2010) showed that the S-CIT could also be applied to a group of mock terrorism suspects and successfully extracted the target, location and date of an upcoming terrorist attack (see also Meijer, Bente, Ben-Shakhar, & Schumacher, 2013). Moreover, S-CIT can also be successfully applied in groups of suspects if this group includes unknowledgeable participants (Breska, Ben-Shakhar, & Gronau, 2012).

### **Key challenge#1. False negatives.**

The false negative rate of the CIT is substantial, particularly in realistic set-ups. Here, we propose four possible solutions for the false negative problem: Selection of salient items; weighing response measures individually; independently assessing the CIT's sensitivity; and combining measures.

First, lack of responding in the CIT may be related to poor memory. Basic memory research has shown that item saliency adds to memorability. For instance, emotionally arousing events (Bradley, Greenwald, Petry, & Lang, 1992) are better remembered than neutral events; and unexpected events are better remembered than expected events (Donchin, 1981). Jokinen et al (2006) indeed found that items that were considered more salient (e.g, the stolen good) led to greater responding in the CIT than those that were considered less salient (e.g., decoration of the room where the crime happened). Using more salient items may help to reduce the impact of testing under realistic conditions, as illustrated by Carmel and colleagues (2003). Participants enacted a mock crime either under more laboratory or more realistic conditions, with the latter

including time pressure to enact the crime and incidental encoding of the crime details rather than memorizing them to perfection. The more realistic mock crime resulted in weaker detection efficiency than the more laboratory mock crime. The important point, however, was that this effect was analysis nullified when restricting the CIT to the salient items. Thus, the use of salient items may increase CIT validity and buffer detrimental effects of testing under realistic conditions (see also Osugi, 2011).

Second, it could be questioned whether the CIT will work in each and every individual. It is well known that there exist substantial individual differences in general physiological responding, and also in the specific process that is believed to underlie the CIT – the orienting reflex (Dawson, Schell, & Fillion, 2000; Hamm, Cuthbert, Globisch, & Vaitl, 1997; Hodes, Cook, & Lang, 1985; Klorman, Weissberg, & Wiesenfeld, 1977; Ogorman, 1990). To accommodate for individual differences, Matsuda et al (2006) proposed to assess in which physiological system the individual is most responsive, and to use this information to weigh the physiological signals in the scoring of the CIT. Thus, when an individual shows primarily reactivity in the respiratory channel, the examiner could rely stronger on that channel and less on the other channels.

Third, Noordraven and Verschuere (2013) tried to independently assess whether an individual responds in a CIT, allowing to predict in whom the CIT is most likely to work. Guilty and innocent mock crime participants were administered a RT-based CIT that required speeded Yes-response to target items, and No-responses to both critical and control items. Guilty individuals showed a marked slowing on the crime items compared to the control items, but this was particularly so for individuals who also showed a marked slowing on the target items, indicating that target responsivity could index the CIT's potential. Screening out individuals without reliable target slowing indeed increased the CIT's discriminative ability.



Fourth, the sensitivity of the CIT may be increased by combining several measures. The fact that autonomic, behavioral and central nervous system measures are typically recorded in dedicated CIT protocols complicates the combination of measures. For instance, ERPs requires averaging across a great number of trials, whereas the number of trials is limited for ANS-based tests because of habituation. Rather than forcing measures into other protocols (Gronau et al., 2005, Kozel et al., 2009; but see e.g. Matsuda, Nittono, & Ogawa, 2011), there may be more merit in combining measures that are optimally recorded in the same protocol (e.g., several ANS measures; Gamer, Verschuere, Crombez, & Vossel, 2008), or in combining paradigms. Meijer and colleagues (E. H. Meijer, Smulders, Johnston, & Merckelbach, 2007), for instance, found that a forced choice paper and pencil test – the Symptom Validity Test (SVT) – can have incremental validity to an ANS-based CIT.

### **Key challenge#2. Faking.**

In real-life, a failed CIT (i.e., 'recognition indicated') can have great consequences. Examinees may therefore be motivated to try to alter the test outcome in their best interest. Such attempts are known as faking or countermeasure strategies (for a review see Ben-Shakhar, 2011). Faking can come in many forms, yet successful faking entails creating a pattern of responding that mimics that of the innocent. The examinee can either try to avoid enhanced responding to the critical items, or increase responding to the control items. Researchers have explored both physical (e.g., biting your tongue) as well as psychological means (e.g., thinking about an upsetting event) to evoke a response to the control items. Honts and colleagues (1996) found that the sensitivity of the electrodermal measure decreased from 80% in the control condition to only 10% and 40%, in the physical and psychological faking strategy condition, respectively. Ben-Shakhar and Dolev (1996) also found that faking significantly reduced the validity of the

electrodermal measure. Remarkably, both studies found that respiratory measures were not affected by faking. The sensitivity of the P300-based CIT was also observed to drop substantially through faking (from 82% to 18%; Rosenfeld et al., 2004). Most recently, the effects of faking on the fMRI-based CIT were investigated, again with clear effects of faking: Sensitivity dropped from 100% to only 33% (Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011). The effects of faking seem robust. Claims of resistance to faking should therefore be regarded with great caution. For instance, it seems hazardous to conclude from the studies above that respiratory measures would be resistant to faking. Surely, dedicated strategies (holding breath or paced breathing) can affect the respiratory measures to the extent that any difference in respiration between the critical and control items is nullified.

