



UvA-DARE (Digital Academic Repository)

Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty

Meijer, E.H.; Verschuere, B.; Gamer, M.; Merckelbach, H.; Ben-Shakhar, G.

DOI

[10.1111/psyp.12609](https://doi.org/10.1111/psyp.12609)

Publication date

2016

Document Version

Final published version

Published in

Psychophysiology

License

CC BY-NC

[Link to publication](#)

Citation for published version (APA):

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, 53(5), 593-604. <https://doi.org/10.1111/psyp.12609>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

REVIEW

Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty

EWOUT H. MEIJER,^a BRUNO VERSCHUERE,^{a,b} MATTHIAS GAMER,^c
HARALD MERCKELBACH,^a AND GERSHON BEN-SHAKHAR^d

^aDepartment of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

^bDepartment of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

^cDepartment of Psychology, University of Würzburg, Würzburg, Germany

^dDepartment of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel

Abstract

The detection of deception has attracted increased attention among psychological researchers, legal scholars, and ethicists during the last decade. Much of this has been driven by the possibility of using neuroimaging techniques for lie detection. Yet, neuroimaging studies addressing deception detection are clouded by lack of conceptual clarity and a host of methodological problems that are not unique to neuroimaging. We review the various research paradigms and the dependent measures that have been adopted to study deception and its detection. In doing so, we differentiate between basic research designed to shed light on the neurocognitive mechanisms underlying deceptive behavior and applied research aimed at detecting lies. We also stress the distinction between paradigms attempting to detect deception directly and those attempting to establish involvement by detecting crime-related knowledge, and discuss the methodological difficulties and threats to validity associated with each paradigm. Our conclusion is that the main challenge of future research is to find paradigms that can isolate cognitive factors associated with deception, rather than the discovery of a unique (brain) correlate of lying. We argue that the Comparison Question Test currently applied in many countries has weak scientific validity, which cannot be remedied by using neuroimaging measures. Other paradigms are promising, but the absence of data from ecologically valid studies poses a challenge for legal admissibility of their outcomes.

Descriptors: Detection of deception, Neuroimaging, Validity, Concealed Information Test, Comparison Question Test, Differentiation of deception

Attempts to use psychophysiological measures to detect deception can be traced back to over a hundred years ago (e.g., Munsterberg, 1908; see Lykken, 1998, for an historical overview). Most notably, this early work led to the development of detection methods based on simultaneous recording of multiple physiological measures such as heart rate, blood pressure, and electrodermal activity, and this gave rise to the term *polygraph* (Grubin & Madsen, 2005). The way the polygraph is employed to infer deception has, however, been criticized by a number of academic scholars throughout the years. In

fact, in its 2003 report, the National Research Council (NRC) noted the field has made little progress over the last decades (National Research Council, 2003; see also Meijer & Verschuere, 2015).

Fuelled by the September 11 terror attack in the United States and subsequent military operations in Afghanistan and Iraq, interest in the detection of deception has gained momentum in the past years. This development coincided with the increased accessibility of modern neuroimaging techniques such as positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), and a growing number of neuroscientists have begun to explore whether measurements of brain functions can help to detect deception (Gamer, 2014; Ganis, 2014; see Figure 1). Moreover, private companies such as Government Works, No Lie MRI, and Truthful Brain Corporation are marketing lie detection tests based on brain function. For example, the No Lie MRI website states that “No Lie MRI Inc. provides unbiased methods for the detection of deception . . .” and that this technology “represents the first and only direct measure of truth verification and lie detection in human history” (www.noliemri.com). The scientific community closely

This work was supported by an NWO-VENI grant (451-11-038) awarded to EHM.

Address correspondence to: Dr. Ewout H. Meijer, Forensic Psychology Section, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: eh.meijer@maastrichtuniversity.nl

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

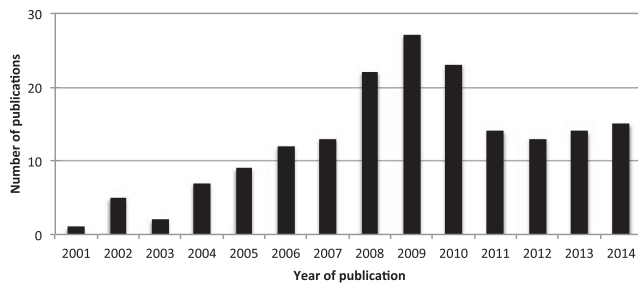


Figure 1. Number of relevant publications about fMRI and lie detection per year, resulting from a Web of Science search using the following terms in the topic field: TS=(“lie detection” OR deception) AND TS=(fMRI OR “functional magnetic resonance”). Results include empirical and review papers, but also papers dealing with ethical and legal evaluations.

monitored these developments, and a number of articles discussing the legal and ethical implications of detection of deception based on neuroimaging have been published in various outlets, including flagship journals such as *Nature Neuroscience* and *Science* (e.g., Editorial, 2008; Farah, Hutchinson, Phelps, & Wagner, 2014; Miller, 2010).

Many of the articles that address the application of neuroimaging to deception detection contrast fMRI with the traditional polygraph tests, and suggest that measuring brain activation can help to overcome the shortcomings of the latter (e.g., Bles & Haynes, 2008; Kozel et al., 2004; Langleben et al., 2005). Farah and colleagues (2014), for example, write that “the appeal of this brain-based lie detection approach is that, in contrast to most previous methods—which detected the emotional arousal resulting from deception—it measures physiological changes associated with cognitive processes during deception and could therefore, in principle, shed light on the process of deception itself” (p. 123). The idea of neuroimaging providing privileged access to the deceptive brain is intuitively appealing. Yet, it ignores one important point: Detecting deceptive behavior always critically depends on a questioning/interrogation paradigm that elicits deceptive behavior. Thus, while in principle neuroimaging research can yield important information about deceptive behavior, it is impossible to evaluate the validity of this research while ignoring the paradigm used for this purpose.

Choosing an appropriate paradigm is important because, in contrast to what is assumed in many articles discussing the neuroimaging of deception, measures cannot be equated with psychological constructs such as deception; a given measure may tap into very different states, depending upon the paradigm in which it is administered. Skin conductance, for example, can index emotional arousal when used in conjunction with a picture-viewing paradigm (Lang, Greenwald, Bradley, & Hamm, 1993), but also attentional orienting when combined with a habituation paradigm (Frith & Allen, 1983). Thus, the paradigm determines the psychological state that any dependent measure reflects. For example, as we will discuss in detail below, the shortcomings of traditional polygraph tests are not related to the reliability and validity of the autonomic nervous system (ANS) measures (e.g., electrodermal activity) on which they rely, but are rather related to weaknesses of the paradigm employed; most paradigms fail at isolating deception because they rely on improper control questions or stimuli. Many recent publications on the merits of neuroimaging for deception detection purposes overlook this point and fail to make proper distinctions between the various paradigms implemented to study deception and the dependent variables used to tap into deceptive behavior.

The distinction between paradigms and dependent measures is important for three reasons. First, the paradigms differ in their dem-

onstrated validity, and each paradigm may be vulnerable to different threats to validity. Second, paradigms vary greatly in what they aim to measure, with some targeting mechanisms involved in deceptive behavior (e.g., executive functioning), while others aim to detect recognition of crime-related knowledge. Third, paradigms and dependent variables may interact such that the validity of the physiological measures may depend on the paradigm used.

In this article, we will stress that the vast majority of neuroimaging literature dealing with deception detection is hampered by a lack of conceptual clarity and a host of methodological problems, primarily because it disregards the distinction between the dependent variable and the paradigm. Clearly, these problems are not unique to the neuroimaging literature, but relate more broadly to the literature on deception and its detection. Below, we will first describe the most frequently used paradigms employed in deception research. Next, we evaluate the assumptions and theories underlying these paradigms, evaluate their validity, and attempt to clarify the distinctions between research designed to shed light on the psychological mechanisms underlying deceptive behavior and research designed to validate techniques that can be applied for the detection of deception. Finally, we discuss the implications for basic and applied research, as well as the legal implications. We have no ambition to provide an exhaustive review of the available literature on the detection of deception with the ANS, ERPs, or neuroimaging-based measures. Such resources are readily available (see, e.g., National Research Council, 2003, and Meijer & Verschuere, 2015, for a discussion of polygraph testing and Gamer, 2014, and Ganis, 2014, for a discussion of neuroimaging-based lie detection). Nor are we aiming to provide elaborate directions for future research. The aim of this paper is to inform readers about the importance of the paradigm in deception research, and we hope that it will serve to guide more in-depth discussions on the scientific status, legal admissibility, and ethical issues surrounding the detection of deception.

Paradigms and Dependent Measures Used to Study Deception

Many deception studies have used ANS measures such as electrodermal activity, respiration, heart rate, and blood pressure. More recently, measures directly related to central nervous system activity, such as fMRI and ERPs, have been introduced (e.g., Farwell & Donchin, 1991; Rosenfeld, Angell, Johnson, & Qian, 1991; Rosenfeld et al., 1988; Spence et al., 2001). A number of studies have also measured behavioral responses such as reaction times (e.g., Seymour, Seifert, Mosmann, & Shafto, 2000). Studies employing these psychophysiological and behavioral measures typically employ more or less controlled paradigms, in which questions and/or stimuli are presented, often in a large number of trials.

