Printed in U.S.A.
Northwestern University Law Review
Vol. 106, No. 3

LIAR, LIAR, JURY'S THE TRIER? THE FUTURE OF NEUROSCIENCE-BASED CREDIBILITY ASSESSMENT IN THE COURT

John B. Meixner

ABSTRACT—Neuroscience-based creditability-assessment tests have recently become increasingly mainstream, purportedly able to determine whether an individual is lying to a certain set of questions (the Control Question Test) or whether an individual recognizes information that only a liable person would recognize (the Concealed Information Test). Courts have hesitated to admit these tests as evidence for two primary reasons. First, following the general standard that credibility assessment is a matter solely for the trier of fact, courts exclude the evidence because it impinges on the province of the jury. Second, because these methods have not been rigorously tested in realistic scenarios, courts rule that they do not meet the *Daubert* criteria for admissibility of expert testimony. This Comment argues that while neuroscience-based credibility-assessment methods should not currently be admissible under the *Daubert* standard, they may become admissible with more research, and the courts should avoid creating precedent that would preclude their admissibility once reliability issues are addressed. Specifically, credibility assessment should not be left entirely to the trier of fact because social science evidence indicates that laypeople are poor at making credibility-assessment judgments based on behavioral cues. Additionally, even if courts continue to rule that evidence assessing whether a witness is telling the truth invades the province of the jury, this should not preclude neuroscience-based credibility assessment that merely shows that an individual recognizes something related to the issue at hand.

AUTHOR—J.D., Ph.D. Candidate, Northwestern University School of Law, Northwestern University Department of Psychology; B.S., University of Michigan, 2006. I am extremely grateful to Professor Shari Seidman Diamond for supporting this project and for continually encouraging my interdisciplinary work; Professor J. Peter Rosenfeld, Michael Winograd, and Xiaoqing Hu for their extremely insightful comments and excellent research on this topic; and Jane Moriarty, Francis X. Shen, Jay Koehler, and Hank Greely for providing wonderful feedback on earlier drafts. I also owe special thanks to the Editorial Board of the *Northwestern University Law Review*, especially Peter Siegal, Jason Cairns, Lyndsie Schmalz, Laura Kolesar, and Elizabeth Uzelac, for their helpful comments and editorial guidance. Finally, I cannot give enough thanks to my wife Kate and my family for their endless love and support. All errors are my own.

INTRODUCTION

In 2010, Dr. Lorne Semrau was indicted for fraud after being accused of intentionally submitting false claims for payment to health insurance providers.[1] Maintaining his innocence, Dr. Semrau hired Cephos,[2] a forensic analysis company, to conduct a neuroscience-based credibility-assessment test to show that he was telling the truth in his denials of knowingly defrauding the government.[3] Dr. Semrau climbed into a functional magnetic resonance imaging (fMRI) scanner and answered a series of yes–no questions about the alleged fraud[4]: "Did you . . . cheat or defraud Medicare?"[5] "Did you enter into a scheme to defraud the government by

[1] United States v. Semrau, No. 07-10074 Ml/P, 2010 WL 6845092, at *2 (W.D. Tenn. May 31, 2010).

[2] Cephos is one of two companies in the United States that provides a neuroscience-based lie-detection service. CEPHOS, http://www.cephoscorp.com (last visited Aug. 4, 2012). The other is No-Lie MRI. NO LIE MRI, http://noliemri.com/ (last visited Aug. 4, 2012). Both of these companies use functional magnetic resonance imaging (fMRI) to attempt to identify brain activity that is associated with lying. CEPHOS, *supra*; NO LIE MRI, *supra*. While Cephos lie-detection tests have been offered (and rejected) as evidence in two cases (*Semrau*, 2010 WL 6845092, at *3, and *Wilson v. Corestaff Servs. L.P.*, 900 N.Y.S.2d 639, 640 (Sup. Ct. 2010)), there have been no recorded instances of parties attempting to introduce No Lie MRI tests as evidence.

[3] *Semrau*, 2010 WL 6845092, at *4.

[4] *Id.* at *4–8. fMRI is a method used to determine the level of activity in various parts of the brain. *See* Nikos K. Logothetis & Brian A. Wandell, *Interpreting the BOLD Signal*, 66 ANN. REV. PHYSIOLOGY 735, 735 (2004).

[5] *Semrau*, 2010 WL 6845092, at *5.

billing for . . . tests conducted by psychiatrists . . . ?"[6] "Have you ever done something illegal?"[7] Dr. Steven Laken, the CEO of Cephos, subsequently analyzed Dr. Semrau's brain responses to the questions and determined that Dr. Semrau was "not deceptive."[8] At trial, Dr. Semrau attempted to introduce these brain scans as evidence through the expert witness testimony of Dr. Laken.[9] The court conducted a *Daubert* hearing to determine the reliability of the test, concluding that it was not sufficiently reliable to be admitted, primarily because its accuracy had not been thoroughly tested in "real-world" settings outside of the laboratory.[10]

Contemporaneously with Dr. Semrau's trial, an individual named Cynette Wilson brought suit against a temporary employment agency, alleging the agency retaliated against her by not placing her in work assignments after she reported an incident of sexual harassment.[11] Ms. Wilson's key witness was Ronald Armstrong, an employee of the defendant company, who testified that the company had instructed him not to place Ms. Wilson in work assignments because of her sexual harassment complaints.[12] Like Dr. Semrau, Ms. Wilson hired Cephos to conduct a lie-detection test, this time on Mr. Armstrong to reinforce his credibility.[13] Dr. Laken conducted the test on Mr. Armstrong and asserted that he could tell "to a very high probability that Armstrong [was] being truthful when he testifie[d]."[14]

Like the court in *United States v. Semrau*,[15] the *Wilson v. Corestaff Services L.P.*[16] court rejected the expert testimony, but relied upon entirely different reasoning. Rather than asserting issues of reliability, the *Wilson* court stated that "credibility is a matter solely for the jury . . . ."[17] Thus,

---

[6] *Id.*

[7] *Id.* at *4.

