CHAPTER 17

Admissibility and Constitutional Issues of the Concealed Information Test in American Courts: An Update*

John B. Meixner, Jr.
Northwestern University, Assistant United States Attorney, Detroit, MI, United States

INTRODUCTION

The use of physiological tools to detect intentionally concealed knowledge about crime-related information has been a controversial and well-researched topic among scholars for well over 100 years. While essentially all the research in the first half of the 20th century focused on detecting *lies* about concealed knowledge, a substantial body of research related to the detection of *recognition* of concealed knowledge has also developed, beginning with David Lykken's seminal work in the 1950s (Lykken, 1959). While that body of work began exclusively by measuring autonomic nervous system (ANS) data using the polygraph, it has now been expanded to a variety of other measures, including reaction times and, perhaps most notably, physiology directly linked to brain activity in the form of event-related potentials (ERPs) or the hemodynamic response as measured via functional magnetic resonance imaging (fMRI).

Tests related to detecting intentionally concealed knowledge have a long history with the American legal system as well, though mostly related to lie detection rather than recognition detection. The seminal case outlining the dominant standard for the admissibility of scientific expert testimony for 70 years, *United States v. Frye* (1923), involved a challenge to the admissibility of the systolic blood pressure deception test, an early precursor to the ANS-based comparison question test (CQT). That case also started a

* The views expressed in this chapter are those of the author and do not necessarily represent the positions of the Department of Justice or the United States.

Detecting Concealed Information and Deception
ISBN 978-0-12-812729-2
https://doi.org/10.1016/B978-0-12-812729-2.00017-3

trend that has endured to the present: skepticism in the legal community as to the use of credibility-assessment tools in court. Indeed, the *Frye* court held that the results of the systolic blood pressure deception test could not be introduced because the test had not yet gained sufficient "standing and scientific recognition among physiological and psychological authorities" (United States v. Frye, 1923, p. 1014). Since then, nearly every major decision to consider a tool potentially falling under the label of deception detection—a label that is dubious in the case of the Concealed Information Test (CIT), as I will discuss next—has rejected that evidence.

In this chapter, I explain the governing legal standards that will determine the potential use of the CIT in court. Those standards come from two distinct sources: (1) the Federal Rules of Evidence (and the cases interpreting those rules), which limit the admissibility of evidence in certain circumstances, and (2) the US Constitution (and the cases interpreting it), which limits the use of evidence obtained in violation of certain individual rights—most notably, for our purposes, the privilege against self-incrimination. I focus primarily on the first category of standards, as the literature is better developed and more accurate predictions can be made as to the current challenges and ways those challenges could be overcome. I briefly discuss the second category, though the way in which courts will conduct that analysis remains largely unknown.

Though these issues may seem esoteric to nonlawyer scholars involved in CIT research, I argue that they should drive the research agenda of every CIT researcher interested in the practical use of their work, at least in the United States. There are serious hurdles that the CIT faces limiting its potential use in American courts, but targeted empirical research can address at least some of the problems.

POTENTIAL ADMISSIBILITY OF THE CONCEALED INFORMATION TEST

Credibility Assessment Tests and the American Legal System

As a general matter, testimony in the American legal system is separated into two categories: lay witness testimony and expert witness testimony. Lay witnesses are those individuals who have firsthand, personal knowledge about matters relevant to the case. For example, a person who witnessed events related to the commission of a crime would be a lay witness who could testify about what her or she saw or heard. As a general matter, lay

witnesses are limited in the extent to which they can testify in the form of an opinion; they must testify only to what they actually saw or heard, and may provide opinions only based on those firsthand perceptions, and not based on any specialized knowledge or expertise they may have (Federal Rule of Evidence 701, 2000).

In contrast, expert witnesses who are sufficiently qualified by specialized knowledge, training, or other expertise are permitted to testify in the form of an opinion, provided that certain circumstances are met. Did the car's poorly maintained breaks give out at the time of the accident, contributing to the collision? Did the DNA sample found at the scene of the crime match the defendant's DNA? Expert witnesses are permitted to opine on such questions based on their specialized knowledge. As you can imagine, such testimony is often at the heart of critical issues in the case, and can be very powerful in influencing the judge or jury tasked with weighing the evidence. Accordingly, the American legal system has long had checks in place to help ensure that only sufficiently valid and reliable opinions are presented.

For most of the 20th century, the dominant test for assessing whether proposed expert testimony is sufficiently valid and reliable was derived from *Frye v. United States* (1923). In that case, the defendant—on trial for second-degree murder—had sought to offer an expert to testify about the results of a systolic blood pressure deception test that he had administered to the defendant and that had presumably indicated that the defendant was being truthful in denying his involvement in the crime. That particular test appears to have been a precursor to the modern CQT. As the court described the systolic blood pressure test,

> *it is asserted that blood pressure is influenced by change in the emotions of the witness, and that the systolic blood pressure rises are brought about by nervous impulses sent to the sympathetic branch of the autonomic nervous system. Scientific experiments, it is claimed, have demonstrated that fear, rage, and pain always produce a rise of systolic blood pressure, and that conscious deception or falsehood, concealment of facts, or guilt of crime, accompanied by fear of detection when the person is under examination, raises the systolic blood pressure in a curve, which corresponds exactly to the struggle going on in the subject's mind, between fear and attempted control of that fear, as the examination touches the vital points in respect of which he is attempting to deceive the examiner (p. 1013).*

The trial court rejected the expert, and the defendant was convicted. The court of appeals agreed with the trial court's decision to reject the

evidence. It explained that it is the court's role to determine whether the scientific principle at issue had "crosse[d] the line between the experimental and demonstrable stages." In order to make this determination, the *Frye* court instructed that courts must decide whether the scientific method or principle at issue is "sufficiently established to have gained general acceptance in the particular field in which it belongs." Applying that standard to the systolic blood pressure deception test, the court found that the test "ha [d] not yet gained such standing and scientific recognition among physiological and psychological authorities," and accordingly determined that the trial court had made the correct decision in rejecting it.

Following *Frye*, nearly all courts adopted some form of what has come to be known as the general acceptance test, determining whether a proposed expert's method is "sufficiently established to have gained general acceptance in the particular field in which it belongs." That standard persisted until at least 1975, when the Federal Rules of Evidence were adopted. The Federal Rules of Evidence sought to codify evidentiary rules that had long been established through cases, and provide a uniform and easy-to-understand set of rules for use in federal courts. Federal Rule of Evidence 702 (which currently governs the admissibility of expert testimony in all federal and most state courts) provided at the time a very broad standard, stating that, where helpful to the body weighing the evidence, a witness qualified as an expert by sufficient skill or expertise is permitted to testify in the form of an opinion (Federal Rule of Evidence 702, 1975). The rule made no mention of the *Frye* general acceptance test, and for some time there was a question as to whether Federal Rule 702 had replaced the *Frye* test, and, if so, how the new test was to be applied.