Although a variety of paradigms have been adopted to study deception and its detection, the bulk of the research can be classified as relying on one of three paradigms. These paradigms, along with the dependent measures used in conjunction with them, are displayed in Table 1¹. They include the Comparison Question Test

1. We focus on the paradigms that have been most frequently used in deception and detection of deception research. Other paradigms have been applied by only a limited number of researchers, such as the autobiographical Implicit Association Test (Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008), and the rapid serial visual presentation task (Ganis & Patnaik, 2009). As the number of studies using these other paradigms is relatively small, we will not discuss them. We will also not discuss variations within each of the three paradigms.

Table 1. *Paradigms and Physiological Measures Used in Studies of Deception and its Detection*

Paradigm	Contrast	Usage in field	Measures
CQT	Relevant vs. comparison questions	Extensive worldwide	ANS
CIT	Correct vs. incorrect details	Limited except for Japan	ANS, ERP, RT, fMRI
DoD	Deceptive vs. truthful responses	Not used	ANS, ERP, RT, fMRI

Note. CQT = Comparison Question Test; CIT = Concealed Information Test; DoD = differentiation of deception; ANS = autonomic nervous system; RT = reaction times; ERP = event-related brain potential; fMRI = functional magnetic resonance imaging.

(CQT),² designed to detect deception by formulating direct (“Did you do it?”) relevant questions. The responses to relevant questions are compared with responses to comparison questions. The latter type of questions is focused on general, nonspecific transgressions of a nature as similar as possible to the issue under investigation (e.g., “Have you ever taken something that did not belong to you?”). To increase the relevance of the comparison questions to the innocent examinee, they are typically introduced with a cover story, for example, that admitting any wrongdoing will negatively affect the credibility of the suspect (Offe & Offe, 2007). The CQT paradigm is typically employed in combination with ANS measures (Raskin, Honts, & Kircher, 2014; Reid & Inbau, 1977), and is widely used by law enforcement agencies around the world (Meijer & van Koppen, 2008; Raskin & Honts, 2002).

The second paradigm was initially labeled the Guilty Knowledge Test (GKT, see Lykken, 1959, 1960), but is nowadays commonly referred to as the Concealed Information Test (CIT, see Verschuere, Ben-Shakhar, & Meijer, 2011). This paradigm is not a deception test per se, but rather a procedure designed to detect whether an examinee possesses pertinent information. When the test shows that an examinee has knowledge of crime-related details such as the weapon used in a murder, involvement in the crime may be inferred. In the CIT, questions presented to the examinee are followed by one relevant alternative (e.g., a feature of the crime under investigation) and several neutral (control) alternatives randomly ordered. These neutral alternatives are chosen such that an innocent suspect would not be able to discriminate them from the relevant alternative. In contrast, a suspect who is familiar with the details of the crime would be able to discriminate between the relevant and the neutral control items.

The CIT has been studied extensively with a variety of dependent measures, including ANS measures (see Gamer, 2011, for a review), ERP measures (see Rosenfeld, 2011, for a review), and reaction times (see Verschuere & de Houwer, 2011, for a review). Several CIT studies using fMRI as the dependent measure have also been published (e.g., Bhatt et al., 2009; Cui et al., 2013; Ding et al., 2012; Gamer, Bauermann, Stoeter, & Vossel, 2007; Gamer, Klimecki, Bauermann, Stoeter, & Vossel, 2012; Langleben et al., 2002, 2005; Nose, Murai, & Taira, 2009; Peth et al., 2015; Suchotzki, Verschuere, Peth, Crombez, & Gamer, 2015). Application of the CIT by law enforcement agencies is limited, except for Japan where the CIT—with ANS measures—is the standard paradigm used by the National Police in criminal investigations (for a description of CIT practices in Japan, see Osugi, 2011, and for a comparison of the [Western] laboratory use of the CIT vs. the

Japanese field application, see Ogawa, Matsuda, Tsuneoka, & Verschuere, 2015).

The third paradigm described in Table 1 is the differentiation of deception (DoD) paradigm. The DoD paradigm was originally developed by Furedy and his colleagues (e.g., Furedy, Davis, & Gurevich, 1988) to study deception by isolating the deceptive response and controlling for other confounding factors. Typically, in the DoD paradigm, examinees are presented with a series of (autobiographical) questions and are instructed to give truthful answers to half of them and deceptive answers to the other half. In a more recent variant of the DoD paradigm, labeled the Sheffield Lie Test (e.g., Spence et al., 2001), participants are typically asked to answer each question twice: once truthfully and once deceptively. Initial research using the DoD paradigm relied mainly on ANS measures (e.g., Furedy et al., 1988; Gödert, Rill, & Vossel, 2001; Vincent & Furedy, 1992; see also Bradley, MacLaren, & Black, 1996). More recently, the DoD paradigm has been widely adopted to study deception with fMRI (e.g., Kozel et al., 2004; Spence et al., 2001), and reaction times (Hu, Chen, & Fu, 2012; Van Bockstaele et al., 2012; Verschuere, Spruyt, Meijer, & Otgaar, 2011) as the dependent measures. A number of studies used the DoD paradigm with ERPs (e.g., Hu, Wu, & Fu, 2011; Johnson, Barnhardt, & Zhu, 2005; Suchotzki, Crombez, Smulders, Meijer, & Verschuere, 2015; Tu et al., 2009). The limited number of legal cases in which fMRI was employed all relied on a variant of the DoD paradigm (Miller, 2010; Spence, Kaylor-Hughes, Farrow, & Wilkinson, 2008).

Assumptions and Theory Underlying the Paradigms

CQT

The CQT is based on the assumption that deceptive examinees will perceive the relevant questions as more threatening than the comparison questions, and that relevant questions will therefore elicit larger ANS responses. Truthful examinees, on the other hand, are expected to perceive the comparison questions as more threatening than the relevant questions (Elaad, 2003; MacNeill, Bradley, Cullen, & Arseneault, 2014). Thus, larger responses to the relevant than to the comparison questions are seen as a red flag of deception, while the reverse pattern (larger responses to the comparison than to the relevant questions) is thought to reflect truth telling. The relevant and comparison questions, however, differ on many dimensions, such as significance, ambiguity, and emotional valence. Both guilty and innocent suspects may easily perceive the differences between the relevant questions specifically related to the issue under investigation and the more broadly framed comparison questions. Importantly, because the ANS parameters monitored during the typical CQT are known to be affected by these other factors, stronger responses to relevant than to comparison items cannot be solely attributed to deception (e.g., National Research Council, 2003). The CQT has also been criticized because the formulation

2. Before the CQT was developed, the relevant/irrelevant (R/I) paradigm was employed as an aid to forensic investigations in the early period of polygraph practice (see Reid, 1947). But nowadays, its obvious weaknesses have been widely recognized (Horowitz, Kircher, Honts, & Raskin, 1997), and it is no longer used for this purpose. Furthermore, insufficient research on the R/I paradigm is available and, consequently, it will not be discussed in this paper.

and presentation of the CQT questions are unstandardized, and because the test is typically conducted in a contaminated manner, that is, by examiners who have a priori knowledge of the case that might bias their interpretation of the subsequent test results (for a general analysis of such observer effects, see Risinger, Saks, Thompson, & Rosenthal, 2002). Thus, it is impossible to determine whether test outcomes actually reflect differential physiological responding to relevant and comparison questions, or prior information available to the examiner (Ben-Shakhar, Bar-Hillel, & Liebllich, 1986).

Several theoretical frameworks have been formulated to explain differential responding in the CQT (e.g., conflict theory or the conditioned response theory), but none of these accounts is convincingly supported by the data (National Research Council, 2003). For these reasons, the CQT has continuously been criticized and is considered by most researchers as lacking scientific foundation (e.g., Ben-Shakhar, 2002; Iacono & Lykken, 2002; Lykken, 1974; National Research Council, 2003).

CIT

As indicated above, the CIT is not a test of deception, but can detect only knowledge of crime-related information. The main assumption underlying this test is that for a knowledgeable suspect the relevant alternatives are significant and consequently elicit differential physiological and behavioral responses in comparison to neutral alternatives, whereas an innocent suspect is unable to distinguish between the relevant and the neutral alternatives. When critical and neutral alternatives are, indeed, indistinguishable to an innocent suspect, the CIT possesses an optimal control condition, and differential responding can be attributed only to knowledge of crime details. In addition, because innocent suspects cannot discriminate between the relevant and the neutral items, their relative responses to the relevant items cannot be affected by factors such as stress and motivation to avoid detection.