[8] *Id.* at *6.

[9] *Id.* at *1.

[10] *See id.* at *9–14. Additionally, because Dr. Semrau was retested after being originally identified as deceptive, Dr. Laken violated Cephos's standard protocol. *See id.* at *13. Under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the controlling standard for admissibility of expert testimony in federal cases, "the existence and maintenance of standards controlling [a] technique's operation" is a factor to consider in ruling on the admissibility of "a particular scientific technique." 509 U.S. 579, 594 (1993). Presumably, the lack of a followed standard by Dr. Laken in *Semrau* would cut against admitting the test.

[11] *Wilson v. Corestaff Servs. L.P.*, 900 N.Y.S.2d 639, 640 (Sup. Ct. 2010).

[12] *Id.*

[13] *Id.*

[14] *Id.* (internal quotation marks omitted).

[15] 2010 WL 6845092.

[16] 900 N.Y.S.2d 639.

[17] *Id.* at 642. It should be noted that the *Wilson* case was a New York state court case, and New York follows the *Frye* standard for admissibility of expert testimony. *See id.* at 640. The *Frye* test requires that evidence "must be sufficiently established to have gained general acceptance in the particular field in which it belongs." Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). In

because the expert testimony concerned the credibility of the witness and therefore "impinge[d] on the province of the jury[,] . . . [it] should be treated with a great deal of skepticism."[18]

These cases indicate two distinct obstacles to the admissibility of neuroscience-based credibility assessment: first, rules that restrict all credibility assessment to the trier of fact and preclude the use of expert testimony as an aid to the jury's credibility assessment (the issue in *Wilson*), and second, the unreliability of the tests that prevents them from gaining admissibility under *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (the issue in *Semrau*).

This Comment argues that while neuroscience-based credibility-assessment methods are not currently admissible under the *Daubert* standard, further research may prove their reliability and therefore render them admissible. Thus, the courts should avoid creating precedent that would preclude their admissibility once these reliability issues are addressed. Specifically, because social science evidence indicates that laypeople are poor at making credibility-assessment judgments based on behavioral cues, credibility assessment should not be left entirely to the trier of fact if other tools, such as neuroscience, can aid them in their decision. Additionally, even if courts continue to rule that evidence that directly assesses whether a witness is telling the truth invades the province of the jury, this should not preclude neuroscience-based credibility assessment that merely shows that an individual possesses knowledge related to the issue at hand.

Part I of this Comment provides an overview of two major types of neuroscience-based credibility assessment, explains the theoretical bases for each of those two methods, and describes the research assessing their accuracy. Part II reviews the social science literature on the ability of laypeople to assess credibility based on behavioral cues and asserts that judges and jurors are unlikely to effectively determine which witnesses are credible, and thus should not be the sole determiners of credibility. Part III applies the four *Daubert* factors to neuroscience-based credibility assessment and concludes that while the currently available tests should be inadmissible, there is promise for neuroscience-based credibility assessment once more realistic testing is conducted.

---

contrast, federal courts follow the *Daubert* standard under the Federal Rules of Evidence, which asks whether the evidence is generally reliable. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594–95 (1993).

[18] *Wilson*, 900 N.Y.S.2d at 642. The court rejected the evidence solely on this basis, though it briefly mentioned that the evidence would also likely be disqualified because it was not derived from a method generally accepted in the scientific community, as required under *Frye*. *Id.*

I. BACKGROUND

A. What Is Neuroscience-Based Credibility Assessment?

This Comment defines "neuroscience-based credibility assessment" as any tool that can be used to assess the veracity of a witness's statement through measurements of brain activity. In particular, this Comment will focus on two very distinct types of neuroscience-based tests: the Control Question Test (CQT), which attempts to measure whether a person is telling the truth, and the Concealed Information Test (CIT), which attempts to measure whether a person has knowledge relating to a crime or other event. Both of these tests were traditionally conducted on the polygraph machine[19] but have since been adapted for use with techniques that measure brain activity.