That question was answered in 1993, when the US Supreme Court decided *Daubert v. Merrell Dow Pharmaceuticals, Inc.* That case involved a claim that the ingestion of a particular antinausea drug marketed by the defendant had caused birth defects in the children of mothers who had taken the drug during their pregnancies. Both sides offered experts supporting their claims as to whether the drug could have caused such defects, but the trial court rejected the plaintiffs' experts on the basis that their studies were based on methods that had not been generally accepted in the relevant scientific field.

On appeal, the Supreme Court explained that Federal Rule 702 made no mention of any general acceptance test, nor were the rules designed to embrace any such standard, and accordingly, the rules displaced the *Frye* test entirely. The Court then went on to explain that Rule 702's discussion of

scientific evidence meant that it included a requirement that the method at issue meet a certain standard of evidentiary reliability, a somewhat murky term that the court noted is akin to trustworthiness, or what scientists would typically call validity (Daubert v. Merrell Dow Pharmaceuticals, 1993). How is a court to determine whether such evidentiary reliability is present in a method? The court largely left this question to the discretion of trial judges, though it outlined four now-famous factors for courts to apply:

- "[W]hether [the theory or technique] can be (and has been) tested,"
- "[W]hether the theory or technique has been subjected to peer review and publication,"
- "[T]he known or potential rate of error" and "the existence and maintenance of standards controlling the technique's operation,"
- The "general acceptance" of the technique, as outlined in *Frye*.[1]

Though the Court was careful to note that these factors are nonexclusive and that other factors may bear on the inquiry, the vast majority of admissibility inquiries under the *Daubert* standard[2] focus on these four[3] factors. With this basic history in mind, I now turn to a discussion of the potential admissibility of the concealed information test under the now-dominant *Daubert* standard.

Prior Accounts of the Concealed Information Test's Admissibility

Daubert was a monumental decision. Expert testimony is common in American trials, and *Daubert* completely upended the method by which judges are to evaluate such testimony's admissibility. Unsurprisingly, *Daubert* immediately engendered substantial scholarly discussion. Scholars have disagreed both as to the merits of the standard (e.g., Bernstein, 1994; Capra, 1998; Faigman, Kaye, Saks, & Sanders, 2000; Fenner, 1996; Heinzerling, 2006; Jonakait, 1994) and whether it has actually had any real impact on admissibility outcomes (Black, Ayala, & Saffran-Brinks, 1994; Chen & Yoon, 2005).

[1] Federal Rule of Evidence 702 was subsequently amended in 2000 in response to *Daubert*.

[2] Although the *Daubert* standard has been adopted by the majority of states, states are not obliged to follow the Federal Rules of Evidence, and some states have continued to apply the *Frye* standard, or have applied some other standard (see Daubert v. Frye).

[3] There is some dispute as to whether the test contains five separate factors (with maintenance of standards as a separate factor) or whether the error rate and maintenance of standards factors combine to form one single factor, yielding a total of four. For discussion, see Meixner and Diamond (2014, p. 1068).

Pertinent to this volume, a number of authors have previously discussed the potential admissibility under *Daubert* of both the CIT and the CQT and its analogues, particularly those tests as conducted using neuroscience-based methods such as ERPs and fMRI. Those authors have largely been critical of the potential admissibility of neuroscience-based CQT-like tests that purport to determine whether a participant is telling the truth or lying, largely for the same reasons that many scholars have been critical of the ANS-based variant of the CQT: the lack of a coherent theoretical underpinning for the tests, the potential for a high number of false positives, and the lack of clearly controlled laboratory studies demonstrating a promising rate of error (Alexander, 2007; Ellenberg, 2009; Kittay, 2007; Meixner, 2012; Moreno, 2009; Moriarty, 2009; Shapiro, 2016; Spence, 2008; Woodruff, 2014). Others have called for regulation limiting the use of fMRI-based lie detection, independent of its admissibility under *Daubert* and the Federal Rules of Evidence (Greely & Illes, 2007).[4] And at least three courts have rejected fMRI-based lie detection evidence, one rejecting it on the basis that it fails to satisfy the *Daubert* standard (United States v. Semrau, 2010, 2012), one on the basis that it fails to meet the *Frye* standard (Smith v. Maryland, 2011), and a third because the lie-detection evidence would impede on the role of the jury as a credibility assessor (Wilson v. Corestaff Services, LLP, 2010).[5]

Perhaps because of those cases and because of the rise of fMRI as a popular imaging tool, discussion of the admissibility of neuroscience-based credibility assessment tools has focused largely on lie detection paradigms, like the ones at issue in the *Semrau*, *Smith*, and *Wilson* cases. Further, numerous commentaries have failed to make any distinction whatsoever between lie detection paradigms and recognition detection paradigms like the CIT, whether using neuroscience-based tools or autonomic-based tools. This is important and dismaying, because, as many of the chapters in this volume outline in detail, there are critically important differences

[4] Some authors, however, have expressed a more positive view of the potential use of neuroscience-based tests using CQT-like methods (Langleben, 2008; Schauer, 2010; Langleben & Moriarty, 2013).

[5] Aside from methodological problems, there is another potentially fatal hurdle facing true lie detection paradigms: the notion that "the *jury* is the lie detector," one of the principles the Supreme Court has referenced in rejecting attempts of parties to admit the results of polygraph exams (United States v. Scheffer, 1998). Empirical data cast at least some doubt on the utility of that principle, but it nonetheless is a well-established legal doctrine that may limit the use of lie detection evidence even if methodological problems are solved. For a more complete treatment of this issue, see Meixner (2012).

between the two, both as a matter of theoretical validity and as a matter of practical accuracy in lab testing. As I and others have written in the past, those differences should radically affect the *Daubert* analysis conducted by a judge willing to examine the science at a deep level (Erickson, 2007; Meixner, 2012; Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013), though the fact that the scholarly literature has been inconsistent in doing this will make judges' jobs more difficult.

In the past, I have argued that the P300-based CIT—a variant of the CIT using the P300 ERP component as the primary measure of recognition—is reasonably close to admissibility, and I think the arguments I outline here also apply to the autonomic-based CIT conducted using the polygraph.[6] The CIT has been subjected to rigorous peer review for more than half a century, and, especially in the past 3 decades with the introduction of the P300-based CIT, the number of peer-reviewed publications of the CIT has skyrocketed, likely satisfying *Daubert*'s aim that methodological flaws be identified by other scientists so they can be remedied. Likewise, these tests are well-regarded in their scientific fields, and survey data have already been collected demonstrating the general acceptance of the CIT among psychologically oriented scientists (Iacono & Lykken, 1997). Further, most of the laboratories regularly conducting CIT research use specific standards and methods in their testing, leading to the consistency and reproducibility that the *Daubert* standard seeks.