As the autonomic nervous system fluctuations used in the CIT are components of the orienting response (OR, see Lynn, 1966; Sokolov, 1963), it is not surprising that orienting theory has been proposed as a framework to understand the differential responding in the CIT. Germane to this, Sokolov (1963) and his followers noted that significant stimuli ("signal-value stimuli," to use Sokolov's terminology) elicit enhanced ORs, and this can account for the enhanced responses to the crime-relevant stimuli observed among knowledgeable (guilty) individuals. The relationship between the CIT effect and the OR was highlighted by Lykken (1974) who wrote that, "... for the guilty subject only, the 'correct' alternative will have a special significance, an added 'signal value' which will tend to produce a stronger orienting reflex than that subject will show to other alternatives" (p. 728).

There is ample evidence supporting the OR account for CIT outcomes. For example, the physiological response pattern elicited by the relevant CIT items in knowledgeable individuals (e.g., increased skin conductance response, Lykken, 1959; heart rate deceleration, Verschuere, Crombez, De Clercq, & Koster, 2004; respiratory suppression, Timm, 1982; and increased pupil dilation, Lubow & Fein, 1996) is typical for the OR. Furthermore, several features characteristic of the OR have been demonstrated using the CIT paradigm. A case in point is response habituation, which has been observed in several CIT studies (e.g., Balloun & Holmes, 1979; Ben-Shakhar, Liebllich, & Kugelmass, 1975; Gamer, Verschuere, Crombez, & Vossel, 2008; Verschuere, Crombez, De Clercq, & Koster, 2005). Related to this, and as predicted by OR

theory, differential responding has been demonstrated to increase when pertinent items are less frequently presented (e.g., Ben-Shakhar, 1977).

The CIT has also been used effectively with the P300 component of the ERPs (e.g., Meijer, Klein Selle, Elber, & Ben-Shakhar, 2014; Rosenfeld, 2011). While the question of whether the P300 is an OR measure has been debated in the literature (Donchin et al., 1984), it is definitely affected by stimulus qualities that elicit ORs, notably stimulus novelty and significance (e.g., Donchin, 1981). Both 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).

Besides orienting, response inhibition (i.e., suppression of dominant responses) can partially explain the differential responding in the CIT (Verschuere & Ben-Shakhar, 2011). Data supporting the idea that a combination of orienting and inhibition underlies the responses observed in the CIT comes from fMRI studies. Gamer (2014) recently argued that the pattern of brain activation in CIT studies primarily reflects the engagement of cognitive mechanisms that are associated with (a) attentional orienting toward relevant alternatives, (b) response inhibition (withholding the prepotent truth response), and (c) selection and planning of the deceptive behavioral response. Importantly, the notion that apart from the OR inhibition plays a role fits with both ANS results and P300 results, as the amplitude of the latter has also been shown to be sensitive to inhibition (Polich, 2007). Two recent studies (Klein Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2015; Suchotzki, Verschuere et al., 2015) disentangled the role of orienting and response inhibition in the CIT, and showed that skin conductance responding could best be explained by orienting theory, while reaction time, heart rate, respiration, and fMRI responses are affected by response inhibition.

In sum, the CIT is a valid paradigm for detecting concealed knowledge. However, for studying deception, the CIT is confounded by a frequency effect. That is, with each question, only one relevant alternative that potentially elicits deceptive behavioral responses is presented, while multiple neutral alternatives—associated with a truthful behavioral response—are presented. An additional confound when trying to isolate deception is that for knowledgeable examinees the relevant alternative has been previously encoded in memory, while the neutral alternatives have not. For these two reasons, the CIT is ill-suited for studying mechanisms underlying deception *per se*.

DoD

The DoD paradigm was designed to examine deception while controlling for all other potentially confounding factors. Because deceptive and truthful responses are elicited equally often, it is not confounded by frequency. In addition, implementations of the DoD paradigm (e.g., Spence et al., 2001) required subjects to respond both deceptively and truthfully to each question (i.e., on a within-subject basis), thereby effectively controlling for the level of significance. Because of this control, any difference in responding between the questions can be attributed solely to deception.

Theoretical accounts that have been offered for the effects found in the DoD paradigm are primarily based on fMRI data, and show that deceptive responses are associated with increased cognitive effort, particularly response inhibition. As noted by Gamer (2014), results of the various studies in this domain yield an activation pattern scattered across the brain, showing that deception is not related to a unique brain region. As is true for the CIT,

activated regions have been linked in previous studies to a number of cognitive mechanisms including response monitoring, cognitive control, response inhibition, and memory. Moreover, the reaction time slowing associated with deceptive responses is consistent with cognitive control (Debey, Verschuere, & Crombez, 2012) and inhibition (Debey, Ridderinkhof, De Houwer, De Schryner, & Verschuere, 2015; but see Verschuere, Schuhmann, & Sack, 2012), playing an important role in the DoD paradigm.

Our discussion of the various paradigms leads us to conclude that, from a forensic application perspective, only the CIT has the demonstrated potential of a scientifically based method. However, the CIT is not suited to shed light on the neurocognitive factors underlying deceptive behavior, and the DoD has the best methodological rigor for enhancing our theoretical understanding of deception.

Accuracy

One of the most important questions—especially from an applied perspective—is “to what extent can a specific paradigm combined with a specific dependent measure differentiate between deceptive and truthful participants?” Accuracy can be expressed in several ways. Many studies report their results in terms of correct detection rates of deceptive (sensitivity) and nondeceptive (specificity) examinees. But because the meaning of the proportion of correctly identified persons depends on the specific cutoff point on the detection measure that was employed in a particular study, this parameter is not very useful for evaluating criterion validity and defies comparisons of detection rates across different studies. One way of dealing with this problem is by using signal detection measures. Indeed, a signal detection approach was recommended by the NRC report (National Research Council, 2003) and adopted by many researchers (e.g., Ben-Shakhar, 1977; Ben-Shakhar & Gati, 1987; Ben-Shakhar, Lieblich, & Kugelmass, 1970; Gamer et al., 2008; Rosenfeld, Soskins, Bosh, & Rayan, 2004).

The signal detection approach provides measures of detection efficiency that do not depend on a single, arbitrary, cutoff point. Rather, statistics are calculated by comparing the entire distributions of the detection scores of guilty (or deceptive) and innocent (or truthful) participants. Based on these distributions, a receiver operating characteristic (ROC) curve can be generated and the area under this ROC curve (a) represents the detection efficiency regardless of any specific cutoff point (for a detailed description of generating ROC curves in CIT experiments, see Lieblich, Kugelmass, & Ben-Shakhar, 1970). The area under the ROC curve ranges between 0 and 1, such that an area of 0.5 means that the two distributions (i.e., the detection score’s distributions for guilty and innocent examinees) are indistinguishable (i.e., detecting whether an examinee is deceptive or not will be at chance level). An area of 1 means that there is no overlap between the two distributions and thus a perfect classification is possible. To allow for comparisons between paradigms and dependent measures, we report the a statistic.³ For studies reporting only sensitivity and specificity, we transformed these values to a using the formula proposed by Grier (1971).

CQT

For the CQT in combination with ANS measures, the best estimate of the criterion validity can be derived from the 2003 report by the

NRC (National Research Council, 2003). The Council evaluated 37 laboratory CQT studies and found a median a value of .85. Accordingly, it concluded that “in populations of examinees such as those represented in the polygraph research literature . . . specific-incident polygraph tests can discriminate lying from truth telling at rates well above chance, though well below perfection” (p. 4). The NRC did, however, also point out that, for the bulk of the CQT research, it is questionable whether its results translate to a real-life situation, a point to which we will return. As the CQT has been exclusively combined with ANS measures, no data are available for any other dependent measure.

CIT

The accuracy of the CIT has been extensively studied in laboratory experiments, with a variety of measures. A recent meta-analysis (Meijer et al., 2014) reported accuracy estimates for three ANS measures (skin conductance response, respiration line length, and heart rate), as well as for the P300 ERP measure. Average a ’s were .85, .77, .74, and .88, respectively. Importantly, accuracy of the P300 was similar to that of skin conductance in the forensically relevant mock crime paradigm.

To estimate the criterion validity of the CIT with reaction times, we rely on a total of nine studies that included a group of participants knowledgeable of crime details and a group of unknowledgeable participants, and reported data from which a could be derived (Hu, Evans, Wu, Lee, & Fu, 2013; Kleinberg & Verschuere, 2015; Noordraven & Verschuere, 2013; Rosenfeld et al., 2004; Sai, Zhou, Ding, Fu, & Sang, 2014; Verschuere & Kleinberg, 2015; Verschuere, Kleinberg, & Theocharidou, 2015; Visu-Petra, Miclea, & Visu-Petra, 2012; Visu-Petra, Varga, Miclea, & Visu-Petra, 2013). Taken together, these nine studies reveal a weighted average a of .82

Finally, only four studies assessed the criterion validity of the fMRI-based CIT, using both crime-knowledgeable and unknowledgeable participants (Cui et al., 2013; Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011; Nose et al., 2009; Peth et al., 2015). Collectively, these studies yield a weighted average a value of .94. The a values for the different dependent measures in the CIT are given in Table 2.