We see two main strategies to cope with faking: *Faking prevention* and *faking detection*. The use of covert measures was proposed to prevent faking (Elaad, 2011a). The reasoning is straightforward: A suspect who is unaware of being assessed, is unlikely to engage in faking strategies. Elaad and Ben-Shakhar (2009) tested innocent, guilty control, and guilty faking participants in two subsequent CITs. Participants were informed that their reactions would be monitored throughout, and that the first CIT involved the use of the polygraph whereas the second involved a human rater. While detached from the polygraph in the second session, respiration was actually still recorded through covert sensors in the seat and the back of the chair. Because guilty fakers seemed to continue their faking strategies even when detached from the polygraph, the use of covert measures did not provide the expected success. Despite the legal and ethical issues that this approach may raise, it seems worthwhile to continue the quest for covert measures (e.g., eyetracking, laser doppler vibrometry, functional near-infrared spectroscopy (fNIRS), thermal imaging), and ways to obscure what is being measured. Recently, Maoz et al

(2012) explored the validity of the CIT with subliminal stimulus presentation. As participants do not know the identity of the stimulus, successful faking (i.e., increasing responding to the control items) may be more difficult. However, it was found that the subliminal presentation of the stimuli reduced the validity of the CIT.

Given it is very difficult to completely prevent faking, the examiner may engage into *faking detection*. One common way to detect faking, is the detection of movement (e.g., by sensors in the polygraph chair). Recently, it was argued that behavioral responding could be used to detect faking in the ERP-based CIT (Rosenfeld et al., 2008). This is an interesting approach, that could be corroborated with other attempts to use behavioral responding to detect faking (Agosta, Ghirardi, Zogmaister, Castiello, & Sartori, 2011; Cvencek, Greenwald, Brown, Gray, & Snowden, 2010).

Although successful faking has been shown to lower the sensitivity of the CIT, it does not affect specificity. It should therefore be acknowledged that many accepted forms of evidence such as fingerprints, DNA or confessions are also vulnerable to simple methods analogous of faking. Wearing gloves will prevent leaving fingerprints. This has never resulted in an argument that these methods should not be acceptable in court. It simple means they should be used as challenge tests. Presence of fingerprints, DNA, or a confession is indicative of guilt, but absence in not indicative of innocence. In line with this reasoning, the CIT outcome should be handled in a similar manner: an information present outcome is indicative of guilt, while an information absent outcome is less indicative of innocence.

**Key challenge#3. Leakage.**

Although the false positive rate of the CIT is low, this does not imply that a 'hit' on the CIT necessarily implies guilt. The main reason is that the CIT assesses recognition, not guilt. Indeed it has been demonstrated that mere recognition in the absence of guilt is sufficient for differential responding in the CIT (Bradley, Barefoot, & Arsenault, 2011). Thus, when the CIT indicates recognition it is important to consider possible sources of that recognition. In case of the crime CIT, being guilty of the crime is one plausible explanation for recognition, yet there are other possible explanations. Innocent bystanders are also likely to recognize the crime details, as are innocents who are exposed to the critical information through other means, through previous interrogations (investigators may have confronted the suspect with some of the crime details), through media (in high profile cases), through accomplices, or the 'word on the street'. It is strongly recommended to inform all involved in the investigation of the CIT procedure and to prevent exposure of certain crime details. This procedure can reduce leakage of the critical details.

Bradley and Warfield (1984) reasoned that innocent suspects who possess crime knowledge lie when instructed to deny crime recognition, but not when denying crime involvement. These authors therefore proposed to adjust the phrasing of the CIT and interrogate the suspect on crime involvement (e.g., *Did you shoot...poison...beat...stabbed... the victim?*) rather than crime recognition (e.g., *Do you know whether the victim was shot?...poisoned?...beaten?...stabbed?*). This modification of the CIT was labeled the 'Guilty Action Test' or GAT as it asks about active involvement in the crime. In a direct comparison of the CIT and the GAT, Bradley et al (1996) found that knowledgeable innocents reacted less strongly in the GAT than in the CIT, whereas guilty suspects peaked in both tests. Unfortunately, Gamer and colleagues (2010) failed to replicate these effects, and found that the vast majority of

knowledgeable innocents showed differential responding in both the CIT as well as the GAT. In the absence of a final solution, it has been recommended to preview the test items with the examinee, asking to provide alternative explanations for possible crime recognition prior to conducting the CIT (Verschuere & Crombez, 2008).