The most difficult factors of the *Daubert* analysis for the CIT to overcome, I think, are the "known or potential rate of error" factor and the testability factor. This may seem counterintuitive at first, since the CIT has undergone extensive controlled laboratory testing across a variety of conditions, and has generally been reported to have a reasonably low false-negative error rate and a very low false-positive error rate. The issue, as I and others have argued, is that the vast majority of these studies have been conducted in conditions that do not sufficiently approximate those of the real world. For example, it is common in the P300-based CIT literature to use what is termed "self-referring information"—items like birth date, address, or telephone number—as an analogue for the critical crime-related information that an individual associated with the crime would recognize in a real-world CIT (e.g., Rosenfeld, Soskins, Bosh, & Ryan, 2004; Rosenfeld et al., 2008). Most other studies use some form of a mock-crime paradigm,

[6] Indeed, the ANS-based CIT has undergone at least some level of field testing, and that would be to its benefit in a *Daubert* analysis.

in which participants are given instructions to steal a particular item or carry out a particular task under controlled conditions (e.g., Ben-Shakhar & Dolev, 1996; Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Lui & Rosenfeld, 2008; Lykken, 1959; Meixner & Rosenfeld, 2011; Mertens & Allen, 2008; Rosenfeld et al., 1988; Winograd & Rosenfeld, 2011).

Both of these designs are problematic when we seek to extrapolate the error rates yielded by the laboratory study to the error rates we might expect when using the test in the field. Self-referring information is rehearsed repeatedly and is particularly meaningful to individuals, likely driving up sensitivity of the test. And mock crimes—while approximating real crimes—suffer from several external validity problems. First, they involve a singular focus on the assigned crime and do not provide the rich array of distracting details that exist in the real world, which may decrease detection sensitivity because of reduced salience at the time of encoding. Second, the items involved are often notable, stand out, or are easily remembered by participants (e.g., the critical item in a test might be the only potentially valuable item encountered during the task, such as a ring or a computer disk). Third, participants in a mock crime CIT study know that they are participating in an experiment, and may be more likely to strongly encode and clearly remember relevant items due to the salience of the experiment itself. And fourth, actions committed by the subject in the lab are generally not voluntary, while actions involved in real crimes typically are.

There have been a few true field tests of the CIT, all using the ANS-based CIT variant. A pair of studies conducted in Israel found excellent accuracy rates in classifying nonknowledgeable/innocent individuals (between 95% and 98% specificity), but significantly lower accuracy rates among knowledgeable/guilty individuals (around 75% total sensitivity, and as low as 50% when using only a single ANS measure) (Elaad, 1990; Elaad, Ginton, & Jungman, 1992). Other articles have attempted to analyze data from field use of the CIT in Japan, where police have regularly used the test in criminal investigations for more than half a century. Kobayashi, Yoshimoto, and Fujihara (2009) reported results of 25%–50% sensitivity for individual autonomic measures in response to single questions, but did not report any specificity results, nor combined sensitivity results that would be indicative of detection rates. At least one other paper has reported some Japanese data with similar results and approximately 95% specificity, but the methods and data are difficult to interpret (Hira & Furumitsu, 2002; for reviews of Japanese methods, see Matsuda, Nittono, & Allen, 2012; Osugi, 2011). And one additional concern is that it may be difficult in many crimes

to cull appropriate details on which to test (Krapohl, 2011); a critique that was borne out in one examination of FBI records (Podlesny, 1993).

These issues would likely influence a judge's *Daubert* analysis of the CIT, whether ANS-based or P300-based, though the ANS-based CIT would have a stronger argument for having known error rates and having been tested. In assessing an fMRI-based CQT variant, the *Semrau* trial court noted that "there are no known error rates for fMRI-based lie detection outside the laboratory setting, i.e., in the 'real-world' or 'real-life' setting" (United States v. Semrau, 2010, p. 11). If a judge cannot trust laboratory studies to provide a realistic approximation of the error rate of the test in the real world, there is no relevant error rate to assess, and the factor would clearly cut against admissibility. These issues may come into play under the testability *Daubert* factor as well: while there has been significant testing of the CIT's general methodology in the lab, an argument can be made that virtually no testing of the CIT has been done in sufficiently realistic settings. Indeed, without a reliable way to determine ground truth in a field test, it may be difficult to truly test the CIT at all.[7]

Because these conclusions are dependent on the literature, they are subject to change, and the most recent substantial commentaries on this issue are now nearly 5 years old, and largely discuss only the P300 literature (Meixner, 2012; Rosenfeld et al., 2013). In the next section, I examine recent CIT studies and assess the extent to which they have contributed to solving the puzzle of the true error rate of the CIT.

How Do More Recent Studies Affect the Analysis?

As an initial matter, I note that a large number of CIT papers published each year are not focused on developing a CIT paradigm that even attempts to mimic real-world conditions. Perhaps the majority of CIT studies in a given year are instead focused on refining a particular technique or aspect of the CIT (such as the ideal timing of items presented or presentation format), tweaking an analysis method (such as the ideal way to score responses, or the statistical measures used to make decision criteria, like bootstrapping), testing the effectiveness of countermeasures against a particular test, or

[7] We might ask whether judges, as nonscientists, will be capable of making these sorts of nuanced assessments of the literature. While the extent to which judges actually assess scientific methods is the subject of a long-running debate since *Daubert* was authored, at least one recent empirical study has demonstrated that judges spend significant space in their opinions assessing the validity of experts' scientific methods and their likely effects on error rates (Meixner & Diamond, 2014).

designing an entirely new CIT protocol (e.g., Deng, Rosenfeld, Ward, & Labkovsky, 2016; Labkovsky & Rosenfeld, 2014; Rosenfeld, Ward, Frigo, Drapekin, & Labkovsky, 2015).

One particularly notable subcategory in this group is studies that have sought to refine the so-called searching CIT (sometimes abbreviated as SCIT). In that test, rather than attempting to determine whether a particular suspect possesses a specific item of concealed information, the investigator knows that the suspect possesses at least *some* concealed information, and is trying to determine specifically what the information is, such as the location of a bomb or the name of a conspirator (e.g., Breska, Ben-Shakhar, & Gronau, 2012; Elaad, 2016; Meijer, Smulders, & Merckelbach, 2010; Meixner & Rosenfeld, 2011). Relatedly, more recent studies have conducted the SCIT on groups, much like individuals in a terrorist cell might be questioned jointly to maximize information (e.g., Elaad, 2016; Meijer, Ben-Shakhar, Verschuere, & Donchin, 2013; Meijer, Bente, Ben-Shakhar, & Schumacher, 2013). While these studies are qualitatively different from other CITs, they are important for our discussion here because they demonstrate the use of the CIT even if admissibility is never achieved—the SCIT is potentially useful to law enforcement in seeking to stop crime before it happens, or in finding new suspects of crimes that have already happened.

However, more pertinent to this chapter are a number of recent studies that have focused specifically on the issue of external validity in the CIT. While none of these studies involve actual field testing of actual crime suspects,[8] they do use unique methods to attempt to make laboratory studies more realistic. I will briefly describe several recent efforts.