DoD

Furedy and his colleagues (Furedy et al., 1988; Furedy, Gigliotti, & Ben-Shakhar, 1994; Furedy, Posner, & Vincent, 1991; Vincent & Furedy, 1992) were the first to demonstrate that deceptive answers elicited enhanced skin conductance responses in a DoD paradigm (see also Gödert et al., 2001). As explained above, in the DoD paradigm, deception and truth telling are manipulated within subjects. As a result, these studies do not report sensitivity and specificity and therefore do not allow for calculating the a statistic. Similarly, neither reaction time nor ERP studies are available that compare deceptive with truthful participants. Several studies investigated the criterion validity of the DoD using fMRI as the dependent measure, but due to the within-subject comparison, most studies report only sensitivity. Only one study reported both sensitivity and specificity allowing for a derivation of a . Kozel et al. (2009) found 93% sensitivity and 38% specificity, corresponding to an a of .79.

In sum, only for the CIT paradigm are reasonably sufficient data available to allow for a comparison between the various dependent measures. This comparison shows that accuracy does not differ much between measures (see also Verschuere, Crombez,

3. This a measure is also referred to as the AUC.

Table 2. Overview of the Detection Accuracy of the Different Measures in the CIT

Measure	<i>n</i>	<i>a</i>
SCR		
Meta-analysis by Meijer et al. (2014)	3,863	.85
RT		
Hu et al. (2013)	63	.91
Kleinberg & Verschuere (2015) Study 1	202	.78
Kleinberg & Verschuere (2015) Study 2	212	.80
Noordraven & Verschuere (2013)	42	.87
Verschuere et al. (2015) Study 2, single probe condition	116	.69
Verschuere et al. (2015) Study 2, multiple probe condition	94	.86
Verschuere & Kleinberg (2015)	73	.98
Visu-Petra et al. (2012)	40	.92
Visu-Petra et al. (2013)	73	.86
Rosenfeld et al. (2004)	22	.93
Sai et al. (2014)	44	.73
Total <i>n</i>/weighted <i>a</i>	981	.82
ERP		
Meta-analysis by Meijer et al. (2014)	646	.88
fMRI		
Cui et al. (2014)	32	.88
Ganis et al. (2011)	24	1
Nose et al. (2009)	38	.90
Peth et al. (2015)	40	.98
Total <i>n</i>/weighted <i>a</i>	134	.94

Note. Numbers in bold denote the total *n* and the weighted *a*. *n* = total number of participants; *a* = area under the ROC curve; SCR = skin conductance response; RT = reaction times; ERP = event-related brain potentials; fMRI = functional magnetic resonance imaging.

Degrootte, & Rosseel, 2010). Two important considerations should be noted here. First, the estimate of the *a* value for ANS measures in Table 2 relies only on the SCR measure. Given that combining several ANS measures has been shown to increase accuracy (Gamer et al., 2008), this estimate of *a* should be treated as the lower bound accuracy estimate of ANS measures in the CIT. Second, it should be noted here that the accuracy estimate of the fMRI-based CIT seems slightly higher compared to the other measures but it is based on relatively few participants, and none of the four studies performed a cross validation on an independent sample, meaning that the accuracy estimate might be an overestimation.

Reverse Inference and Countermeasures

The assumptions and theories underlying the different paradigms described earlier make clear that none of them assesses deception directly. In the CQT, deception is inferred from emotional arousal, in the DoD from cognitive control and inhibition, while in the CIT recognition of crime-relevant information is inferred from an enhanced orienting response, potentially combined with response inhibition. The problems associated with such inferences have been discussed extensively in the fMRI literature as the *fallacy of the reverse inference* (Poldrack, 2006; Sip, Roepstorff, McGregor, & Frith, 2007). That is, even if deceptive responses are differentially associated with brain activation in areas associated with cognitive control, we cannot conclude that differential activation in these areas necessarily implies that the subject is deceptive (i.e., responses to questions may be associated with enhanced cognitive control even when they are truthful). Similarly, the fallacy of reverse inference applies to the absence of differential activation: a lack of activation in areas associated with inhibition does not necessarily imply that the subject is responding truthfully.

Of course, the fallacy of the reverse inference problem is not unique to brain imaging studies, and other dependent measures (e.g., skin conductance or heart rate measures) are similarly vulnerable to it. In fact, logically this fallacy is at the core of the shortcoming of the CQT: deception may be associated with emotional arousal, but this does not mean that the presence of emotional arousal necessarily indicates the subject is deceptive.

The reverse inference fallacy is closely related to another threat, namely, that of countermeasures. The theoretical account of the CIT, for example, implies that neither ANS measures nor the P300 are necessarily direct indicators of the presence of a memory trace (Satel & Lilienfeld, 2013). They reflect stimulus significance and/or deviance, and only under the assumption that all alternatives are equally plausible can the presence of a memory trace be inferred. As a consequence, both the P300 and the ANS-based CIT are susceptible to countermeasures (for a review, see Ben-Shakhar, 2011). For example, Rosenfeld and his colleagues showed that any of the irrelevant items in a CIT can be made significant by giving simple instructions such as “wiggle your toe upon presentation of this stimulus.” As a result, irrelevant items become significant and the P300 amplitude elicited by these items increases, thereby reducing diagnostic accuracy (Rosenfeld et al., 2004). A more recent P300-based CIT protocol (the complex trial protocol; Rosenfeld et al., 2008) has been shown to be resistant to these specific types of countermeasures (see also Meixner & Rosenfeld, 2010; Rosenfeld & Labkovsky, 2010). Yet, recent research has shown that simple memory suppression instructions reduced the P300 to mock crime information in this protocol (Hu, Bergstrom, Bodenhausen, & Rosenfeld, 2015, but see Rosenfeld Ward, Drapekin, Labkovsky & Tullman, 2015). Importantly, Ganis et al. (2011) demonstrated that comparable countermeasures also worked in an fMRI setting and reduced CIT detection accuracy from 100% to only 33%. This illustrates nicely that the reverse inference problem discussed earlier applies to the fMRI-based CIT as well.

From the Lab to the Field

In the section about accuracy, we noted that the NRC pointed out that it is questionable whether the results reported in the bulk of the CQT research would translate into real-life situations. Data showing that when the CQT is used in mock crime research with participants (typically college students) who have nothing to lose, the comparison questions are indeed perceived as more threatening than the relevant questions (e.g., MacNeill et al., 2014) by no means imply that this would also be the case for real suspects.

To determine the accuracy in the field, properly executed field studies are necessary. An inherent problem to field studies, however, is selection bias (Begg & Greenes, 1983; Iacono, 1991; Patrick & Iacono, 1991). This selection bias occurs because the limited number of field studies on both the ANS-based CQT (e.g., National Research Council, 2003), and ANS-based CIT (Elaad, 1990; Elaad, Ginton, & Jungman, 1992) have relied on confessions made by suspects at a later point in time as a ground truth criterion. However, in many cases, such confessions are intimately related to the outcomes of the deception test. That is, a confession is typically obtained only in interrogations that follow positive (i.e., deception-indicated) test outcomes. Thus, the basic requirement of an independent criterion of ground truth is not met in this research area, and, as a consequence, the accuracy estimates generated by these studies might be inflated. A case in point here is Iacono (1991), who showed that when using confessions elicited after a failed test as a measure of ground truth, even a chance level accurate

procedure yields near perfect accuracy (see also Iacono & Lykken, 2002). In addition, the contaminated administration of the CQT in the field confounds the outcomes of this test with prior information available to the investigators (Ben-Shakhar et al., 1986). In sum—despite the extensive field use of the CQT worldwide, and the CIT in Japan—no solid field studies exist, while for the DoD paradigm, no field studies have been conducted for any of the dependent measures.

For the CIT, external validity is also a concern. Importantly, the CIT rests on the assumption that (a) individuals committing a crime will remember several critical details, and (b) these details are unknown to innocent suspects. These assumptions can be easily checked in laboratory experiments, but in real life, where crimes are often committed under great stress and time constraints, it may be less straightforward to determine whether the perpetrator actually perceived and encoded all crime-related items. Also, as the test usually takes place weeks and sometimes months or even years after the crime, it is doubtful whether all crime-related items will be remembered during the test. Indeed, several recent laboratory studies using ANS measures revealed that, when the CIT is administered 1 or 2 weeks after the mock crime, certain critical items are not recalled and fail to elicit differential responses (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Gamer, Kosiol, & Vossel, 2010; Nahari & Ben-Shakhar, 2011; Peth, Vossel, & Gamer, 2012). Needless to say, the problem of false negatives caused by presenting critical details that are not remembered by the suspect will not be resolved by changing from ANS to P300 or fMRI measures.

The CIT also requires that none of the critical items have been leaked either through the media or during the investigation. Several studies, most of which were conducted by Bradley and his colleagues, demonstrated how information leakage may compromise the outcomes of the CIT (see Bradley, Barefoot, & Arsenault, 2011, for a review). False positive errors caused by information leakage equally apply to the P300 (Winograd & Rosenfeld, 2014), and there is no evidence showing that fMRI can tackle this problem any better than traditional ANS measures or the P300 (Peth et al., 2015).