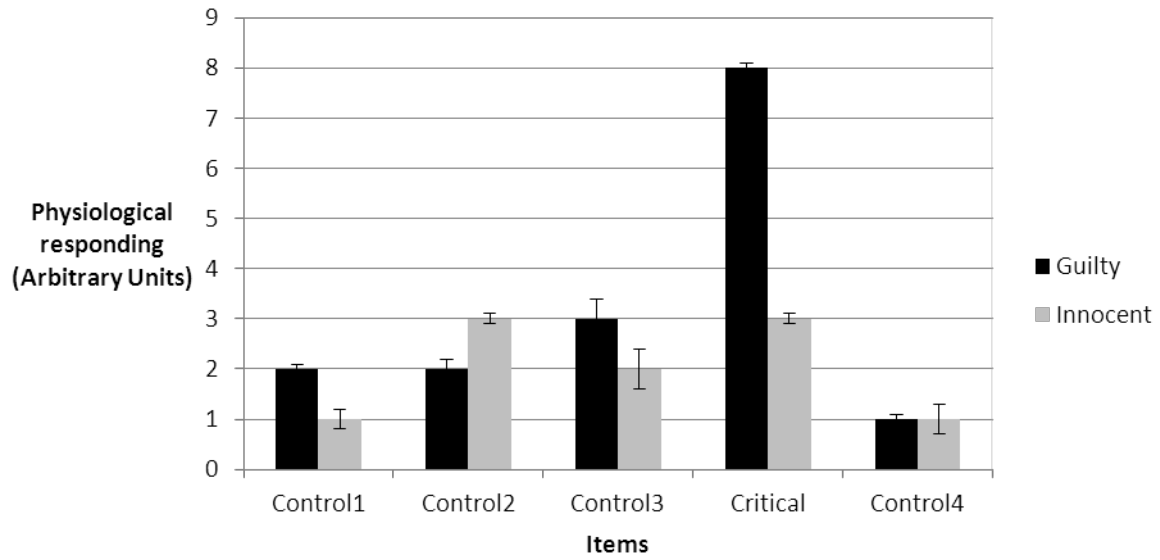
Perhaps additional tests could be used to determine the source of recognition in a CIT. For instance, the autobiographical IAT (aIAT; Sartori et al., 2008) evaluates which of two contrasting autobiographical events is true. Contrasting 'I witnessed the [crime]' with 'I enacted the [crime]' should in principle help to determine whether crime recognition implies crime involvement.

## **Summary**

Hundreds of studies have confirmed the validity of the CIT, whether measuring autonomic, behavioral, or brain responses. Effect sizes are very large on the group level, and several procedures have been proposed that allow to protect the innocent, while detecting a substantial amount of guilty suspects. We reviewed existing and novel applications of the CIT. False negatives, faking, and leakage are important challenges for the CIT, and we have suggested several ways to cope with these challenges.

## Figures

Figure 1. Expected physiological responding in the Concealed Information Test (CIT).



## References

- Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2011). Detecting Fakers of the autobiographical IAT. *Applied Cognitive Psychology, 25*(2), 299-306. doi: Doi 10.1002/Acp.1691
- Allen, J., & Movius, H. L. (2000). The objective assessment of amnesia in dissociative identity disorder using event-related potentials. *International Journal of Psychophysiology, 38*(1), 21-41.
- Ambach, W., Dummel, S., Luer, T., & Vaitl, D. (2011). Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. *International Journal of Psychophysiology, 81*(3), 275-282. doi: DOI 10.1016/j.ijpsycho.2011.07.010
- Ben-Shakhar, G. (1980). Habituation of the orienting response to complex sequences of stimuli. *Psychophysiology, 17*(6), 524-534.
- Ben-Shakhar, G. (2008). The case against the use of polygraph examinations to monitor post-conviction sex offenders. *Legal and Criminological Psychology, 13*, 191-207. doi: Doi 10.1348/135532508x298577
- Ben-Shakhar, G. (2011). Countermeasures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 200-230). Cambridge, UK: Cambridge University Press.
- Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology, 81*, 273-281.

- Ben-Shakhar, G., & Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination. *Journal of Applied Psychology, 87*, 972-977.
- Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology, 88*, 131-151.
- Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception*. New York: Springer-Verlag.
- Ben-Shakhar, G., & Gati, I. (2003). The effects of serial position and frequency of presentation of common stimulus features on orienting response reinstatement. *Psychophysiology, 40*, 139-145.
- Bogaard, G., Meijer, E. H., Vrij, A., Merckelbach, H., & Broers, N. (In press). SCAN is largely driven by 12 criteria: Results from sexual abuse statements. . *Psychology, Crime and Law*.
- Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures - Pleasure and Arousal in Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(2), 379-390. doi: Doi 10.1037//0278-7393.18.2.379
- Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage to innocent suspects. In B. Verschuere, G. Ben-Shakhar & E. H. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 187-199). London: Cambridge University Press.
- 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*, 153-160.