First, while many early CIT experiments involved repeated rehearsal of the critical items in order to ensure that they would be recognized during the CIT, which is a method with low ecological validity (e.g., Farwell & Donchin, 1991), more recent studies have sought to measure the effects of such rehearsal (Bradley, Malik, & Cullen, 2011), finding that it does lead to stronger sensitivity. More recent CITs have sought to eliminate such priming of the information to be tested, and in some circumstances, have tested for both central details, which relate directly to the assignment given

[8] It is worth noting here that true field testing on criminal suspects will be extremely difficult to conduct, for a number of reasons, including: law enforcement may be reluctant to work with experimenters on real cases; IRB approval for research on criminal suspects will be very difficult to obtain; and any such research would require the cooperation of criminal suspects, who may not be inclined to participate voluntarily.

to the participant (e.g., the item to be stolen), and peripheral details, which may be noticed but are incidental to the crime (e.g., the color of the wall in the room where the crime was committed). As a general matter, these studies have found that knowledge of peripheral details is not detected with as much sensitivity as central details (e.g., Gamer, Kosiol, & Vossel, 2010; Nahari & Ben-Shakhar, 2011; Peth, Vossel, & Gamer, 2012; for review, see Meijer, Ben-Shakhar, et al., 2013; Meijer, Bente, et al., 2013). How might this information influence a *Daubert* analysis? If peripheral details cannot be detected with the same sensitivity as central details (a question that certainly still warrants further research), the critical question becomes: how often will peripheral details need to be relied upon in order to make detection decisions? The answer to this question remains unknown, though at least one study has come to the conclusion that strongly encoded details will be difficult to find in many case records (Podlesny, 1993). A judge deciding the admissibility of a CIT might then conclude that laboratory studies involving detection of only central details inflate the sensitivity of the CIT as compared to what could reasonably be expected in the field.

This concern might be at least in part alleviated by a recent study that did not employ a mock-crime paradigm, but rather attempted to detect recognition of events experienced during normal daily life (Meixner & Rosenfeld, 2014). In that study, subjects wore a video-recording device for a 4-h period, and then returned to the lab on the following day, when they were presented with CIT blocks containing information related to events that were recorded by the camera that participants wore. At the individual subject level, the investigators were able to correctly classify all 24 participants as either knowledgeable or nonknowledgeable. From a legal perspective, the study might be useful in demonstrating that the CIT is capable of detecting purely incidentally acquired information, and it is also notable in that it is one of the few CIT studies that detects information acquired purely voluntarily by participants, rather than requiring them to memorize information or commit a mock crime. However, given the novelty of this method, replication and extension are still necessary before the results can be given significant weight.

Other recent studies have sought to examine and address different ecological validity limitations common in many CIT experiments. For example, in Peth et al. (2015), investigators had three groups of participants either commit a mock crime, plan (but not commit) the same mock crime, or fulfill a noncriminal task that exposed them to information related to the mock crime. Following a CIT in which both fMRI and skin conductance

response data were collected, the authors found virtually no differences between the three groups. While the result is potentially discouraging in that they "indicate[] a high risk for innocents with crime related knowledge to be misclassified as guilty," (p. 170), it is encouraging in that it indicates that CIT studies with poor ecological validity may not necessarily have poor external validity—that is, even though CIT studies do not approximate field-like conditions, the artificial conditions they employ may not actually influence the results.

One other recent result in this vein was published by Zaitsu (2016). That study involved an artificial card test paradigm in which the test sought to determine which of five numbered cards the participant had drawn. What makes the study interesting is the participants: the study compared performance of voluntary participants and actual criminal suspects, finding virtually no differences in performance between the groups. The result, as in Peth et al. (2015), is encouraging in that it indicates that laboratory participant populations may not be so different from field populations.

Another area of poor ecological validity for most CIT studies is motivation: an individual taking a CIT with his potential livelihood on the line would have a strong incentive to avoid detection, while typical student participants in CIT studies do not have the same motivation. This shortcoming could cut in either direction: participants motivated to avoid detection might be more effective in avoiding detection (perhaps though effective performance of countermeasures) but, on the other hand, their motivation could also have a detrimental effect of making the critical information even more salient, and thus make them stand out even more from irrelevant items.

While some studies assessing polygraph CIT data have found mixed results as to the effect of motivation (e.g., Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Zvi, Nachson, & Elaad, 2012), many of those studies are now nearly 20 years old, and nearly all use exclusively polygraph data. One recent study examined the effect of motivation using a reaction time-based CIT, finding that an incentive of up to $5 to avoid detection had no effect on detection rates (Kleinberg & Verschuere, 2016). While it may be questionable whether such a small reward sufficiently motivated participants, manipulation checks in the study indicated that those in the motivation group did seek to avoid detection more than the control group. A new set of studies from the Rosenfeld group, discussed in Chapter 6 of this volume, has similarly found no significant effects of modest ($10) financial motivation to avoid detection in several P300 CIT

applications. More work, especially with potentially stronger motivators, would be helpful in this area, along with extension of this line of research to P300-based and fMRI-based CITs.

Other, more traditional, recent CIT studies have also continued to define the contours of the CIT's accuracy and limits in the lab context, and while these studies do not solve the ecological validity problems described earlier, they will still be useful if and when CITs are subjected to a *Daubert* analysis. Recent studies have continued to demonstrate very good sensitivity and specificity of the CIT using a variety of dependent measures (e.g., Lukács et al., 2016). One of the primary topics of interest continues to be countermeasures (e.g., in the memory suppression context, Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013; Hu, Bergström, Bodenhausen, & Rosenfeld, 2015; Rosenfeld, Ward, Drapekin, Labkovsky, & Tullman, 2017; Ward & Rosenfeld, 2017). While the evidence of the effectiveness of countermeasures is mixed and some protocols appear to be more resistant to them than others, they continue to be a serious concern for the overall accuracy of the CIT.

What effect are countermeasures likely to have on a *Daubert* analysis? It is difficult to know for certain, but I expect that the effect will be relatively minimal, largely because of the strong specificity of the CIT. This is somewhat paradoxical, as countermeasures threaten the sensitivity of the test. However, as the false-positive rate approaches zero (as should theoretically occur as the number of categories of questions is increased in a CIT), even if the sensitivity is not close to 100%, the test still provides strong probative value because when a knowledgeable result occurs, the person weighing the evidence can trust that it is almost certainly valid (as a false-positive error is exceedingly rare). Of course, if such strong specificity cannot be guaranteed, countermeasures become a much greater threat; where the judge or jury considering the evidence cannot strongly rely on either a knowledgeable or nonknowledgeable test result, the probative value of the test is substantially reduced and its potential prejudicial effect is greater (see Federal Rule of Evidence 403, 2011, which requires that a court exclude evidence "if its probative value is substantially outweighed by a danger of…unfair prejudice").