B. The Control Question Test

The test that tends to come to mind most quickly when a layperson envisions a lie-detector test is the CQT. The CQT seeks to determine whether an individual is lying about a particular question by comparing the individual's physiological response to that question with the individual's physiological response to a "control" question.[20]

The test consists of two critical items: relevant questions and control questions.[21] Relevant questions are germane to the subject of the investigation (e.g., "Did you shoot your wife on the night of September 16th, 2004?").[22] In contrast, control questions are deliberately vague questions about past actions that relate to the complementary relevant question (e.g., "Prior to September 16th, 2004, had you ever hurt anyone?").[23] Control questions are designed such that nearly any honest examinee's truthful answer would be "yes."[24] However, through the interrogation process, the examinee is led to believe that he *should* be able to truthfully answer all of these control questions "no," (because a "yes"

---

[19] *See* David C. Raskin, *The Polygraph in 1986: Scientific, Professional, and Legal Issues Surrounding Application and Acceptance of Polygraph Evidence*, 1986 UTAH L. REV. 29, 31. The polygraph measures "palmar sweating, blood pressure increases and respiratory changes" to determine the veracity of a statement or the presence of concealed information. *Id.*

[20] *See* William G. Iacono & David T. Lykken, *The Case Against Polygraph Tests*, *in* 5 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 40:47 (David L. Faigman et al. eds., 2011–2012). The "control" question is sometimes instead referred to as a "comparison question" because it is not a true control in the scientific sense of the word. *Id.* at § 40:48. The control question is not as emotionally charged as the relevant question, making it difficult to compare the two. *See id.* § 40:73. The CIT does not have the same problem, as is discussed *infra* Part I.C.

[21] *See id.* § 40:47.

[22] *Id.*

[23] *Id.*

[24] *See* William G. Iacono, *Detection of Deception*, *in* HANDBOOK OF PSYCHOPHYSIOLOGY 688, 688 (John Cacioppo et al. eds., 3d ed. 2007).

answer would be indicative of poor moral character, which the examinee is led to believe may be used against him in the investigation), and thus the examinee feels obligated to lie in response to the control questions.[25] These control questions are then used as an approximation of the innocent examinee's deceptive response, which is then compared to the examinee's response to the relevant question.[26] If the physiological response to the relevant question is stronger than the physiological response to the control question, deception is inferred.[27]

Though popular among police departments and government agencies for eliciting confessions and screening potential employees, the CQT has come under criticism for its low accuracy rates.[28] A recent report from the National Research Council stated that the theoretical rationale for the CQT is flawed, as an elevated physiological response to the relevant question would be expected even from innocent individuals because the implications of the relevant question are clear.[29] For this reason, the CQT elicits many false positives—cases in which truthful individuals are classified as deceptive.[30]

For example, suppose an individual is given a CQT test to determine whether he cheated on his 2003 tax return. A control question designed to elicit a lie might be "Have you ever taken something that didn't belong to you?" Even though this question may elicit a lie (presumably a "no" response is a lie), and therefore some physiological arousal, when the participant hears the critical question, "Did you cheat on your 2003 tax return?," he knows that this is the critical question that will influence whether he is prosecuted for a serious crime. Even if he is honest when he answers "no" to this question, the extreme salience of its relationship to the issue at hand may cause elevated physiological responses beyond that of the control question, leading the examiners to believe that the accused is lying when he is in fact telling the truth. Thus, though the polygraph-based CQT is sometimes proffered in court, it is rarely admitted, with a majority of states and federal courts disallowing it.[31] Moreover, the 1923 landmark case regarding the admissibility of scientific evidence, *Frye v. United States*, rejected a polygraph-based precursor to the CQT because the test was not generally accepted in the scientific community.[32]

---

[25] *See id.*

[26] *See id.* at 688–89.

[27] *See id.*

[28] *See, e.g.*, NAT'L RESEARCH COUNCIL, THE POLYGRAPH AND LIE DETECTION xiii, 19, 124–29 (2003); Iacono, *supra* note 24, at 690, 693–94.

[29] *See* NAT'L RESEARCH COUNCIL, *supra* note 28, at 92–95; Iacono, *supra* note 24, at 694.

[30] *See* Iacono, *supra* note 24, at 693.

[31] *See* Jane Campbell Moriarty, *Visions of Deception: Neuroimages and the Search for Truth*, 42 AKRON L. REV. 739, 743–44 (2009).

[32] 293 F. 1013, 1014 (D.C. Cir. 1923). In *Frye*, the court rejected a "systolic blood pressure deception test" used to detect deception that was highly related to the modern polygraph-based CQT. *Id.*

For more than fifty years, the CQT was conducted on the polygraph machine.[33] The key assumption of the test was that the increased anxiety associated with lying to the relevant questions would result in increased sweating.[34] Recently, however, the CQT has been adapted for use with fMRI.[35] Rather than measuring an indirect indicator of anxiety such as sweating or heart rate, fMRI allows one to measure blood flow to the brain, which is a better proxy for brain activity (and perhaps actual deception itself) than are external measurements. fMRI measures brain function by recording the "changes in blood flow that correspond to changes in local brain activity."[36] Thus, when a particular area of the brain receives a greater quantity of oxygenated blood than other areas, one can assume that area has been more active.[37] By identifying particular areas of the brain that are associated with lying, an examiner could theoretically distinguish truth from lies by examining which areas of the brain become active when a subject answers a question.

Attempts to use this method to distinguish liars from truth tellers have achieved mixed results. At least one study has reported accuracy rates as high as 90%,[38] though the design used some unrealistic methods,[39] and another study found only 78% accuracy.[40] The relative dearth of this type of research and the lack of replication of past studies make it difficult to determine the true accuracy of fMRI-based lie detection. The use of fMRI or other neuroimaging methods in general has rarely been admitted in the

---

at 1013–14. The defendant in the case offered the deception test as evidence of her innocence. *Id.* The court rejected the test because it "ha[d] not yet gained . . . scientific recognition." *Id.* at 1014.