- 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: DOI 10.1111/j.1469-8986.1984.tb00257.x
- Breska, A., Maoz, K., & Ben-Shakhar, G. (2011). Interstimulus intervals for skin conductance response measurement. *Psychophysiology*, *48*(4), 437-440. doi: DOI 10.1111/j.1469-8986.2010.01084.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*, 261-269.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Lawrence Erlbaum.
- Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the Implicit Association Test Is Statistically Detectable and Partly Correctable. *Basic and Applied Social Psychology*, *32*(4), 302-314. doi: Pii 929933922
- Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. B. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 200-223). NY: Cambridge University Press.
- Donchin, E. (1981). Society for Psychophysiological Research - Presidential-Address, 1980 - Surprise ... Surprise. *Psychophysiology*, *18*(5), 493-513. doi: DOI 10.1111/j.1469-8986.1981.tb01815.x
- Doob, A. N., & Kirschenbaum, H. M. (1973). Bias in police lineups - Partial remembering. . *Journal of Police Science and Administration*, *1*, 187-293.

- Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics and marriage*. New York, US: W. W. Norton & Company.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, *75*, 521-529.
- Elaad, E. (1998). The challenge of the concealed knowledge polygraph test. *Expert evidence*, *6*, 161-187.
- Elaad, E. (2011a). New and old covert measures in the Concealed Information Test. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 114-127). Cambridge, UK: Cambridge University Press.
- Elaad, E. (2011b). Validity of the CIT in realistic settings. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 171-186). Cambridge, UK: Cambridge University Press.
- Elaad, E., & Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *International Journal of Psychophysiology*, *61*, 226-234.
- Elaad, E., & Ben-Shakhar, G. (2009). Countering Countermeasures in the Concealed Information Test Using Covert Respiration Measures. *Applied Psychophysiology and Biofeedback*, *34*(3), 197-208. doi: DOI 10.1007/s10484-009-9090-5
- Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, *77*, 757-767.
- Eriksson, A., & Lacerda, F. (2007). Charlatanry in forensic speech science: A problem to be taken seriously. *Journal of Speech, Language, and the Law*, *14*, 169-193.

- Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (lie detection) with event-related brain potentials. *Psychophysiology*, *28*, 531-547.
- 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: DOI 10.1016/j.ijpsycho.2010.01.009
- Gamer, M. (2011). Detecting of deception and concealed information using neuroimaging techniques. In B. Verschuere, G. Ben-Shakhar & E. H. Meijer (Eds.), *Memory Detection: Theory and application of the concealed information test* (pp. 90-113). Cambridge, UK Cambridge University Press.
- Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariations Among fMRI, Skin conductance and behavioral data during processing of concealed information. *Human Brain Mapping*, *28*, 1287-1301.
- Gamer, M., Rill, H. G., Vossel, G., & Godert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology*, *60*, 76-87.
- Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior*, *95*, 333-340.
- Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011). Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage*, *55*(1), 312-319. doi: DOI 10.1016/j.neuroimage.2010.11.025

- Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, *75*(2), 332-346. doi: Doi 10.1037/0022-3514.75.2.332
- 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.
- Hamm, A. O., Cuthbert, B. N., Globisch, J., & Vaitl, D. (1997). Fear and the startle reflex: Blink modulation and autonomic response patterns in animal and mutilation fearful subjects. *Psychophysiology*, *34*(1), 97-107.
- Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Applications of the guilty knowledge test in forensic investigations. *International Journal of Police Science and Management*, *4*, 16-27.
- Hodes, R. L., Cook, E. W., & Lang, P. J. (1985). Individual differences in autonomic response: Conditioned association or conditioned fear. *Psychophysiology*, *22*(5), 545-560.
- Honts, C. R. (2005). The psychophysiological detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 103 - 123). Cambridge: University Press.
- Honts, C. R., & Amato, S. L. (2002). Countermeasures. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 251-264). London: Academic Press.
- Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, *33*(1), 84-92. doi: DOI 10.1111/j.1469-8986.1996.tb02111.x
- Honts, C. R., Raskin, D. C., & Kircher, J. C. (2002). The scientific status of research on polygraph techniques: The case for polygraph tests. In D. L. Faigman, D. H. Kaye, M. J.

- Saks & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Vol. 2, pp. 446-483). ST. Paul, MN: West Publishing.
- Huntjens, R. J. C., Verschuere, B., & McNally, R. J. (2012). No Inter-identity Autobiographical Amnesia in Patients with Dissociative Identity Disorder. *Plos One*, 7, e405080.
- Iacono, W. G. (2006). Detection of deception. In J. T. Cacioppo, L. G. Tassinary & G. B. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 688-703). NY: Cambridge University Press.
- Iacono, W. G. (2011). Encouraging the use of the guilty knowledge test (GKT): What the GKT has to offer law enforcement. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 12-23). Cambridge, UK: Cambridge University Press.
- Jokinen, A., Santtila, P., Ravaja, N., & Puttonen, S. (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. *International Journal of Psychophysiology*, 62(1), 175-184.
- Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-Analysis of Mock Crime Studies of the Control Question Polygraph Technique. *Law and Human Behavior*, 12(1), 79-90.
- Klorman, R., Weissberg, R. P., & Wiesenfeld, A. R. (1977). Individual-Differences in Fear and Autonomic Reactions to Affective Stimulation. *Psychophysiology*, 14(1), 45-51.
- Kozel, F. A., Johnson, K. A., Laken, S. J., Grenesko, E. L., Smith, J. A., Walker, J., & George, M. S. (2009). Can simultaneously acquired electrodermal activity improve accuracy of fMRI detection of deception? *Social Neuroscience*, 4(6), 510-517. doi: Pii 790717993