The DoD paradigm has been examined only in experiments conducted under conditions that are substantially different from the typical forensic situation. For example, in most DoD studies, participants are asked to give deceptive answers to simple semantic (e.g., “Is Rome the capital of Italy?”) or autobiographic questions (e.g., “Is your name X?”). These items are markedly different from relevant questions that may be presented in a criminal investigation (e.g., “Did you murder Mr. X?”). For example, they differ in their capacity to elicit emotional arousal and motivation to avoid detection. Importantly, research using the DoD with reaction times has shown that lying becomes easier with practice (Hu et al., 2012; Van Bockstaele et al., 2012; Verschuere et al., 2011). As it is likely that, in real life, suspects are extensively repeating their lies, the results of laboratory studies may not generalize to field applications.

As outlined above, the DoD paradigm is the only paradigm that attempts to isolate deceptive responding. Translating the findings from DoD research to the field is, however, cumbersome for another reason. Real-life deception typically contains an element of choice. Yet, in DoD research, participants are instructed by the experimenter to deceive, meaning that the essential choice component is missing (see also Farah et al., 2014; Kanwisher, 2009; Sip et al., 2007). Some DoD studies (e.g., Furedy, et al., 1994; Sip et al., 2010; Spence et al., 2008) tried to overcome this difficulty

by giving participants a choice regarding the specific questions to which they would give deceptive responses, although they had to give deceptive responses to approximately half of the questions. However, even under these artificial choice conditions, where participants are instructed to deceive to some questions, deceptive responses of experimental subjects are unlikely to match real deceptive behavior in natural social interactions, where individuals freely choose whether to lie or tell the truth.

Implications

We emphasized the distinction between paradigms designed to shed light on the mechanisms underlying deception by isolating deceptive behavior and those designed to diagnose guilt through knowledge. Below, we will discuss the implications of this distinction for both basic research and applications.

Implications for Basic Research

As we argued above, the CQT and the CIT are ill-suited to study the mechanisms that underlie deception. The CQT is unsuitable because it lacks the basic requirement of proper controls, whereas the CIT is ill-suited because deceptive responses are confounded by frequency (i.e., deceptive responses occur on the minority of trials) and memory effects (i.e., deceptive responses occur only to information present in memory). The DoD paradigm, on the other hand, was designed to examine deceptive behavior while controlling for confounding factors, but it is questionable whether the findings translate to realistic situations mainly because one could argue that participants follow instructions rather than actually respond deceptively. A number of recent studies have already attempted to resolve these issues by having participants choose whether to be deceptive or not (e.g., Kozel et al., 2005; Sip et al., 2010, 2012), but as of yet, construct validity (i.e., whether the construct of deception is captured by this paradigm) of the DoD still remains questionable.

Applied Implications

Because the CQT lacks proper controls, it is ill-suited to distinguish between guilty and innocent suspects. The DoD procedure may be adopted for this purpose, but currently there is no sufficient evidence regarding the accuracy of discrimination between deceptive and truthful subjects with this particular paradigm. The CIT, on the other hand, is an effective paradigm to discriminate between people with knowledge of intimate crime details and those without such knowledge. As long as innocent suspects are unable to discriminate between the relevant and the neutral items, they will not show a differential pattern of behavioral or physiological responses to the relevant items, irrespective of any potential confounding psychological mechanism unrelated to deception (e.g., being nervous). Several recommendations for dealing with threats to the validity of the CIT and for increasing its forensic application were made by Ben-Shakhar (2012).

One of the factors that seem to hinder a large-scale application of the CIT is its limited applicability (e.g., Krapohl, 2011; Podlesney, 1993, 2003). Whereas in the CQT and the DoD questions about virtually any event can be formulated (e.g., “Did you shoot X?” in the CQT or “I shot X” in the DoD), the CIT can only be applied if sufficient details known exclusively to the offender are present. This may be true for well-planned offences, but is less likely to be the case in impulsively committed crimes. Importantly,

employing the P300 or fMRI, rather than ANS measures, will not resolve the limited applicability of the CIT due to insufficient probe material.

A practical consideration is that both the CIT and the DoD paradigms can be combined with easy and cost-effective measures such as ANS parameters and reaction times. Table 2 demonstrates that, in the CIT, the detection accuracy of fMRI is largely similar to that obtained with ERPs (Meijer et al., 2014) and to accuracy rates based on ANS measures. Thus, given that ANS, ERP, and reaction time data can be obtained much more easily and cheaply than fMRI data and that the former measures can be acquired in many participants that cannot undergo fMRI testing (e.g., because of claustrophobia or metallic implants), it is not to be expected that fMRI would have practical utility as a field detection measure in the near future.

Legal Implications

Because of its shortcomings outlined above, the CQT has no scientific validity, and should not be admitted as evidence in court. Gallai (1999) as well as Saxe and Ben-Shakhar (1999) discussed the admissibility of the CQT in light of the four Daubert criteria (testability, or falsifiability), error rate, peer review and publication, and general acceptance (*Daubert v. Merrell Dow Pharmaceuticals Inc.*, 1993, 113 S. Ct. Supp. 2786) and reached similar conclusions. Gallai concluded that “the results of polygraph examinations as practiced today (i.e., the CQT) should not be admissible evidence in federal courts” (Gallai, 1999, p. 88). In analyzing the four Daubert criteria, Gallai (1999) demonstrated that the CQT does not satisfy these criteria, with the possible exception of peer review and publications (but it should be noted that many of the publications related to the CQT are critiques of the technique). Saxe and Ben-Shakhar (1999) focused on the modern concept of scientific validity (e.g., Messick, 1995) and demonstrated that the CQT lacks several critical components of construct validity and discriminant validity. Given that all these threats to the validity of the CQT are related to weaknesses of the paradigm, they will not be resolved by using fMRI rather than ANS measures in a CQT.

In contrast to the CQT, both the CIT and the DoD paradigm have the potential of meeting the Daubert criteria, yet the crucial component lacking is research on their accuracy in field situations. Although the outcomes of the CIT can serve, under certain conditions, as admissible evidence in judicial proceedings in Japan (Nakayama, 2002), they do not meet one of the four Daubert criteria because error rates in realistic situations remain unknown (Ben-Shakhar & Kremnitzer, 2011; Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013).

The Daubert criteria and related guidelines (*Kumho Tire Co., Ltd. v. Carmichael*, 1999, 1526 U.S. 137) allow for a relative straightforward discussion of admissibility of scientific evidence. In many other legal systems, admissibility may be less straightforward. A case in point here is a review about the admissibility of emerging forensic neuroscience technologies in Canada (Frederiksen, 2011). This author attempts to answer the question whether “brain fingerprinting” (i.e., the P300-based CIT) and fMRI lie detection escape the prohibition imposed by jurisprudence regarding the polygraph. Frederiksen concludes that “since the fMRI technique claims to detect falsehood, it is likely inadmissible due to its similarity to the polygraph: both claim to detect lies” (p. 127), and “While the current ban on the polygraph would apply to any technique that is explicitly a lie detector, brain fingerprinting, however, does not detect lying but instead the presence of memory.

This difference may allow it to bypass the broad prohibition on lie detectors” (p. 115). As our review above has pointed out, however, detecting memory of crime scene details can also be done with fMRI or other measures. Frederiksen’s confusion of technique and procedure highlights the importance of taking into account the paradigm for in-depth discussions about legal admissibility. Noteworthy here is also the U.S. Supreme Court’s decision, *United States v. Scheffer* (2003). In this case, the court decided that credibility determinations is the jury’s task, rendering a CQT polygraph test inadmissible (see also *Wilson v. Corestaff Services, L.P.*, 2010). If one were to follow this line of reasoning, the DoD would be inadmissible as well because it detects deception, whereas the CIT could be admissible as it detects concealed knowledge rather than deception (Rosenfeld et al., 2013).

Methodologically sound field studies are essential to establish accurate error rates. But what do such field studies look like? Research has shown that prior information about the guilt of a suspect—for example, about a confession—influences the outcomes of subsequent forensic tests, including the outcomes of lie detection tests (e.g., Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2014; Elaad, Ginton, & Ben-Shakhar, 1994; Kassin, Dror, & Kukucka, 2013). Given the large number of degrees of freedom, especially when it comes to interpreting fMRI data, analysis of deception detection results may be prone to biasing observer effects. Kassin (2012) argues that these effects also occur in the other direction: when suspects (falsely) confess, this confession taints the interpretation and collection of subsequent evidence, in turn corroborating the (false) confession. The literature on the observer biases surrounding confessions then illustrates the importance of an independent measure of ground truth: Even if the outcome of a test is not used directly to elicit a confession, as is standard practice with the forensic application of the CIT in Japan, dependence between test outcome and ground truth may still be present (investigative authorities may, for example, invest more resources in crime scene analysis once a suspect failed a test). Ideally, to prevent such dependence, the test should be conducted before any information relevant to the ground truth has been collected, and the test outcome should only be determined after the ground truth has been established independently. Recently, Langleben and Moriarty (2013) called for clinical trials of fMRI deception tools. Clinical trials may be informative, but without a clear specification of the paradigm that will be employed, as well as a specification of how to deal with the dependence of the test outcome and the ground truth, such trials are destined to result in a discussion similar to that about the accuracy of the CQT.