[33] The original CQT was developed by John Reid in 1947 in response to the shortcomings of a previous similar polygraph technique called the relevant–irrelevant test. *See* Charles R. Honts et al., *The Case for Polygraph Tests*, *in* 5 MODERN SCIENTIFIC EVIDENCE, *supra* note 20, § 40:22. This method was developed for the polygraph and continues to be used primarily with the polygraph. *See id.*

[34] Though palmar sweating tends to be the primary measure used in polygraph tests, blood pressure increases and respiratory changes are also frequently considered. *See, e.g.*, Raskin, *supra* note 19, at 31.

[35] *See, e.g.*, F. Andrew Kozel et al., *Detecting Deception Using Functional Magnetic Resonance Imaging*, 58 BIOLOGICAL PSYCHIATRY 605, 605 (2005); Daniel D. Langleben et al., *Telling Truth from Lie in Individual Subjects with Fast Event-Related fMRI*, 26 HUM. BRAIN MAPPING 262, 262–63 (2005).

[36] Daniel D. Langleben et al., *True Lies: Delusions and Lie-Detection Technology*, 34 J. PSYCHIATRY & L. 351, 359 (2006).

[37] *See, e.g.*, Logothetis & Wandell, *supra* note 4, at 760 ("The BOLD response provides us unprecedented visibility of the neural activity in the human brain . . . ."). The term "BOLD" stands for blood oxygen level dependent—it is a signal that measures blood flow to various areas of the brain. *See id.* at 735–36.

[38] Kozel et al., *supra* note 35, at 610.

[39] The study involved the use of a "model building group" to identify approximately which neural networks were involved in deception for the experiment's particular paradigm. *Id.* at 605–06. This method could not likely be used in field scenarios, where individual questions must be unique for each person tested.

[40] Langleben et al., *supra* note 35, at 267.

courtroom,[41] and a lie-detection test using fMRI has never been ruled admissible.

C. The Concealed Information Test

The Concealed Information Test (CIT)[42] is a credibility-assessment protocol of an entirely different nature than the CQT. Instead of attempting to detect actual lying (the goal of the CQT), the goal of the CIT is to determine whether an individual possesses knowledge of specific details of a crime or event.[43] For example, if a murder was committed at 800 Church Avenue using a .38 caliber revolver, the CIT seeks to determine whether a suspect recognizes the address and type of weapon:

> The CIT presents subjects with various stimuli, one of which is a crime-related item (the *probe*, such as the gun used to commit a murder). Other stimuli consist of control items that are of the same class (*irrelevants*, such as other potentially deadly weapons: a knife, a bat, etc.) such that an innocent person would be unable to discriminate them from the crime-related item. If the subject's physiological response is greater for the probe item than for irrelevant items, then knowledge of the crime or other event is inferred.[44]

Like the CQT, the CIT was initially conducted using the polygraph; the subject was expected to show elevated arousal when viewing crime-related probe items.[45] Since the original development of the CIT in the early 1960s, the test has been adapted to be used through the measurement of brainwaves. In 1965, Samuel Sutton and his colleagues discovered a particular pattern of neuronal firing that signaled an individual's recognition of a unique or meaningful item.[46] Because the response occurred approximately 300 milliseconds after the presentation of the meaningful item, the brainwave was eventually termed the "P300."[47] Just over twenty

---

[41] *See* Jane Campbell Moriarty, *Flickering Admissibility: Neuroimaging Evidence in the U.S. Courts*, 26 BEHAV. SCI. & L. 29, 39–40 (2008). For a case in which a PET scan was admitted to establish brain trauma, however, see Brown v. Allerton Assocs., No. 17917/03, 2006 WL 3102331 (N.Y. Sup. Ct. Oct. 5, 2006). There are other contexts in which neuroimaging has been used in the courtroom. For a comprehensive review, see Owen D. Jones & Francis X. Shen, *Law and Neuroscience in the United States*, *in* INTERNATIONAL NEUROLAW: A COMPARATIVE ANALYSIS 349 (Tade Matthias Spranger ed., 2012).

[42] The CIT is also frequently referred to in the literature as the "Guilty Knowledge Test" or GKT. *See* Honts et al., *supra* note 33, § 40:24.

[43] *See* Iacono & Lykken, *supra* note 20, § 40:107.

[44] John B. Meixner & J. Peter Rosenfeld, *A Mock Terrorism Application of the P300-Based Concealed Information Test*, 48 PSYCHOPHYSIOLOGY 149, 149 (2011).

[45] *See* David T. Lykken, *The GSR in the Detection of Guilt*, 43 J. APPLIED PSYCHOL. 385, 385 (1959) (using the polygraph to conduct the first published CIT).