- Kozel, F. A., Johnson, K. A., Mu, Q. W., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58(8), 605-613.
- Krapohl, D. J. (2011). Limits of the CIT in criminal cases. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 151-170). Cambridge, UK: Cambridge University Press.
- Krapohl, D. J., McCloughan, J. B., & Senter, S. M. (2006). How to use the Concealed Information Test. *Polygraph*, 35(3), 123-138.
- Liberman, A. (2012). TED Lecture: The Emotional Microscope, from [http://www.youtube.com/watch?feature=player\\_embedded&v=x\\_AD0wzf6\\_c](http://www.youtube.com/watch?feature=player_embedded&v=x_AD0wzf6_c)
- Lieblich, I., Shakhar, G. B., & Kugelmass, S. (1976). Validity of guilty Knowledge technique in a prisoners sample. *Journal of Applied Psychology*, 61, 89-93.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385-388.
- Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist*, 29, 725-739.
- MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology*, 86, 674-683.
- Maoz, K., Breska, A., & Ben-Shakhar, G. (2012). Orienting response elicitation by personally significant information under subliminal stimulus presentation: Demonstration using the Concealed Information Test. *Psychophysiology*, 49(12), 1610-1617. doi: DOI 10.1111/j.1469-8986.2012.01470.x

- Matsuda, I., Hirota, A., Ogawa, T., Noriyoshi, T., & Shigemasu, K. (2006). A new discrimination method for the Concealed Information Test using pretest data and within-individual comparisons. *Biological Psychology*, *73*(2), 157-164.
- Matsuda, I., Hirota, A., Ogawa, T., Takasawa, N., & Shigemasu, K. (2009). Within-individual discrimination on the Concealed Information Test using dynamic mixture modeling. *Psychophysiology*, *46*(2), 439-449. doi: DOI 10.1111/j.1469-8986.2008.00781.x
- Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., & Takasawa, N. (in press). Event-related potentials during the standard autonomic-based concealed information test. *Int J Psychophysiol.*
- Matsuda, I., Nittono, H., & Ogawa, T. (2011). Event-related potentials increase the discrimination performance of the autonomic-based concealed information test. *Psychophysiology*, *48*(12), 1701-1710. doi: DOI 10.1111/j.1469-8986.2011.01266.x
- Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology*, *48*(2), 149-154. doi: DOI 10.1111/j.1469-8986.2010.01050.x
- Meijer, E., Ben-Shakhar, G., Elber, L., & Klein Stelle, N. (2012). *Skin Conductance, Respiration, Heart Rate, and P300 in the Concealed Information Test: A Meta Analysis* Paper presented at the World Organization of Psychophysiology, Italy.
- Meijer, E. H., Bente, G., Ben-Shakhar, G., & Schumacher, A. (2013). Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback. *Frontiers in Cognitive Psychology*, *4*, 68.

- Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., & Merckelbach, H. L. G. J. (2007). Combining psychophysiology and forced choice in the detection of concealed information. *Psychophysiology*, *44*, 814-822.
- Meijer, E. H., Smulders, F. T. Y., & Merckelbach, H. L. G. J. (2010). Extracting Concealed Information from Groups. *Journal of Forensic Sciences*, *55*(6), 1607-1609. doi: DOI 10.1111/j.1556-4029.2010.01474.x
- Meyer, P. (2011). Pamela Meyer: How to spot a liar [video file], from [http://www.ted.com/talks/pamela\\_meyer\\_how\\_to\\_spot\\_a\\_liar.html](http://www.ted.com/talks/pamela_meyer_how_to_spot_a_liar.html)
- Münsterberg, H. (1908). *On the witness stand*. New York: Page.
- Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*(6), 733-744. doi: DOI 10.1111/j.1469-8986.2010.01148.x
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigations in Japan. In K. M. (Ed.), *Handbook of Polygraph Testing* (pp. 49-86). London: Academic Press.
- National Research Council. (2003). *The polygraph and lie detection. Committee to review the scientific evidence on the Polygraph*. Washington: The National Academies Press.
- Noordraven, E., & Verschuere, B. (2013). Predicting the Sensitivity of the Reaction Time-based Concealed Information Test. *Applied Cognitive Psychology*, *27*, 328 -335.
- Ogorman, J. G. (1990). Individual differences in the orienting response: Nonresponding in nonclinical samples. *Pavlovian Journal of Biological Science*, *25*(3), 104-110.