This chapter would be remiss not to discuss the controversial studies that have recently been published by Lawrence Farwell and his colleagues. Farwell was a coauthor on one of the very first P300-based CIT studies (Farwell & Donchin, 1991), and eventually sought to market a commercially available variant of the P300-based CIT that he terms

"Brain Fingerprinting." That test was raised in two court cases—the only such US cases to discuss the admissibility of a CIT paradigm. Those cases, *Harrington v. State* (2001) and *Slaughter v. State* (2005), do not, in my view, shed much light on how the *Daubert* analysis would apply to a CIT in future instances. This is for several reasons. First, these cases involved a unique procedural situation. In both cases, criminal defendants had been convicted at trial, and then—years later—took a Brain Fingerprinting test that yielded a null result: they indicated that the defendants lacked knowledge about particular aspects of the crimes (and, in Harrington's case, the results indicated that Harrington did respond to stimuli related to his alibi defense). Both defendants then sought to obtain a new trial, alleging that the Brain Fingerprinting tests, among other things, were newly discovered evidence that they could not have obtained previously and that indicated their innocence. Defendants seeking such relief generally must pass a high hurdle—they must show not only that the newly discovered evidence is admissible, but also that it would have changed the outcome of the trial if it were then known.

Because of this unique procedural situation, a full Daubert analysis was not completed in either case. In *Harrington*, an Iowa court opined that while the P300 component itself is generally accepted among psychophysiologists, the MERMER effect—Brain Fingerprinting's proprietary analysis method—was not generally accepted. It also discussed a number of other methodological issues related to the fact that the conclusion was based on a null result, with subjectively chosen stimuli, years after the actual crime was committed. But it never made a clear determination as to the admissibility of the evidence because it found that the defendant did not demonstrate that the evidence would have changed the result of his trial. On appeal, the Iowa Supreme Court reversed Harrington's conviction on entirely independent grounds, making no statement as to the admissibility of the Brain Fingerprinting evidence (*Harrington v. State*, 2003).

Similarly, in *Slaughter v. State* (2005), an Oklahoma state appellate court found no evidence that Brain Fingerprinting is generally accepted in the psychological community, and could not even complete a *Daubert* analysis because Farwell failed to provide any report to support his affidavit. Ultimately, the court found that the Brain Fingerprinting evidence was not newly available because the test could have been conducted at the time of the defendant's original appeal. It also stated that, in part based on the lack of a complete report regarding the test, the defendant had not demonstrated that the test would survive a *Daubert* analysis.

While Farwell went for a period of time in the early 2000s without publishing any significant work, he and his colleagues have been active more recently, publishing a broad description of Brain Fingerprinting (Farwell, 2012) as well as two sets of purported field studies (Farwell, Richardson, & Richardson, 2013; Farwell, Richardson, Richardson, & Furedy, 2014).

In Farwell et al. (2013), the authors report results from four studies, two of which used acronyms well known to FBI agents or explosive device dismantlers as the critical information to be detected (studies 3 and 4, pp. 274–276). A third study (study 1) used information known to CIA operatives about their investigations, and a fourth (study 2) used "information regarding real crimes, in circumstances where the outcome of the test could produce major, life-changing consequences" presented to "suspects in criminal investigations or convicted prisoners who claimed innocence and were appealing their convictions" (p. 272). Among all four groups, the authors report perfect accuracy, with no indeterminate results. Similarly, in Farwell et al. (2014) the authors presented military medical experts with terms "known only to experts in military medicine" among other irrelevant items, along with the same items presented to nonexperts, and again reported perfect discriminability between experts and nonexperts.

These results would be very important to the field and to the issue of admissibility, especially the one true field study in Farwell et al. (2013) (study 2), if not for a number of shortcomings. First, as is a common criticism of Farwell's studies, the methods are not described in sufficient detail that they could be independently replicated. Accordingly, there have not been any independent studies conducted by any groups other than Farwell's that use Farwell's methods. Such independent replication is especially important when the claims are extraordinary, as Farwell's are. Second, the results are so uniformly perfect (as is virtually every result reported by Farwell since Farwell & Donchin, 1991) that they are difficult to believe. As others have noted, the use of highly specific standards could lead to selection bias, influencing the result (Meijer, Ben-Shakhar, et al., 2013). Because of these and other issues, Farwell's studies have been strongly criticized by experts in the field (e.g., Guadet, 2011; Meijer, Ben-Shakhar, et al., 2013; Rosenfeld, 2005).

How would a judge facing an admissibility decision interpret the Farwell line of P300-based CIT studies? While Farwell's work has been examined previously by courts, as just discussed, it has become so broadly criticized in the field that it is hard to imagine how any court could consider

it generally accepted, either under *Frye* (in which general acceptance is the only inquiry) or *Daubert* (in which general acceptance is a single, but important, factor). Other shortcomings in the Farwell studies, as discussed in the critiques outlined earlier, would further cut against its admissibility under *Daubert*.

The effect of Farwell's studies on an admissibility determination leads into another, broader, question: to what extent would CIT studies using one dependent measure (e.g., ANS measurements) be considered in determining the admissibility of a CIT conducted using a different dependent measure (e.g., P300)? When discussing the four *Daubert* factors, the *Daubert* Court instructed trial judges to assess them through the lens of the "theory or technique" or the "particular scientific technique" at issue. How broadly to frame the question of what scientific technique was at issue or what constituted the relevant scientific community was long a central argument in Frye determinations of admissibility, and remains an issue under *Daubert*. However, I think the more natural reading of *Daubert* asks the trial judge to make determinations based on the specific method sought to be admitted. The *Semrau* court's analysis largely bears this out—both the trial-level court and the court of appeals assessed research on fMRI-based lie detection, not lie detection research generally.

This does not mean, however, that ANS-based CIT research will necessarily be irrelevant to the admissibility of a P300-based CIT, or vice versa. To the extent that the research demonstrates consistently similar results across modalities, studies focusing on one modality may inform capabilities in another modality. Some recent studies have begun to combine data from multiple modalities into single CIT paradigms (e.g., Langleben et al., 2016; for meta-analysis of various modalities, see Meijer, Selle, Elber, & Ben-Shakhar, 2014), and while those have shown some differences in the capabilities of different modalities, research in this area has promise in better allowing the various subfields of CIT research to benefit each other.

Ultimately, a review of the recent CIT literature leads me to the following recommendations if a goal of the field is to eventually attain admissibility of the CIT in American courts:

- As I and others have written in the past (Meixner, 2012; Rosenfeld et al., 2013), the top priority for CIT researchers should be field testing, especially P300-based CITs, which have not undergone any field testing. Until such testing is done, I think it will be exceedingly difficult to admit a P300-based CIT under the *Daubert* standard. An ANS-based

CIT would have a stronger case, given the field testing that has occurred in Israel and the regular use of the ANS-based CIT in Japan, though the data presented from Japan to date remain limited, and the Elaad field studies from the early 1990s report some concerning issues with sensitivity. More detailed descriptions of error rates in the field in Japan (with specific explanations of the methods used to ascertain ground truth) would be particularly useful.