[46] *See* Samuel Sutton et al., *Evoked-Potential Correlates of Stimulus Uncertainty*, 150 SCIENCE 1187, 1187 (1965). The item must also be infrequently presented in order to generate the P300 component. *See id.*

[47] *See id.* at 1188.

years later, researchers realized that the P300 could be used as a marker of recognition of crime-relevant information in a CIT—when an individual recognizes something she knows she used or saw during the commission of a crime, she should have a large P300 response when presented with that item.[48]

P300-based CITs have evolved over the past twenty years and have achieved accuracy levels beyond those of the CQT, typically correctly detecting individuals that possess concealed information at a rate at or above 90% while keeping false positives (innocent individuals not possessing concealed information that are misclassified) below 10%.[49] Though the CIT is not currently used by American police forces, it is used in Japan to help identify the perpetrators of crimes.[50] One recent CIT experiment reached 100% accuracy,[51] and the CIT has received media exposure for a number of its potential uses, including counterterrorism.[52] Despite these benefits over the CQT, the CIT has rarely been submitted into evidence, and it has never actually influenced the outcome of a trial.

To briefly summarize, we have now discussed two very different types of credibility-assessment paradigms: (1) the Control Question Test (CQT), which actually seeks to determine whether a statement or series of statements is true or false, and (2) the Concealed Information Test (CIT), which seeks to determine whether an individual possesses knowledge of information relevant to a crime or other event. It is important here to recognize that the CQT actually purports to determine whether an individual is *lying*, while the CIT involves no lying at all—it instead determines whether an individual *recognizes* an item that only a person involved in a crime would recognize. With these two paradigms as a foundation, we now turn to some of the potential hurdles that these tests must clear before they could be admitted in a court of law.

---

[48] *See, e.g.*, Lawrence A. Farwell & Emanuel Donchin, *The Truth Will Out: Interrogative Polygraphy ("Lie Detection") with Event-Related Brain Potentials*, 28 PSYCHOPHYSIOLOGY 531, 531–32 (1991); J. Peter Rosenfeld et al., *A Modified, Event-Related Potential-Based Guilty Knowledge Test*, 42 INT'L J. NEUROSCIENCE 157, 157–58 (1988).

[49] *See, e.g.*, Meixner & Rosenfeld, *supra* note 44, at 152–53 (reporting high levels of detection accuracy among twenty-four participants in a mock terrorist attack without any advance knowledge of the crime-related details); J. Peter Rosenfeld et al., *The Complex Trial Protocol (CTP): A New, Countermeasure-Resistant, Accurate, P300-Based Method for Detection of Concealed Information*, 45 PSYCHOPHYSIOLOGY 906, 913 tbl.3b (2008) (reporting 92% detection accuracy with only 8% false positives).

[50] *See* Gershon Ben-Shakhar et al., *Trial by Polygraph: Reconsidering the Use of the Guilty Knowledge Technique in Court*, 26 LAW & HUM. BEHAV. 527, 528, 536 (2002).

[51] Meixner & Rosenfeld, *supra* note 44, at 151 tbl.1.

[52] Eben Harrell, *Fighting Crime by Reading Minds*, TIME SCIENCE (Aug. 7, 2010), http://www.time.com/time/health/article/0,8599,2009131,00.html (covering Meixner & Rosenfeld, *supra* note 44).

II. THE JURY'S CREDIBILITY-ASSESSING ROLE AS A HURDLE TO ADMISSIBILITY

A. The Current Standard of "the Jury as Lie Detector"

Perhaps the most serious threat to the admissibility of neuroscience-based credibility-assessment evidence is the concept that "the *jury* is the lie detector,"[53] which leads to a viewpoint that "credibility is a matter *solely* for the jury."[54] This concern is regularly raised in cases in which one of the parties attempts to admit lie-detection evidence based on the polygraph.[55] Such a position would prevent the admission of any expert testimony regarding credibility, regardless of its reliability or probative value to the trier of fact.[56] Thus, even if a lie-detection tool achieved 100% accuracy when used in the hands of an expert, it would likely be precluded from use because it would "invade the . . . province of the jury"[57] and "[b]y its very nature . . . diminish the jury's role in making credibility determinations."[58]

The most recent Supreme Court case in which the role of the jury was addressed with regard to a credibility-assessment tool was *United States v. Scheffer*.[59] In *Scheffer*, an Air Force officer was court-martialed for methamphetamine use after failing a drug test.[60] Prior to receiving his test results, the defendant took a polygraph test,[61] which "indicated no deception."[62] By an 8–1 vote, the Court concluded that Military Rule of

---

[53] United States v. Scheffer, 523 U.S. 303, 313 (1998) (quoting United States v. Barnard, 490 F.2d 907, 912 (9th Cir. 1973), *cert. denied*, 416 U.S. 959 (1974)).

[54] Wilson v. Corestaff Servs. L.P., 900 N.Y.S.2d 639, 642 (Sup. Ct. 2010) (emphasis added).

[55] *See, e.g.*, *Scheffer*, 523 U.S. at 306–07, 313; State v. Porter, 698 A.2d 739, 742, 769 (Conn. 1997) (asserting that the principal problem in introducing polygraph-based lie-detection evidence is that it will "invade the fact-finding province of the jury"). It should be noted that the jury could still be the *final* determiner of credibility (and guilt) even if lie-detection evidence like the CQT were allowed into evidence. That is, the jury could still weigh the lie-detection evidence with all of the other evidence (e.g., demeanor, consistency of statements, etc.) and make a final decision as to whether they think the witness was being truthful. The argument made in *Scheffer*, however, is that the jury must be the *only* determiner of credibility, which would preclude lie-detection evidence, or any other evidence that directly speaks to the credibility of one of the witnesses. For extended discussion on this in the context of both the CQT and CIT, see *infra* Part II.D.