Finally, to illustrate the problems with internally and externally valid studies, we would like to mention the study by Ginton, Daie, Elaad, and Ben-Shakhar (1982). These authors administered an aptitude test to policemen, who were under the impression that failure to perform well on this test would have severe consequences for their career. Participants could cheat, but were unaware that this cheating could be detected by the experimenters. As such, this study demonstrated how an ideal study to examine the validity of lie detection tests under realistic settings could be designed: individuals freely chose whether to lie or not and ground truth criterion was available independently of the test’s outcomes—and consequences were perceived to be severe. However, under this realistic setup, most guilty participants either refused to take the test or confessed before taking the test and consequently the results could not yield any reliable conclusions. In addition, it is doubtful whether this type of study would be approved by the current standards of ethics committees and institutional review boards (see also Rosenfeld et al., 2013).

Conclusion

Great hopes and expectations were expressed regarding the potential use of brain imaging techniques for the detection of deception. Contrary to what has been advocated by many researchers as well as practitioners (e.g., Bles & Haynes, 2008; Farwell, 2012; Langleben et al., 2005), the introduction of new measures such as P300 and fMRI is by no means a solution to the problems associated with the ANS-based CQT polygraph test. The CQT has been criticized for lacking proper controls and being unstandardized. In addition, its outcome is often contaminated by prior information available to the examiner. None of these criticisms can be resolved by replacing ANS recordings with fMRI measures.

Moreover, all paradigms face a similar logical problem: deception cannot be directly inferred either from the presence of

emotional arousal in the CQT or from attentional orienting or inhibition in the CIT or DoD, regardless of whether ANS, reaction times, ERPs, or fMRI measures have been used.

We discussed several methodological and conceptual shortcomings associated with the paradigms designed to study deception and detection of deception. Importantly, these shortcomings reflect the research paradigms used, and not the dependent measures. New paradigms may be able to overcome these shortcomings—at least to some extent. However, to prevent superficial discussions driven by the illusion of explanatory depth provided by fMRI scans (Brown & Murphy, 2009; Rhodes, Rodriguez, & Shah, 2014), academic, legal, and ethical communities should consider the validity of the paradigms, including the various threats to validity, rather than focusing on the technology of the dependent measures.

References

- Balloun, K. D., & Holmes, D. C. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology, 64*, 316–322. doi: 10.1037/0021-9010.64.3.316
- Ben-Shakhar, G. (1977). A further study of the dichotomization theory in detection of information. *Psychophysiology, 14*, 408–413. doi: 10.1111/j.1469-8986.1977.tb02974.x
- Ben-Shakhar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103–126). Waltham, MA: Academic Press
- Ben-Shakhar, G. (2011). Countermeasures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer, E. *Memory detection: Theory and application of the Concealed Information Test*. Cambridge, UK: Cambridge University Press.
- Ben-Shakhar, G. (2012). Current research and potential application of the Concealed Information Test: An overview. *Frontiers in Psychology, 3*, Article 342. doi: 10.3389/fpsyg.2012.00342
- Ben-Shakhar, G., Bar-Hillel, M., & Lieblich, I. (1986). Trial by polygraph: Scientific and juridical issues in lie detection. *Behavioral Sciences and the Law, 4*, 459–479. doi: 10.1002/bsl.2370040408
- Ben-Shakhar, G., & Gati, I. (1987). Common and distinctive features of verbal and pictorial stimuli as determinants of psychophysiological responsivity. *Journal of Experimental Psychology: General, 116*, 91–105. doi: 10.1037/0096-3445.116.2.91
- Ben-Shakhar, G., & Kremnitzer, M. (2011). The CIT in the courtroom: Legal aspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer. *Memory detection: Theory and application of the Concealed Information Test* (pp. 276–290). Cambridge, UK: Cambridge University Press.
- Ben-Shakhar, G., Lieblich, I., & Kugelmass, S. (1970). Guilty knowledge technique: of application of signal detection measures. *Journal of Applied Psychology, 54*, 409–413. doi: 10.1037/h0029781
- Ben-Shakhar, G., Lieblich, I., & Kugelmass, S. (1975). Detection of information and GSR habituation: An attempt to derive detection efficiency from two habituation curves. *Psychophysiology, 12*, 283–288. doi: 10.1111/j.1469-8986.1975.tb01291.x
- Begg, C. B., & Greenes, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics, 207*–215. doi: 10.2307/2530820
- Bhatt, S., Mbwana, J., Adeyayemo, A., Sawyer, A., Hailu, A., & Vannmeter, J. (2009). Lying about facial recognition: An fMRI study. *Brain and Cognition, 69*, 382–390. doi: 10.1016/j.bandc.2008.08.033
- Bles, M., & Haynes, J. D. (2008). Detecting concealed information using brain-imaging technology. *Neurocase, 14*, 82–92. doi: 10.1080/13554790801992784
- Bogaard, G., Meijer, E. H., Vrij, A., Broers, N. J., & Merckelbach, H. (2014). Contextual bias in verbal credibility assessment: Criteria-based content analysis, reality monitoring and scientific content analysis. *Applied Cognitive Psychology, 28*, 79–90. doi: 10.1002/acp.2959
- Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In B. Verschuere, G. Ben-Shakhar, & E. Meijer *Memory detection: Theory and application of the Concealed Information Test* (pp. 187–199). Cambridge, UK: Cambridge University Press.
- Bradley, M. T., MacLaren, V. V., & Black, M. E. (1996). The control question test in polygraphic examinations with actual controls for truth. *Perceptual and Motor Skills, 83*, 755–762. doi: 10.2466/pms.1996.83.3.755
- Brown, T., & Murphy, E. (2009). Through a scanner darkly: Functional neuroimaging as evidence of a criminal defendant's past mental states. *Stanford Law Review, 62*, 1119–1208.
- 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. doi: 10.1037/1076-898X.9.4.261
- Cui, Q., Vanman, E. J., Dongtao, W., Yang, W., Lei, J., & Zhangg, Q. (2014). Detection of deception based on fMRI activation patterns underlying the production of a deceptive response and receiving feedback about the success of the deception after a mock murder crime. *Social Cognitive and Affective Neuroscience, 9*, 1472–1480. doi: 10.1093/scan/nst134
- Debey, E., Ridderinkhof, R. K., De Houwer, J., De Schryver, M., & Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: Delta plots reveal the role of response inhibition in lying. *Consciousness and Cognition, 37*, 148–159. doi: 10.1016/j.concog.2015.09.005
- Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation using ego depletion and goal neglect. *Acta Psychologica, 140*, 133–141. doi: 10.1016/j.actpsy.2012.03.004
- Ding, X. P., Du, X., Lei, D., Hu, C. S., Fu, G., & Chen, G. (2012). The neural correlates of identity faking and concealment: An fMRI study. *PLoS ONE, 7*, e48639. doi: 10.1371/journal.pone.0048639
- Donchin, E. (1981). Surprise! ... surprise? *Psychophysiology, 18*, 493–513. doi: 10.1111/j.1469-8986.1981.tb01815.x
- Donchin, E., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Ohman, A., & Siddle, D. (1984). Cognition and event-related potentials II. The orienting reflex and P300. *Annals of the New York Academy of Sciences, 425*, 39–57. doi: 10.1111/j.1749-6632.1984.tb23522.x
- Editorial. (2008) *Nature Neuroscience, 11*, 1231. doi: 10.1038/nn1108-1231
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology, 75*, 521–529. doi: 10.1037/0021-9010.75.5.521
- Elaad, E. (2003). Is the inference rule of the “control question polygraph technique” plausible? *Psychology, Crime and Law, 9*, 37–47. doi: 10.1080/10683160308143
- Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making, 7*, 279–292. doi: 10.1002/bdm.3960070405
- Elaad, E., Ginton, A., & Jungman N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77*, 757–767. doi: 10.1037/0021-9010.77.5.757
- Farah, M. J., Hutchinson, B., Phelps, E. A. & Wagner, A. D. (2014). Functional MRI-based lie-detection: Scientific and societal challenges. *Nature Reviews Neuroscience, 15*, 123–131. doi: 10.1038/nrn3665