- Osugi, A. (2011). Daily application of the CIT in Japan. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory Detection: Theory and application of the Concealed Information Test* (pp. 253-275). Cambridge, UK: Cambridge University Press.
- Patrick, C. J., & Iacono, W. G. (1991). Validity of the Control Question Polygraph Test - the Problem of Sampling Bias. *Journal of Applied Psychology*, *76*(2), 229-238.
- Pavlidis, I., Eberhardt, N. L., & Levine, J. A. (2002). Seeing through the face of deception (vol 415, pg 35, 2002). *Nature*, *415*(6872), 602-602.
- Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*(3), 381-390. doi: DOI 10.1111/j.1469-8986.2011.01313.x
- Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, *15*, 344-359.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *42*(1-2), 157-161.
- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate P300-based method for detection of concealed information. *Psychophysiology*, *45*, 906-919.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205-219.
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately assess autobiographical events. *Psychological Science*, *19*, 772-780.

- Schnur, D. B., Bernstein, A. S., Yeager, A., Smith, S., & Bernstein, P. (1995). The Relationship of the Skin-Conductance and Finger Pulse Amplitude Components of the Orienting Response to Season of Birth in Schizophrenia and Depression. *Biological Psychiatry*, 37(1), 34-41.
- 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.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Macmillan.
- Turpin, G. (1986). Effects of Stimulus-Intensity on Autonomic Responding - the Problem of Differentiating Orienting and Defense Reflexes. *Psychophysiology*, 23(1), 1-14.
- Verschuere, B., & Ben-Shakhar, G. (2011). Theory of the Concealed Information Test *Memory detection: Theory and application of the Concealed Information Test* (pp. 128-148). Cambridge, UK Cambridge University Press.
- Verschuere, B., & Crombez, G. (2008). Deja Vu! The effect of previewing test items on the accuracy of the Concealed Information Test. *Psychology, Crime, & Law*, 14, 287-297.
- Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. H. W. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology*, 41, 461-466.
- Verschuere, B., Crombez, G., Degrootte, T., & Rosseel, Y. (2009). Detecting concealed information with reaction times: Validity and comparison with the polygraph. *Applied Cognitive Psychology*, 23, 1-11.
- Verschuere, B., Crombez, G., & Koster, E. (2004). Orienting to guilty knowledge. *Cognition & Emotion*, 18, 265-279.

- Verschuere, B., Crombez, G., Koster, E., Van Bockstaele, B., & De Clercq, A. (2007). Startling Secrets: Startle eye blink modification by concealed crime information. *Biological Psychology, 76*, 52-60.
- Verschuere, B., Crombez, G., Smolders, L., & De Clercq, A. (2009). Differentiating orienting and defensive responses to concealed information: The role of verbalization. *Applied Psychophysiology and Biofeedback, 34*, 237-244.
- Verschuere, B., Prati, V., & De Houwer, J. (2009). Cheating the lie detector: Faking in the autobiographical IAT. *Psychological Science, 20*, 410-413.
- Vrij, A. (2008). *Detection of lies and deceit: Pitfalls and Opportunities. Second Edition.* Chichester: John Wiley & Sons Ltd.
- Wagenaar, W. A. (1988). *Identifying Ivan: A Case Study in Legal Psychology.* Harvester, Wheatsheaf: Harvard University Press

## Box1. Scoring the CIT

ANS measures. The oldest and most simple method to make an individual judgment based upon these measures is the Lykken scoring (Lykken, 1959). Lykken proposed to score each question with either a 0, a 1 or a 2. The critical item is assigned a score of 2 if it elicits the greatest response, a score of 1 if it elicits the second biggest reaction, and otherwise a 0. Using  $n$  questions, the total score will vary between 0 and 2 times  $n$ , with  $n$  being the typical cut-off. Thus, scores of  $n$  or above are taken as an indication that there is systematic stronger responding to the correct alternatives than to the control alternatives, hence point to recognition.

ERPs. The Lykken scoring cannot be used to compare P300 amplitude for correct versus incorrect alternatives. The size of the P300 can only be reliably determined by averaging across a great number of repetitions, leading to a single average for the critical items and a single average for the control items. To be able to do a classical statistical test, one needs a series of critical-P300's and control-P300's rather than a single score for each of them. The elegant solution to this problem is the bootstrapping technique (Farwell & Donchin, 1991), which involves repeatedly - with replacement - drawing a sample from the critical and control items. For instance, if the ERP CIT consists of 50 critical items, one can draw a 1000 samples, each consisting of 20 critical items, providing a 1000 critical-P300 estimates. The same is done for the control items, allowing to conduct a classic statistical test on any size difference in critical-versus control-P300.

fMRI. The earliest method to score an fMRI-CIT is to specify the brain regions of interest, and to determine whether there is greater activity in those regions in response to the correct items as compared to the incorrect items, using a simple voxel count (Kozel et al., 2005). More

sophisticated scoring algorithms are now also available (Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011).

RTs. To determine whether any RT-difference between correct and incorrect items is substantial enough to conclude that there is recognition, Noordraven and Verschuere (2013) proposed to use individual effect size. Specifically, it was proposed to use the standardized mean difference (Cohen's *d*) with .20 as a cutoff to consider a difference as reliable (0.2, 0.5, and 0.8 are often used as benchmarks to label an effect as small, moderate, or large, respectively; Cohen, 1988). One advantage of this classifier is that it could be applied for all other measures as well.