[56] *See, e.g.*, *Barnard*, 490 F.2d at 912 (excluding the testimony of a psychiatrist and psychologist called to testify to the likelihood that another witness was telling the truth).

[57] *Porter*, 698 A.2d at 769.

[58] *Scheffer*, 523 U.S. at 313.

[59] 523 U.S. 303.

[60] *Id.* at 306.

[61] The polygraph test that the defendant underwent was a CQT. *Id.* at 306 & n.1. The Court did not elaborate on the differences between the CQT and the CIT in the opinion, leaving it ambiguous as to whether the holding applies to polygraph tests of all varieties or merely to the polygraph-based CQT. For a discussion of this, see Megan J. Erickson, Note, Daubert*'s Bipolar Treatment of Scientific Expert Testimony—From* Frye*'s Polygraph to Farwell's Brain Fingerprinting*, 55 DRAKE L. REV. 763, 784–86 (2007).

[62] *Scheffer*, 523 U.S. at 306.

Evidence 707, which bars the admission of polygraph evidence in court-martial proceedings, does not violate a defendant's right to present a defense.[63]

However, the Justices did not agree on the reasoning by which they reached this holding.[64] Justice Thomas, joined by Chief Justice Rehnquist and Justices Scalia and Souter, opined that Rule 707 serves the government's legitimate interest in "[p]reserving the court members' core function of making credibility determinations in criminal trials."[65] Without citing any empirical work or even anecdotal evidence, Justice Thomas simply asserted that the members of the jury "are presumed to be fitted [to assess credibility] by their natural intelligence and their practical knowledge of men and the ways of men."[66] Justice Thomas stated that an expert who testifies regarding credibility is not providing "factual matters outside the jurors' knowledge," but is instead "supply[ing] the jury only with another opinion, in addition to its own, about whether the witness was telling the truth."[67]

The concurring Justices disagreed, stating that the principal opinion "overreach[ed]" in deciding that the introduction of lie-detection evidence diminishes the jury's role as the arbiter of credibility.[68] These Justices instead focused on the lack of scientific evidence regarding the accuracy of polygraph-based lie detection as the reason to prevent admissibility.[69] Indeed, the concurring opinion explicitly stated that a per se rule excluding expert testimony was unwise, as "some later case might present a more compelling case for introduction of the testimony than this one does."[70] Justice Stevens's dissent went even further, stating that potential injustice may result from blanket rules against the admissibility of entire categories of relevant but potentially unreliable evidence.[71]

Despite the fact that the principal opinion did not provide substantive support for its assertion that the jury has a unique monopoly on credibility

---

[63] *Id.* at 305, 308.

[64] *Id.* at 312–14, 318 (concurring Justices joining only in the judgment and in Parts I, II.A, and II.D of the opinion).

[65] *Id.* at 312–13. This was not the only consideration by which the Court reached its holding. The Court also considered the broad latitude of lawmakers in restricting the presentation of evidence without restricting due process, *id.* at 308, as well as the lack of scientific consensus regarding the reliability of polygraph tests, *id.* at 309–11. Though the Court never distinguished between the CQT and CIT, both of which may be conducted using the polygraph, one may assume that the Court was referring only to the CQT, as the literature cited is regarding the CQT.

[66] *Id.* at 313 (quoting Aetna Life Ins. Co. v. Ward, 140 U.S. 76, 88 (1891)).

[67] *Id.*

[68] *Id.* at 318.

[69] *Id.* ("Given the ongoing debate about polygraphs, I agree the rule of exclusion is not so arbitrary or disproportionate that it is unconstitutional.").

[70] *Id.*

[71] *Id.* at 327–30 (Stevens, J., dissenting).

assessment, the notion has continued to be popular. In *Wilson v. Corestaff Services L.P.*, the court rejected the plaintiff's offering of fMRI-based credibility evidence in large part because of a concern that the evidence would "impinge[] on the province of the jury."[72] Applying the *Frye* test for admissibility of scientific evidence,[73] the court did not cite a single peer-reviewed study of fMRI lie detection to indicate that the method had not reached general acceptance in the field.[74] Instead, the court asserted that the trial would be "complex and confusing . . . for the jury if it were faced with conflicting expert opinions, each with scientific authority to support it, upon the collateral matter of credibility."[75] The court assumed, without empirical assessment, that jurors are perfectly capable of assessing the credibility of witnesses.[76]

One question remains unanswered in these opinions: Is it a *bad* thing to "diminish the jury's role in making credibility determinations[?]"[77] In other words, *should* the jury be the sole decider regarding matters of credibility? This determination is inevitably based on the accuracy of jurors' credibility determinations. Are juries "unable or incompetent to evaluate the evidence and draw inferences and conclusions"[78] when it comes to issues of witness credibility? If one of the chief goals of the trial is to resolve disputes accurately,[79] the best approach is to admit whatever evidence will increase the accuracy of credibility judgments. If jurors are more accurate in their own credibility judgments than they would be with the assistance of expert

---

[72] 900 N.Y.S.2d 639, 642 (Sup. Ct. 2010).

[73] The *Frye* test outlines a standard where scientific evidence "must be sufficiently established to have gained general acceptance in the particular field in which it belongs." Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). For more discussion of this test, see *infra* Part III.A.