- To the extent that field testing is not possible, laboratories should seek to maximize external validity in their studies. In mock-crime studies, researchers should focus on making mock crimes as complex and realistic as possible. A major concern of courts assessing the value of mock-crime studies is likely to be the fact that crime-related information in mock-crime CIT studies is readily encoded, whereas in the field that same information is frequently learned during the often-chaotic and unrehearsed commission of a crime. To the extent that we can model our studies after that environment, they will be more useful in a *Daubert* analysis.
- Researchers should seek to empirically assess the opinions of the psychological community regarding CIT methods and other tests in the same domain, such as the CQT. Iacono and Lykken (1997) conducted a survey examining this, but that was before the broad proliferation the P300-based CIT and the variety of new methods that are now being explored. The CIT sits in an odd position: it appears to be widely accepted among the scientific community as a valid method and a superior tool for detection of crime-related knowledge as compared to lie-detection methods such as a CQT, and yet it is not well known to anyone outside of that community. Because judges assessing admissibility fall outside of that community, the more concrete evidence can be gathered to demonstrate the scientific community's view of the validity of the CIT, the better.
- Researchers should seek to make extremely clear in their papers the distinctions between memory detection and lie detection. A judge applying *Daubert* is likely to have very little familiarity with the credibility assessment field, and the term "polygraph" is so intertwined with lie detection that judges are likely to confuse the two classes of tests (Meixner, 2012; Ogawa, Matsuda, & Tsuneoka, 2015). This may lead to judges projecting many of the CQT's validity problems onto the CIT. Critically, true lie detection methods may never be admissible because their role overlaps with the traditional credibility-assessment

role of the jury. Memory detection should not suffer from the same pitfall, but without making the distinction between the tests clear in the literature, it will be more difficult for nonexperts to properly treat the tests separately.

- Researchers should seek collaboration with law enforcement whenever possible. Law enforcement personnel are in the ideal position to explain likely problems that the CIT would encounter in practice, such as the difficulty of finding appropriate probe items. Some scholars associated with law enforcement have already written on this topic (e.g., Krapohl, 2011), but there remains a dearth of information.
- Researchers should seek to conduct experiments using multiple dependent measures (such as ANS, ERP, and fMRI measures). Such experiments would have multiple benefits. To the extent that various dependent measures yield results that are independent of each other, combining those measures could increase the sensitivity and specificity of the test. And to the extent that the dependent measures yield results that overlap, those results can be used to argue that results of one measure can be extrapolated to other measures. This may impact the *Daubert* analysis—if a party seeks to admit a P300-based CIT, for example, being able to argue that ANS-based field tests are relevant to the admissibility analysis would be a major benefit.

In sum, the *Daubert* analysis of the CIT has not, in my view, changed in any major way over the past few years, but there has been an incremental step made toward admissibility. While we still lack field studies that are likely critical to admissibility, experimental studies have become more numerous and more externally valid, and the general acceptance, methodological consistency and sophistication, and rigor of the results have continued to strengthen.

OTHER CONSTITUTIONAL ISSUES WITH CONCEALED INFORMATION TEST USE

The previous discussion addressed the complex issue of what would be necessary for the CIT to pass muster under American evidentiary principles, but those are not the only requirements that the CIT would have to satisfy before it could be used in criminal trials. Constitutional protections also limit the extent to which evidence can be involuntarily seized from individuals and then later used against them at trial. There are two Amendments to the US Constitution that are most relevant: the Fourth Amendment,

which protects an individual from "unreasonable" government searches and seizures, and the Fifth Amendment, which protects an individual from being "compelled in any criminal case to be a witness against himself." How a court would consider a compelled CIT under these principals is an entirely novel question, but scholars have begun to consider these questions, and I will briefly summarize some of the positions taken.

The Fourth Amendment guarantees "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures." In determining whether a search is reasonable, courts first determine whether there was a search at all. The US Supreme Court has developed a two-part test in which there is a search when an individual exhibits an actual, subjective expectation of privacy in the thing searched and, most importantly, that expectation, "viewed objectively, is justifiable under the circumstances" (Smith v. Maryland, 1979). Where there is an objectively justifiable expectation of privacy held by the individual subject to the search, government agents must generally obtain a warrant prior to conducting the search, or the evidence found in the search will be suppressed.

Though no court has addressed whether conducting a CIT on an individual would be considered a search under the Fourth Amendment, scholars have generally agreed that compelled mental tests are likely to be considered searches. The Supreme Court has found that compulsion of a physical substance from an individual, such as obtaining a sample of blood or urine, is a search under the Fourth Amendment (Schmerber v. California, 1966; Skinner v. Railway Labor Executives' Association, 1989; for discussion, see Pardo, 2006). And retrieval of information that is emitted outside of a location, such as heat waves from a house (or, by analogy, brain activity measured outside the skull), can still be considered a search where there is a reasonable expectation of privacy in the information searched (Kyllo v. United States, 2001).

However, even if a compelled CIT would be considered a search, scholars appear to agree that government actors would be able to obtain a warrant for that search. In order to obtain a warrant, the government must show that "there is a fair probability that contraband or evidence of a crime will be found in a particular place" (Illinois v. Gates, 1983). And it would likely not be difficult to demonstrate that the evidence of the crime—in a CIT, the presence of crime-related knowledge held by a suspect—would probably be found if a CIT were conducted (e.g., Pardo, 2006; Shen, 2013).

The Fifth Amendment issue is far more complex. That amendment protects an individual from being "compelled in any criminal case to be a witness against himself." However, what type of evidence is considered compelled, and whether compelled evidence is protected by the privilege, is the subject of substantial legal doctrine. For our purposes, the most important distinction is between physical materials, such as a blood or fingerprint sample (which can be compelled without violating the Fifth Amendment) and communicative acts or statements (which are protected under the privilege).

As a number of scholars have observed, CIT evidence, whether obtained through ANS-based methods or neuroscience-based methods, does not neatly fit into this dichotomy. As Dov Fox succinctly put it, "[b]rain imaging is difficult to classify because it promises distinctly testimonial-like information about the content of a person's mind that is packaged in demonstrably physical-like form, either as blood flows in the case of fMRI, or as brainwaves in the case of EEG" (Fox, 2009, p. 791). And the CIT, in many cases, does not even require a voluntary response on the part of the participant—in many ways, it is similar to a blood sample, but instead, it is a sample of neural activity under certain conditions. And yet, the Supreme Court has opined that "[t]o compel a person to submit to testing in which an effort will be made to determine his guilt or innocence on the basis of physiological responses, whether willed or not, is to evoke the spirit and history of the Fifth Amendment" (Schmerber v. California, 1966, p.764).