## Box 2. Signal detection

Intuitively, a test's accuracy is often described in percentage correct decisions. This includes the proportion of *true positives* (proportion of guilty participants classified as guilty), the proportion of *true negatives* (proportion of innocent participants classified as innocent), the proportion of *false positives* (proportion of innocent participants incorrectly classified as guilty) and *false negatives* (proportion of guilty participants incorrectly classified as innocents). An important shortcoming of reporting accuracy in terms of percentages correct decisions is that it relies on an arbitrary cut-off point. According to situational demands, such a cut-off point can be chosen to either minimize *false positives* (e.g., in penal law) or false negatives (e.g., when trying to prevent a terrorist attack). Several measures exist that allow for reporting accuracy regardless of the cut-off point. Most widely used is the *d* statistic. This statistic represents the degree of separation between the distribution of the detection measure (e.g., the relative difference in SCR between the critical and the neutral alternative) in guilty and innocent participants, expressed in standard deviation (National Research Council, 2003).

Editor

*1. First of all, I think it is important to have a clear idea of the target audience for the special issue we are preparing. I think most readers of this issue will not be experts in the field of legal psychology, let alone of lie detection in any form. The target audience will expect to be informed about the state of the art in the topic without taking for granted any detailed knowledge on techniques and research paradigms or terminology. I therefore agree with Reviewer 1 that it would be good to explicitly introduce the two different paradigms CQT and CIT in the beginning of the manuscript.*

In the opening paragraph, we now illustrate the problems of the arousal-based approach by focusing in detail on the CQT. In doing so, we introduce the polygraph, the CQT, the CQT test assumptions, and the most important critiques.

*2. I agree also with Reviewer 2 that the paper is too long. However, several parts contain too much detail for the purposes of an overview article, especially on scoring and subtle differences in assessment methods. I think it is not the purpose to explain to readers how to do an optimal assessment on their own, but rather how valid the approach is, what the problems are, and how existing problems could be addressed in the future. Therefore it should not be a problem to shorten the manuscript by about a quarter to approximately 7500 words. Please consider the suggestions of Reviewer 2 how to rearrange and shorten the manuscript.*

Following the suggestions of the editor and Reviewer2, we substantially shortened the manuscript (to about 6000 words). In particular, we drastically shortened the 'Validity' Chapter and the 'Key Challenges' chapters. Furthermore, we moved information on scoring details to a Box such that that information is available for the interested reader, but it does not hinder the flow of the manuscript.

*3. I do not understand the claim that stimulus novelty plays a role in the CIT (in the "Theory" section). Although critical and control items normally have a ratio of 1:4, all items are novel for truth tellers, and even for liars all items as such are novel, except perhaps the critical item that is the only one the culprit may have knowledge about. But this should decrease the standard CI effect. So please explain in more detail how novelty of items can account for CIT effects.*

This is a good point, and something we would like to clarify. In the context of orienting theory, novelty refers to a change in the environment. In research on novelty orienting, this could be a deviant tone (e.g., 65dBA) among a series of standard tones (e.g., 55dBA), or a colored figure among a series of black-and-white figures. These stimuli are not completely novel in the sense that the examinee has never been confronted with them in their life. Rather, the deviant stimulus is relatively novel, in the sense that it deviates from the others and a such represents a 'change in the environment'. Within the context of the CIT, the critical item differs from the other items and presents a change in environment, only for the guilty examinee. Item homogeneity assures that there is no change in the environment for the innocent. As such, novelty orienting contributes to the CIT effect. Furthermore, the more the stimulus deviates from the other stimuli, the greater the orienting reflex towards it. The greater the number of standard tones, the greater the orienting reflex to the deviant tone. Likewise, the greater the number of control items, the greater the orienting reflex to the critical item.

4. You mention "orienting theory" but you do not explain what it is, nor do you mention a reference. You claim that the CIT has a solid theoretical basis, but I did not find anything on it in the manuscript (see also Reviewer 2).

We did explain orienting theory in the original manuscript, but given comments 4 and 5 of the editor and comment 4 of reviewer2, it becomes clear that we needed to clarify the orienting theory. We rewrote the Theory Chapter, aiming to clarify Orienting theory, providing referenes for it, and pointing to the key findings that empirically support it.

5. I did not understand how the biographical test works. Please explain in a bit more detail.

We tried to clarify how the autobiographical test works, by giving an example of a typical question in such as test. To be consistent, we now provide an example for all four paradigms.

6. I did not understand how it could help to present the examinee with all items before the CIT proper (Nakayama, 2002). In truth tellers you can detect inhomogeneous items, but in culprits this inhomogeneity is confounded with concealed knowledge of the critical items, and in my understanding the presentation of all items should ruin the test, shouldn't it?