[74] *See Wilson*, 900 N.Y.S.2d at 641–42.

[75] *Id.* (quoting People v. Williams, 159 N.E.2d 549, 554 (N.Y. 1959)).

[76] *See id.* at 642. In rejecting the credibility-assessment evidence, the court stated that "it is well established that unless the jurors are unable or incompetent to evaluate the evidence and draw inferences and conclusions, the opinion of an expert, which intrudes on the province of the jury, is both unnecessary and improper." *Id.* (citing Kulak v. Nationwide Mut. Ins. Co., 351 N.E.2d 735 (1976)).

[77] United States v. Scheffer, 523 U.S. 303, 313 (1998).

[78] *Wilson*, 900 N.Y.S.2d at 642.

[79] There are other important potential goals of the trial system that may not favor outcome accuracy. For example, there may be a legitimate social goal of procedural justice in the trial system; that is, giving people the opportunity to be heard and maximizing the perceived legitimacy of the judicial system. *See* Tom R. Tyler & E. Allan Lind, *Procedural Justice*, *in* HANDBOOK OF JUSTICE RESEARCH IN LAW 65, 74–88 (Joseph Sanders & V. Lee Hamilton eds., 2001). Maximizing procedural justice could, in some cases, be in conflict with maximizing outcome accuracy. For example, if jurors make less accurate credibility assessments without the aid of expert testimony regarding credibility than they would with the aid of such testimony, their accuracy will be reduced by not admitting the expert testimony. However, the procedural justice may be increased by rejecting the testimony if perceived procedural fairness is strongly tied to receiving a judgment from peers regarding the truthfulness of testimony, rather than from an expert. While it is not clear how procedural justice may be influenced by experts testifying with regard to credibility, a complete discussion of this is beyond the scope of this Comment. The remainder of this Comment will focus on policies that maximize outcome accuracy.

testimony on the matter, then the testimony should not be allowed because it would confuse an already-accurate trier of fact. However, if jurors are inaccurate in their credibility assessments, a universal rule restricting credibility assessment to the jurors' own evaluations—thereby excluding expert testimony on the matter—could lead to less accurate decisions.

Thus, the merit of a rule that categorically excludes expert testimony regarding witness credibility turns on the ability of the trier of fact to assess witness credibility on its own. Because there is a well-developed literature on the ability of laypeople to assess credibility, empirical analysis can inform decisions regarding the rule. The remainder of Part II reviews the literature on laypeople's ability to assess credibility without the aid of assessment tools like the polygraph or fMRI, and concludes that jurors are likely to be inaccurate in assessing credibility based on demeanor cues. Additionally, while jurors may potentially be better when using contextual cues, the research is not sufficiently well-developed to reach a strong conclusion.

B. Credibility Assessment Based on Demeanor

It seems only natural to assume that individuals can detect lies of others based on factors like facial expressions, tone of voice, aversion of gaze, and general nervousness. After all, nearly everyone has had the experience of lying to a friend or parent and being unable to look that person in the eye. The notion that we are able to distinguish between truth and lies is ubiquitous in our media as well. For example, the television show *Lie to Me* focused on Dr. Cal Lightman, a fictional deception expert whose character is loosely based on deception researcher Dr. Paul Ekman.[80] According to the show's website:

> If you lie to Lightman, he'll see it in your face and your posture or hear it in your voice. If you shrug your shoulder, rotate your hand or even just slightly raise your lower lip, Lightman will spot the lie. By analyzing facial expressions and involuntary body language, he can read feelings ranging from hidden resentment to sexual attraction to jealousy.[81]

Even the Supreme Court has perpetuated the idea that one's demeanor is likely to reveal his lies. In *Coy v. Iowa*, a Confrontation Clause case in which the defendant was charged with sexual assault, the Court stated, "It is always more difficult to tell a lie about a person 'to his face' than 'behind his back.' In the former context, even if the lie is told, it will often be told less convincingly."[82]

---

[80] *See The (Real!) Science Behind Fox's Lie to Me*, POPULAR MECHANICS (Oct. 1, 2009, 12:00 AM), http://www.popularmechanics.com/science/4300722.

[81] *Lie to Me*, PAUL EKMAN GROUP, http://awstats.paulekman.com/drupal/?q=Lie-to-me (last visited Aug. 4, 2012) (summarizing the former television show's website).

[82] 487 U.S. 1012, 1019 (1988).

It would be wonderful if judges and jurors were able to determine the credibility of witnesses by carefully staring at the witnesses' shoulders and lower lips during testimony, but an abundance of research into laypeople's and trained individuals' ability to detect lies has shown that people are simply not very good at detecting lies by analyzing demeanor.[83] Additionally, training individuals to look for certain demeanor-based cues does not significantly improve accuracy; in fact, it increases misplaced confidence in one's own abilities to detect lies and leads to a bias toward suspecting others' untruthfulness.[84]

Early studies that sought to determine what types of behavioral cues were associated with deception typically utilized referees who would simply count the instances of a particular behavior when an individual (called the sender) was either lying or telling the truth.[85] For example, a listener would record each instance that a sender blinked while the sender was lying, and again while the sender was telling the truth. These studies found some relationships between certain types of demeanors and deception, such as eye movements, mouth movements, stiffness of postures, and relaxation.[86] However, when individuals in these early studies attempted to differentiate between equally true and false statements, they only achieved accuracy levels between 45% and 60%.[87]