Some have argued that CIT-like evidence should be considered testimonial because it is *evoked* by a question or stimulus, unlike a blood sample or other physical evidence that was already present prior to any questioning or the presentation of any stimuli (Farahany, 2012; Pustilnik, 2013; for related arguments see Barillare, 2006; Murphy & Greely, 2011). Others have argued that the evidence is not testimonial because it does not put the suspect in the "cruel trilemma" (of either perjuring himself, putting himself in contempt of court by refusing to testify, or incriminating himself) that the Fifth Amendment was designed to protect against (Hurd, 2012) or because the test does not compel any action or behavior on the part of the suspect (Holley, 2009).

Given the Court's long history of mistrust of polygraph evidence and the compulsion of thoughts, I think it is likely that, when confronted with the issue, courts are likely to either fit CIT-like evidence within the current testimonial framework or modify the test so that such evidence is protected from compulsion under the Fifth Amendment. But that does not mean that

the test will not be useful in criminal investigations. For example, in Japan, consent from the suspect is required before a CIT is given, and yet hundreds of CITs are still conducted annually (Osugi, 2011). The test may also be useful outside the court as well; as Danaher (2015) suggests, the CIT may serve as a signaling tool to help parties better evaluate an appropriate plea bargain.

On the whole, these issues are still extremely new, and we lack direction from courts as to how the analysis will likely proceed. As technology continues to develop, it will become more and more likely that courts will be confronted with these and similar issues, and they will help to shape Fourth and Fifth Amendment jurisprudence going forward.

REFERENCES

Alexander, A. (2007). Functional magnetic resonance imaging lie detection: Is a brainstorm heading toward the gatekeeper? *Houston Journal of Health Law and Policy, 7*, 1.

Barillare, J. C. (2006). As its next witness, the state calls… the defendant: Brain fingerprinting as "testimonial" under the fifth amendment. *Temple Law Review, 79*, 971–1004.

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

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology, 94*, 1–11.

Bernstein, D. E. (1994). The admissibility of scientific evidence after Daubert v. Merrell Dow Pharmaceuticals, Inc. *Cardozo Law Review, 15*, 2139–2181.

Black, B., Ayala, F. J., & Saffran-Brinks, C. (1994). Science and the law in the wake of Daubert: A new search for scientific knowledge. *Texas Law Review, 72*, 715–802.

Bradley, M. T., Malik, F. J., & Cullen, M. C. (2011). Memory instructions, vocalization, mock crimes, and concealed information tests with a polygraph. *Perceptual and Motor Skills, 113*, 840–858.

Breska, A., Ben-Shakhar, G., & Gronau, N. (2012). Algorithms for detecting concealed knowledge among groups when the critical information is unavailable. *Journal of Experimental Psychology: Applied, 18*, 292–300.

Capra, D. J. (1998). The Daubert puzzle. *Georgia Law Review, 32*, 669–782.

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.

Chen, E. K., & Yoon, A. H. (2005). Does Frye or Daubert matter? A study of scientific admissibility standards. *Virginia Law Review, 91*, 471–512.

Danaher, J. (2015). The comparative advantages of brain-based lie detection: The P300 concealed information test and pre-trial bargaining. *The International Journal of Evidence and Proof, 19*, 52–66.

Daubert v. Frye, A State-by-State Comparison. https://www.theexpertinstitute.com/daubert-v-frye-a-state-by-state-comparison/.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 US 579 (1993).

Deng, X., Rosenfeld, J. P., Ward, A., & Labkovsky, E. (2016). Superiority of visual (verbal) vs. auditory test presentation modality in a P300-based CIT: The complex trial protocol for concealed autobiographical memory detection. *International Journal of Psychophysiology, 105*, 26–34.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology, 75*, 521.

Elaad, E. (2016). Extracting critical information from group members' partial knowledge using the searching concealed information test. *Journal of Experimental Psychology: Applied, 22*, 500–509.

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology, 26*, 442–451.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77*, 757.

Ellenberg, C. (2009). Lie setection: A changing of the guard in the quest for truth in court? *Law and Psychology Review, 33*, 139–149.

Erickson, M. J. (2007). Daubert's bipolar treatment of scientific expert testimony—from Frye's polygraph to Farwell's brain fingerprinting. *Drake Law Review, 55*, 763–812.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (2000). How good is good enough? Expert evidence under Daubert and Kumho. *Case Western Reserve Law Review, 50*, 645–667.

Farahany, N. A. (2012). Incriminating thoughts. *Stanford Law Review, 64*, 351–408.

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.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*, 531–547.

Farwell, L. A., Richardson, D. C., & Richardson, G. M. (2013). Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information. *Cognitive Neurodynamics, 7*, 263–299.

Farwell, L. A., Richardson, D. C., Richardson, G. M., & Furedy, J. J. (2014). Brain fingerprinting classification concealed information test detects US Navy military medical information with P300. *Frontiers in Neuroscience, 2014*, 1–21.

Federal rule of evidence 403.(2011).

Federal rule of evidence 701.(2000).

Federal rule of evidence 702.(1975).

Fenner, G. M. (1996). The Daubert handbook: The case, its essential dilemma, and its progeny. *Creighton Law Review, 29*, 939–1089.

Fox, D. (2009). The right to silence as protecting mental control. *Akron Law Review, 42*, 763–801.

Furedy, J. J., & Ben-Shakhar, G. (1991). The role of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. *Psychophysiology, 28*, 163–171.

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the guilty actions test. *Biological Psychology, 83*, 101–107.

Greely, H. T., & Illes, J. (2007). Neuroscience-based lie detection: The urgent need for regulation. *American Journal of Law and Medicine, 33*, 377–421.

Guadet, L. M. (2011). Brain fingerprinting, scientific evidence, and daubert: A cautionary lesson from India. *Jurimetrics, 51*, 293–319.

Harrington v. State, No. PCCV 073247 (Iowa Dist. Ct. Mar. 5, 2001).

Harrington v. State, 659 N.W.2d 509 (Iowa 2003).

Heinzerling, L. (2006). Doubting Daubert. *Journal of Law and Policy, 14*, 65–84.

Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Application of the guilty knowledge test in forensic investigations. *International Journal of Police Science and Management, 4*, 15–27.

Holley, B. (2009). It's all in your head: Neurotechnological lie detection and the fourth and fifth amendments. *Developments in Mental Health Law, 28*, 1–23.

Hu, X., Bergström, 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, 26*(7).

Hurd, A. J. (2012). Reaching past fingertips with forensic neuroimaging—non-"testimonial" evidence exceeding the fifth amendment's grasp. *Loyola Law Review, 58*, 213–248.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82*, 426.

Illinois v. Gates, 462 U.S. 213 (1983).

Jonakait, R. N. (1994). The meaning of Daubert and what it means for forensic science. *Cardozo Law Review, 15*, 2103–2117.

Kittay, L. (2007). Admissibility of fMRI lie detection: The cultural bias against "mind reading" devices. *Brooklyn Law Review, 72*, 1351–1399.

Kleinberg, B., & Verschuere, B. (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. *Journal of Applied Research in Memory and Cognition, 5*, 43–51.