Presentation of the items before running the test has several advantages. Most importantly, it allows to verify that the test is unbiased, preventing the subject to come up with a post hoc explanation for enhanced responding. Say the examinee in a robbery case reacts stronger to the correct item (Ipad) than to the other items (laptop, smartphone, 200euro, credit card), and is confronted with the test result. At that time, the examinee might state: 'Well, that must be because I just lost my IPod, and I felt distressed when you asked me about it'. Presenting the items before the test allows to avoid such post hoc reasoning. The examinee could for instance determine to drop any item or question that may be biased. Of course, there is the possibility that the examinee systematically points to the correct items as biased items, but this in itself can be an indication of crime recognition (Meijer, Smulders, Johnston, & Merckelbach, 2007). There are also legal and ethical reasons to advocate presenting the items before the test: The examinee has the right to be informed on how the test will look like before providing informed consent (Verschuere & Crombez, 2008). Despite the practical, legal, and ethical advantages, it is of course possible that 'exposure to the relevant questions prior to the examination would tend to decrease the differential orienting response to the relevant and comparison questions and weaken the test's ability to discriminate (National Research Council, 2003; p. 77)'. It is unclear, however, why this would be the case as prior exposure will not affect the critical item's significance or relative novelty within the test. In two studies, we nonetheless tested this possibility, and did not find any detrimental effect of prior exposure to CIT validity (Verschuere & Crombez, 2008). An elaborate discussion of the advantages of prior exposure seems to fall out of the scope of the review given the request to shorten our manuscript substantially. Yet, we rewrote the sentence to clarify the advantages of prior exposure.

7. "civilized countries" is not a good label, unless you want to allude to the glorious imperialistic past.

We deleted this sentence.

8. The Chapter "Validity: From the group..." contains a lot on scoring details that could be shortened.

See reply to comment 2 of the editor.



*9. On one hand it is claimed that the CIT can be used to detect simulated amnesia, on the other hand that subliminal presentation of critical stimuli still have an effect in CIT. If subliminal stimuli can have an effect, implicit knowledge (that is not introspectively accessible due to amnesia) could also.*

Implicit memory is indeed a possibility that has been raised in the literature (Huntjens et al., 2012), and there are case studies (in prosopagnostic patients) that indicate that the CIT may be reactive to implicit memory. Further empirical research is needed to establish whether the CIT may function as a tests of implicit memory (Allen, 2011), and this discussion seems to fall out of the scope of the present manuscript.

*10. Please omit bullet points in the key challenge #1 section, a simple list will do.*

OK

*11. Also the key challenge sections can be drastically shortened.*

See reply to comment 2 of the editor.

## Reviewer 1

*1. Because most readers will not be very familiar with this topic it would be helpful to add a bit more background information to the first paragraph. Many people will probably not be aware of the fact that different tests are subsumed under the heading of “lie detector” or “polygraph.” If the author(s) would shortly outline the rationale behind the Control Question Test, it might be easier to grasp that the two tests are completely different.*

See reply to comment 1 of the editor.

*2. I wonder whether one should really suggest the application of CIT in immigration cases. Immigration officers’ assumptions are often based on information about the situation or specific incidents in the emigration country that are very often difficult to verify. These are not good preconditions for a CIT.*

Immigrants sometimes present false information in order to try to enhance their chance of acquiring the right to reside. For instance, they may falsely tell that they come from a country where they would fear prosecution for their political conviction, sexuality or ethnicity. Immigrant officers try to check the plausibility of the story, among others by checking factual knowledge on the claimed county of origin. We think that the CIT may also be of use for that purpose. We have tried to clarify our reasoning in the revised manuscript.

*3. I found the discussion on “false negatives” and “faking” very interesting but very detailed and here and there a bit redundant. Maybe this discussion could be shortened a bit.*

See reply to comment 2 of the editor.

## Reviewer 2

*1. But the review is sometimes tedious and the central theme seems to be lost, especially after the half of the paper. The review is too long and some chapters will get better by shortening (e. g. chapter “validity: From the group to the individual”: in the reviewers opinion the authors lost the central theme and it is difficult for the reader to get the main message of this chapter). All in all, the review should be shortened to 7500 words at all.*

See reply to comment2 of the editor.

*2. Please provide more information regarding the neurobiology of the CIT. Which brain regions are activated in more detail? What is about guilty knowledge tests in the scanner?*

We now elaborate on the brain regions that were found to be differentially reactive in the CIT, and point to their possible functional relevance.

*3. Chapter “measures”: Perhaps it would improve this chapter if the statistical basis for the decision (cut-off etc.) from chapter “validity: from group to the individual” are mentioned here.*

We shortened the scoring section and moved it to a Box. The remainder of the ‘Validity: from group to the individual’ regarding individual hit rates was moved to the Validity section.

*4. Chapter “theory”: Please describe the alternative theories in more detail. What empirical data exists for the several different theories?*

See reply to comment 5 of the editor.

*5. Please provide the literature for the meta-analysis at the beginning of the chapter “validity”.*

We rearranged the Validity chapter. We start the chapter with a description of the paradigms used to determine the validity. Next, we summarize the results of laboratory validity research. We end by discussing the need for more field research.

*6. Chapter “key challenge 1”, “key challenge 2” and “key challenge 3”: Please shorten the chapter and work out you central theme. Provide your (new) ideas and suggestions for the researcher and practitioner in a clearer and more distinct way.*

See reply to comment2 of the editor.