- Farwell, L. A. (2012). Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, *6*, 115–154. doi: 10.1007/s11571-012-9192-2
- Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (“lie detection”) with event-related potentials. *Psychophysiology*, *28*, 531–547. doi: 10.1111/j.1469-8986.1991.tb01990.x
- Frederiksen, S. (2011). Brain fingerprint or lie detector: Does Canada’s polygraph jurisprudence apply to emerging forensic neuroscience technologies? *Information & Communications Technology Law*, *20*, 115–132. doi: 10.1080/13600834.2011.578930
- Frith, C. D., & Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological Psychology*, *17*, 27–39. doi: 10.1016/0301-0511(83)90064-9
- Furedy, J. J., Davis, C., & Gurevich, M. (1988). Differentiation of deception as a psychological process: A psychophysiological approach. *Psychophysiology*, *25*, 683–688. doi: 10.1111/j.1469-8986.1988.tb01908.x
- Furedy, J. J., Gigliotti, F., & Ben-Shakhar, G. (1994). Electrodermal differentiation of deception: The effect of choice versus no choice of deceptive items. *International Journal of Psychophysiology*, *18*, 13–22. doi: 10.1016/0167-8760(84)90011-4
- 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*, 91–97. doi: 10.1016/0167-8760(91)90376-9
- Gallai, D. (1999). Polygraph evidence in federal courts: Should it be admissible? *American Criminal Law Review*, *36*, 87–116.
- Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer. *Memory detection: Theory and application of the Concealed Information Test* (pp. 27–45). Cambridge, UK: Cambridge University Press.
- Gamer, M. (2014). Mind reading using neuroimaging: Is this the future of deception detection? *European Psychologist*, *19*, 172–183. doi: 10.1027/1016-9040/a000193
- 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. doi: 10.1002/hbm.20343
- Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., & Vossel, G. (2012). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Social Cognitive and Affective Neuroscience*, *7*, 506–515. doi: 10.1093/scan/nsp005
- Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the guilty action test. *Biological Psychology*, *83*, 101–107. doi: 10.1016/j.biopsycho.2009.11.005
- Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Psychology & Behavior*, *95*, 333–340. doi: 10.1016/j.physbeh.2008.06.011
- Ganis, G. (2014). Deception detection using neuroimaging. In P. A. Granhag, A. Vrij, & B. Verschuere, *Detecting deception: Current challenges and cognitive approaches*. Chichester, UK: John Wiley & Sons, Ltd.
- Ganis, G., & Patnaik, P. (2009). Detecting concealed knowledge using a novel attentional blink paradigm. *Applied Psychophysiology and Biofeedback*, *34*, 189–196. doi: 10.1007/s10484-009-9094-1
- 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*, 312–319. doi: 10.1016/j.neuroimage.2010.11.025
- Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real life situation. *Journal of Applied Psychology*, *67*, 131–137. doi: 10.1037/0021-9010.67.2.131
- Gödert, H. W., Rill, H. G., & Vossel, G. (2001). Psychophysiological differentiation of deception: The effects of electrodermal lability and mode of responding on skin conductance and heart rate. *International Journal of Psychophysiology*, *40*, 61–75. doi: 10.1016/S0167-8760(00)00149-5
- Grier, J. B. (1971). Non-parametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429. doi: 10.1037/h0031246
- Grubin, D., & Madsen, L. (2005). Lie detection and the polygraph: A historical review. *Journal of Forensic Psychiatry & Psychology*, *16*, 357–369. doi: 10.1080/14789940412331337353
- Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, *34*, 108–115. doi: 10.1111/j.1469-8986.1997.tb02421.x
- Hu, X., Bergstrom, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, 1–9. doi: 10.1177/0956797615575734
- Hu, X., Chen, H., & Fu, G. (2012). A repeated lie becomes a truth? The effect of intentional control and training on deception. *Frontiers in Psychology*, *3*, 488. doi: 10.3389/fpsyg.2012.00488
- Hu, X. Q., Evans, A., Wu, H. Y., Lee, K., & Fu, G. Y. (2013). An interfering dot-probe task facilitates the detection of mock crime memory in a reaction time (RT)-based concealed information test. *Acta Psychologica*, *142*, 278–285. doi: 10.1016/j.actpsy.2012.12.006
- Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self-and other-referential information: An ERP study. *Brain Research*, *1369*, 149–157. doi: 10.1016/j.brainres.2010.10.106
- Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (pp. 201–201). Greenwich, CT: JAI Press.
- Iacono, W. G., & Lykken, D. T. (2002). The scientific status of research on polygraph techniques: The case against 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. 483–538). St. Paul, MI: West Publishing, Co.
- Johnson, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research*, *24*, 386–404. doi: 10.1016/j.cogbrainres.2005.02.011
- Kanwisher, N. (2009). The use of fMRI in lie detection: What has been shown and what has not. In E. Bizzi, S. E. Hyman, M. E. Raichle, N. Kanwisher, E. A. Phelps, S. J. Morse . . . H. T. Greely, *Using imaging to identify deceit: Scientific and ethical questions* (pp. 7–13). Cambridge, MA: The American Academy of Arts and Sciences.
- Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist*, *67*, 431–445. doi:10.1037/a0028212
- Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*, 42–52. doi: 10.1016/j.jarmac.2013.01.001
- Klein Selle, N., Verschuere, B., Kindt, M., Meijer, E. H., & Ben-Shakhar, G. (2015). Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures. *Psychophysiology*. Advance online publication. doi: 10.1111/psyp.12583
- Kleinberg, B., & Verschuere, B. (2015). Memory detection 2.0: The first web-based memory detection test. *PLoS ONE*, *10*, e0118715. doi: 10.1371/journal.pone.0118715
- Kozel, F. A., Johnson, K. A., Grenesko, E. L., Laken, S. J., Kose, S., Lu, X., & George, M. S. (2009). Functional MRI detection of deception after committing a mock sabotage crime. *Journal of Forensic Sciences*, *54*, 220–231. doi: 10.1111/j.1556-4029.2008.00927.x
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, *58*, 605–613. doi: 10.1016/j.biopsycho.2005.07.040
- Kozel, F. A., Ravell, L. J., Lorenbaum, J. P., Shastri, A., Elihai, J. D., Horner, M. D., . . . George, M. S. (2004). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *Journal of Neuropsychiatry and Clinical Neuroscience*, *16*, 295–305. doi: 10.1176/jnp.16.3.295
- Krapohl, D. J. (2011). Limitations of the concealed information test in criminal cases. In B. Verschuere, G. Ben Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). Cambridge, UK: Cambridge University Press.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x
- Langleben, D. D., Loughhead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I. & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*, 262–272. doi: 10.1002/hbm.20191
- Langleben, D. D., & Moriarty, J. C. (2013). Using brain imaging for lie detection: Where science, law, and policy collide. *Psychology, Public Policy, and Law*, *19*, 222–234. doi:10.1037/a0028841

- Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., ... Childress, A. R. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *NeuroImage*, *15*, 727–732. doi: 10.1006/nimg.2001.1003
- Lieblisch, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR detection of information as a function of stimulus set size. *Psychophysiology*, *6*, 601–608. doi: 10.1111/j.1469-8986.1970.tb02249.x
- Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, *2*, 164–177. doi: 10.1037/1076-898X.2.2.164
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*, 385–388. doi: 10.1037/h0046060
- Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*, 258–262. doi: 10.1037/h0044413
- Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist*, *29*, 725–739. doi: 10.1037/h0037441
- Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. Berlin, Germany: Plenum Press.
- Lynn, R. (1966). *Attention, arousal and the orienting reaction*. New York, NY: Pergamon.
- MacNeill, A. L., Bradley, M. T., Cullen, M. C., & Arsenault, A. M. (2014). Cognitive and emotional reactions to questions in the comparison question test. *Perceptual and Motor Skills*, *118*, 429–445. doi: 10.2466/22.03.PMS.118k20w9
- Meijer, E. H., Klein Selle, N., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*, 879–904. doi: 10.1111/psyp.12239
- Meijer, E. H., & van Koppen, P. J. (2008). Lie detectors and the law: The use of the polygraph in Europe. In D. Canter & R. Zuckauskiene (Eds.), *Psychology and law: Bridging the gap* (pp. 31–50). Aldershot, UK: Ashgate Publishing.
- Meijer, E. H., & Verschuere, B. (2015). The polygraph: Current practice and new approaches. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches*. Chichester, UK: John Wiley & Sons, Ltd.
- Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based Concealed Information Test. *Psychophysiology*, *47*, 57–65. doi: 10.1111/j.1469-8986.2009.00883.x
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, *14*, 5–8. doi: 10.1111/j.1745–3992.1995.tb00881.x
- Miller, G. (2010). fMRI lie detection fails a legal test. *Science*, *328*, 1336–1337. doi: 10.1126/science.328.5984.1336-a
- Munsterberg, H. (1908). *On the witness stand*. New York, NY: The McClure Company.
- Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology*, *48*, 733–744. doi: 10.1111/j.1469-8986.2010.01148.x
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 49–86). Waltham, MA: Academic Press.
- National Research Council (2003). *The polygraph and lie detection*. Committee to review the scientific evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*, 162–175. doi: 10.1111/j.1469-8986.2010.01057.x
- Noordraven, E., & Verschuere, B. (2013). Predicting the sensitivity of the reaction time-based Concealed Information Test. *Applied Cognitive Psychology*, *27*, 328–333. doi: 10.1002/acp.2910
- Nose, I., Murai, J., & Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *NeuroImage*, *44*, 1380–1386. doi: 10.1016/j.neuroimage.2008.11.002
- Offe, H., & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior*, *31*, 291–303. doi: 10.1007/s10979-006-9059-3
- Ogawa, T., Matsuda, I., Tsunooka, M., & Verschuere, B. (2015). The concealed information test in the laboratory versus Japanese field practice: Bridging the scientist-practitioner gap. *Archives of Forensic Psychology*, *1*, 16–27. doi: 10.16927/afp.2015.1.2
- Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp.253–275). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511975196.015
- Patrick, C. J., & Iacono, W. G. (1991). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology*, *28*, 632–638. doi: 10.1111/j.1469-8986.1991.tb01006.x
- Peth, J., Sommer, T., Hebart, M. N., Vossel, G., Büchel, C., & Gamer, M. (2015). Memory detection using fMRI—Does the encoding context matter? *NeuroImage*, *113*, 164–174. doi: 10.1016/j.neuroimage.2015.03.051
- Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*, 381–390. doi: 10.1111/j.1469-8986.2011.01313.x
- Podlesney, J. A. (1993). Is the guilty knowledge technique applicable in criminal investigations? A review of FBI records. *Crime Laboratory Digest*, *20*, 57–61.
- Podlesney, J. A. (2003). A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations. *Forensic Science Communications*, *5*. Retrieved from <https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2003/index.htm/podlesny.htm>
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63. doi: 10.1016/j.tics.2005.12.004
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 1–47). Waltham, MA: Academic Press.
- Raskin, D. C., Honts, C. R., & Kircher, J. C. (Eds.), (2014). *Credibility assessment: Scientific research and applications*. Philadelphia, PA: Elsevier.
- Reid, J. E. (1947). A revised questioning technique in lie-detection tests. *Journal of Criminal Law and Criminology*, *37*, 542–547. doi: 10.2307/1138979
- Reid, J. E., & Inbau, F. E. (1977). *Truth and deception: The polygraph ("Lie detector") technique* (2nd ed.). Baltimore, MD: Williams & Wilkins.
- Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1432–1440. doi: 10.1037/a0036844
- Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, *1–56*. doi: 10.2307/3481305
- Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). Cambridge, UK: Cambridge University Press.
- Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. H. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *28*, 319–335. doi: 10.1111/j.1469-8986.1991.tb02202.x
- Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdacz, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *24*, 157–161. doi: 10.3109/00207458808985770
- Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*, 118–134. doi: 10.1016/j.ijpsycho.2013.08.012
- Rosenfeld, J. P., & Labkovsky, E. (2010). New P300-based protocol to detect concealed information: Resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. *Psychophysiology*, *47*, 1002–1010. doi: 10.1111/j.1469-8986.2010.01024.x

- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenberg, 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. doi: 10.1111/j.1469-8986.2008.00708.x
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Rayan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219. doi: 10.1111/j.1469-8986.2004.00158.x
- Rosenfeld, J. P., Ward, A., Drapekin, J., Labkovsky, E., & Tullman, S. (2015). *P300 evidence does not support suppression of semantic memory*. Manuscript in preparation.
- Sai, L., Zhou, X., Ding, X. P., Fu, G., & Sang, B. (2014). Detecting concealed information using functional near-infrared spectroscopy. *Brain Topography*, *27*, 652–662. doi: 10.1007/s10548-014-0352-z
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science*, *19*, 772–780. doi: 10.1111/j.1467-9280.2008.02156.x
- Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. New York, NY: Basic Books.
- Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. *Psychology, Public Policy and Law*, *5*, 203–223. doi: 10.1037/1076-8971.5.1.203
- Seymour, T. L., Seifert, C. M., Mosmann, A. M., & Shafiq, M. G. (2000). Using response time measures to assess guilty knowledge. *Journal of Applied Psychology*, *85*, 30–37. doi: 10.1037/0021-9010.85.1.30
- Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., & Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia*, *48*, 3619–3626. doi: 10.1016/j.neuropsychologia.2010.08.013
- Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2007). Detecting deception: The scope and limits. *Trends in Cognitive Sciences*, *12*, 48–53. doi: 10.1016/j.tics.2007.11.008
- Sip, K. E., Skewes, J. C., Marchant, J. L., McGregor, W. B., Roepstorff, A., & Frith, C. D. (2012). What if I get busted? Deception, choice, and decision-making in social interaction. *Frontiers in Neuroscience*, *6*, 58. doi: 10.3389/fnins.2012.00058
- Sokolov, E. N. (1963). *Perception and the Conditioned Reflex*. Oxford, UK: Pergamon Press.
- Spence, S. A., Farrow, T. F., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport*, *12*, 2849–2853. doi: 10.1097/00001756-200109170-00019
- Spence, S. A., Kaylor-Hughes, C., Farrow, T. F., & Wilkinson, I. D. (2008). Speaking of secrets and lies: The contribution of ventrolateral prefrontal cortex to vocal deception. *NeuroImage*, *40*, 1411–1418. doi: 10.1016/j.neuroimage.2008.01.035
- Suchotzki, K., Crombez, G., Smulders, F., Meijer, E., & Verschuere, B. (2015). The cognitive mechanisms underlying deception: An event-related potential study. *International Journal of Psychophysiology*, *95*, 395–405. doi: 10.1016/j.ijpsycho.2015.01.010
- Suchotzki, K., Verschuere, B., Peth, J., Crombez, G., & Gamer, M. (2015). Manipulating item proportion and deception reveals crucial dissociation between behavioral, autonomic, and neural indices of concealed information. *Human Brain Mapping*, *36*, 427–439. doi: 10.1002/hbm.22637
- 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*, 391–400. doi: 10.1037/0021-9010.67.4.391
- Tu, S., Li, H., Jou, J., Zhang, Q., Wang, T., Yu, C., & Qiu, J. (2009). An event-related potential study of deception to self preferences. *Brain Research*, *1247*, 142–148. doi: 10.1016/j.brainres.2008.09.090
- Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in Psychology*, *3*, 526. doi: 10.3389/fpsyg.2012.00526
- Verschuere, B., & Ben-Shakhar, G. (2011). Theory of the concealed information test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer. *Memory detection: Theory and application of the Concealed Information Test* (pp. 128–148). Cambridge, UK: Cambridge University Press.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge, UK: Cambridge University Press.
- Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. H. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology*, *41*, 461–466. doi: 10.1111/j.1469-8986.00167.x
- Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. (2005). Psychopathic traits and autonomic responding to concealed information in a prison sample. *Psychophysiology*, *42*, 239–245. doi: 10.1111/j.1469-8986.2005.00279.x
- Verschuere, B., Crombez, G., Degrootte, T., & Rosseel, Y. (2010). Detecting concealed information with reaction times: Validity and comparison with the polygraph. *Applied Cognitive Psychology*, *24*, 991–1002. doi: 10.1002/acp.1601
- Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: Response latency-based measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer. *Memory detection: Theory and application of the Concealed Information Test* (pp. 128–148). Cambridge, UK: Cambridge University Press.
- Verschuere, B., & Kleinberg, B. (2015). ID-check: Online concealed information test reveals true identity. *Journal of Forensic Science*. doi: 10.1111/1556-4029.12960
- Verschuere, B., Kleinberg, B., & Theocharidou, K. (2015). Rt-based memory detection: Item saliency effects in the single-probe and the multiple-probe protocol. *Journal of Applied Research in Memory and Cognition*, *4*, 59–65. doi: 10.1016/j.jarmac.2015.01.001
- Verschuere, B., Schuhmann, T., & Sack, A. T. (2012). Does the inferior frontal sulcus play a functional role in deception? A neuronavigated theta-burst transcranial magnetic stimulation study. *Frontiers in Human Neuroscience*, *6*. doi: 10.3389/fnhum.2012.00284
- Verschuere, B., Spruyt, A., Meijer, E. H., & Otgaar, H. (2011). The ease of lying. *Consciousness and Cognition*, *20*, 908–911. doi: 10.1016/j.concog.2010.10.023
- Vincent, A., & Furedy, J. J. (1992). Electrodermal differentiation of deception: Potentially confounding and influencing factors. *International Journal of Psychophysiology*, *13*, 129–136. doi: 10.1016/0167-8760(92)90052-D
- Visu-Petra, G., Miclea, M., & Visu-Petra, L. (2012). Reaction time-based detection of concealed information in relation to individual differences in executive functioning. *Applied Cognitive Psychology*, *26*, 342–351. doi: 10.1002/acp.1827
- Visu-Petra, G., Varga, M., Miclea, M., & Visu-Petra, L. (2013). When interference helps: Increasing executive load to facilitate deception detection in the concealed information test. *Frontiers in Cognitive Psychology*, *4*, 146. doi: 10.3389/fpsyg.2013.00146
- Winograd, M. R., & Rosenfeld, J. P. (2014). The impact of prior knowledge from participant instructions in a mock crime P300 concealed information test. *International Journal of Psychophysiology*, *94*, 473–481. doi: 10.1016/j.ijpsycho.2014.08.002

(RECEIVED September 8, 2015; ACCEPTED December 18, 2015)