Kobayashi, T., Yoshimoto, K., & Fujihara, S. (2009). The contemporary situation of field polygraph tests. *Japanese Journal of Physiological Psychology and Psychophysiology, 27*, 5–15.

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. 151–170). Cambridge: Cambridge University Press.

Kyllo v. United States, 533 U.S. 27 (2001).

Labkovsky, E., & Rosenfeld, J. P. (2014). A novel dual probe complex trial protocol for detection of concealed information. *Psychophysiology, 51*, 1122–1130.

Langleben, D. D. (2008). Detection of deception with fMRI: Are we there yet? *Legal and Criminological Psychology, 13*, 1–9.

Langleben, D. D., Hakun, J. G., Seelig, D., Wang, A. L., Ruparel, K., Bilker, W. B., & Gur, R. C. (2016). Polygraphy and functional magnetic resonance imaging in lie detection: A controlled blind comparison using the concealed information test. *The Journal of Clinical Psychiatry, 77*, 1372–1380.

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.

Lui, M., & Rosenfeld, J. P. (2008). Detection of deception about multiple, concealed, mock crime items, based on a spatial temporal analysis of ERP amplitude and scalp distribution. *Psychophysiology, 45*, 721–730.

Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology, 110*, 56–65.

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

Matsuda, I., Nittono, H., & Allen, J. B. (2012). The current and future status of the concealed information test for field use. *Frontiers in Psychology, 2012*, 1–11.

Meijer, E. H., Ben-Shakhar, G., Verschuere, B., & Donchin, E. (2013). A comment on Farwell (2012): Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics, 7*, 155–158.

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 Psychology, 2013*, 1–6.

Meijer, E. H., Selle, N. K., 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.

Meijer, E. H., Smulders, F. T., & Merckelbach, H. L. (2010). Extracting concealed information from groups. *Journal of Forensic Sciences, 55*, 1607–1609.

Meixner, J. B. (2012). Liar liar, jury' the trier: The future of neuroscience-based credibility assessment in the court. *Northwestern University Law Review, 106*, 1451–1488.

Meixner, J. B., & Diamond, S. S. (2014). The hidden Daubert factor: How judges use error rates in assessing scientific evidence. *Wisconsin Law Review, 2014*, 1063–1134.

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology, 48*, 149–154.

Meixner, J. B., & Rosenfeld, J. P. (2014). Detecting knowledge of incidentally acquired, real-world memories using a P300-based concealed-information test. *Psychological Science, 25*, 1994–2005.

Mertens, R., & Allen, J. B. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual mock crime scenarios. *Psychophysiology, 45*, 286–298.

Moreno, J. A. (2009). The future of neuroimaged lie detection and the law. *Akron Law Review, 42*, 717–737.

Moriarty, J. C. (2009). Visions of deception: Neuroimages and the search for truth. *Akron Law Review, 42*, 739–761.

Murphy, E. R., & Greely, H. T. (2011). What will be the limits of neuroscience-based mindreading in the law. In J. Illes, & B. J. Sahakian (Eds.), *Oxford handbook of neuroethics* (p. 642).

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.

Ogawa, T., Matsuda, I., & Tsuneoka, M. (2015). The comparison question test versus the concealed information test? That was the question in Japan: A comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology, 95*, 29–30.

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: Cambridge University Press.

Pardo, M. S. (2006). Neuroscience evidence, legal culture, and criminal procedure. *American Journal of Criminal Law, 33*, 301–337.

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.

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.

Podlesny, J. A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest, 20*, 57–61.

Pustilnik, A. C. (2013). Neurotechnologies at the intersection of criminal procedure and constitutional law. In S. Richardson, & J. Parry (Eds.), *The constitution and the future of the criminal law*.

Rosenfeld, J. P. (2005). Brain fingerprinting: A critical analysis. *The Scientific Review of Mental Health Practice, 4*, 20–37.

Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 42*, 157–161.

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.

Rosenfeld, J. P., Labkovsky, E., Lui, M. A., Winograd, M., Vandenboom, C., & Chedid, K. (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.

Rosenfeld, J. P., Ward, A., Drapekin, J., Labkovsky, E., & Tullman, S. (2017). Instructions to suppress semantic memory enhances or has no effect on P300 in a concealed information test (CIT). *International Journal of Psychophysiology, 113*, 29–39.

Rosenfeld, J. P., Ward, A., Frigo, V., Drapekin, J., & Labkovsky, E. (2015). Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, complex trial protocol for concealed autobiographical memory detection. *International Journal of Psychophysiology, 96*, 16–22.

Schauer, F. (2010). Can bad science be good evidence? Neuroscience, lie detection, and beyond. *Cornell Law Review, 95*, 1191–1219.

Schmerber v. California, 384 U.S. 757 (1966).

Shapiro, Z. E. (2016). *Harvard Journal of Law and Technology, 29*, 527–549.

Shen, F. X. (2013). Neuroscience, mental privacy, and the law. *Harvard Journal of Law and Public Policy, 36*, 653–713.

Skinner v. Railway Labor Executives' Assocation, 489 U.S. 602 (1989).

Slaughter v. State, 105 P.3d 832 (Okla. Crim. App. 2005).

Smith v. Maryland, 442 U.S. 735 (1979).

Smith v. Maryland, 16 A.3d 977 (Md. Ct. App. 2011).

Spence, S. A. (2008). Playing devil's advocate: The case against fMRI lie detection. *Legal and Criminological Psychology, 13*, 11–25.

United States Constitution, Fourth Amendment.

United States Constitution, Fifth Amendment.

United States v. Frye, 293 F. 1013 (D.C. Cir. 1923).

United States v. Scheffer, 523 US 303 (1998).

United States v. Semrau, 693 F.3d 510 (6th Cir. 2012).

United States v. Semrau, No. 07–10074 Ml/P, 2010 WL 6845092 (W.D. Tenn. June 1, 2010).

Ward, A. C., & Rosenfeld, J. P. (2017). Attempts to suppress episodic memories fail but do produce demand: Evidence from the P300-based complex trial protocol and an implicit memory test. *Applied Psychophysiology and Biofeedback, 42*, 13–26.

Wilson v. Corestaff Services, L.P., 900 N.Y.S.2d 639, 640 (Sup. Ct. 2010).

Winograd, M. R., & Rosenfeld, J. P. (2011). Mock crime application of the complex trial protocol (CTP) P300-based concealed information test. *Psychophysiology, 48*, 155–161.

Woodruff, W. A. (2014). Evidence of lies and rules of evidence: The admissibility of fMRI-based expert opinion of witness truthfulness. *North Carolina Journal of Law and Technology, 16*, 105–252.

Zaitsu, W. (2016). External validity of concealed information test experiment: Comparison of respiration, skin conductance, and heart rate between experimental and field card tests. *Psychophysiology, 53*, 1100–1107.

Zvi, L., Nachson, I., & Elaad, E. (2012). Effects of coping and cooperative instructions on guilty and informed innocents' physiological responses to concealed information. *International Journal of Psychophysiology, 84*, 140–148.