These studies typically restrict the listener's interaction such that the only cues that she could use to detect deception are verbal and nonverbal demeanor cues, as opposed to cues gleaned from the content of the message (e.g., consistency, detail). For example, Ekman and Maureen O'Sullivan conducted an experiment in which listeners were shown ten one-minute samples from interviews of college-aged women describing their feelings about a nature film they were watching.[88] Half of these women were actually watching the nature film and truthfully describing their positive feelings, while the other half were watching a gruesome film that was upsetting to them and were lying about having positive feelings.[89] Groups of individuals observed these interviews to determine the veracity of the statements.[90]

---

[83] *See infra* Part II.C.

[84] *See infra* text accompanying notes 95–112.

[85] *See* Miron Zuckerman et al., *Verbal and Nonverbal Communication of Deception*, 14 ADVANCES EXPERIMENTAL SOC. PSYCHOL. 1, 11 (1981). Extensive review of these early demeanor studies is beyond the scope of this Comment. For a review of these studies, see Bella M. DePaulo et al., *Cues to Deception*, 129 PSYCHOL. BULL. 74 (2003), and, Zuckerman, *supra*.

[86] *See* Zuckerman, *supra* note 85, at 41–44 apps. 1A–1B.

[87] *Id.* at 26.

[88] Paul Ekman & Maureen O'Sullivan, *Who Can Catch a Liar?*, 46 AM. PSYCHOLOGIST 913, 915 (1991).

[89] *Id.*

[90] *Id.* at 914–15. The different groups included college students, psychiatrists, judges, robbery investigators, federal polygraphers, and Secret Service agents. *Id.*

Aside from Secret Service agents, every group was near chance accuracy, ranging from 52% to 57% accurate.[91] Additionally, prior to viewing the tape, each listener was asked to estimate his ability to tell when other people are lying, and then was asked to estimate how well he had done on assessing the interviews.[92] Neither estimate was significantly correlated with accuracy on the task, indicating that the listeners were unable to discern how effectively they determined credibility based on demeanor evidence.[93]

One interesting result of the Ekman and O'Sullivan study, however, was that Secret Service agents were considerably more accurate than the other groups, achieving 64% accuracy.[94] This leads to an important question: Could laypeople, such as jurors, be trained to better recognize deception based on demeanor evidence, thereby making the jury an efficient lie detector and alleviating the need for expert testimony regarding credibility?

A series of studies by Saul Kassin and colleagues involving mock crime interrogations sheds some light on this question. In a first study, eight college students were instructed to commit a mock crime, while eight others committed a related innocent act.[95] All of the students then underwent a five-minute interrogation, during which the students were instructed to deny having committed any crime.[96] The interrogations were taped and shown to forty students who were asked to assess the veracity of the senders' denials.[97] Prior to watching the interrogations, half of the forty students were trained for thirty minutes in the Reid Technique[98]—a method that uses verbal and nonverbal cues to detect deception.[99] Those students who were not trained in the Reid Technique had an accuracy rate of 52%, while those individuals that were trained achieved an accuracy of only 45%.[100] This lower accuracy rate was compounded by the fact that those individuals

---

[91] *Id.* at 916 & tbl.2.

[92] *Id.* at 915.

[93] *Id.* at 916–17.

[94] *Id.* at 916 tbl.2.

[95] Saul M. Kassin & Christina T. Fong, *"I'm Innocent!": Effects of Training on Judgments of Truth and Deception in the Interrogation Room*, 23 LAW & HUM. BEHAV. 499, 501–03 (1999). Subjects committed one of four possible crimes, such as vandalizing the wall of a university building. *Id.* Innocent acts involved going to the same location, but not committing any crime. *Id.*

[96] *Id.* at 503–04.

[97] *Id.* at 506.

[98] *Id.* at 505.

[99] The Reid Technique is the leading method of training law enforcement officials to conduct interrogations, detect deception, and elicit confessions. The method is outlined in a book co-authored by John Reid. *See* FRED E. INBAU ET AL., CRIMINAL INTERROGATION AND CONFESSIONS 187–90 (5th ed. 2013).

[100] Kassin & Fong, *supra* note 95, at 508. It should be noted that this accuracy is less than chance, indicating that the method is actually reducing lie-detection ability, rather than augmenting it.

trained in the Reid Technique had higher confidence than those not trained, despite their reduced accuracy.[101]

This phenomenon has been replicated in several follow-up studies examining highly trained police officers' abilities to assess the veracity of interview statements. In a 2002 study, Christian Meissner and Kassin presented highly trained and experienced police investigators with the same interrogation videos used in the 1999 Kassin and Christina Fong study described above.[102] Like the students of the Kassin and Fong study, the investigators were no better than chance at detecting deception in the interrogations and yet were more confident in their judgments than were the trained students.[103] Once again, investigator confidence was not correlated with accuracy.[104] Notably, both the trained officers and trained students were more likely to misdiagnose truthful senders as deceptive,[105] reflecting a bias in trained individuals towards assuming that senders will be deceptive.[106]

Although these are only a few of the many studies conducted regarding laypeople's ability to assess demeanor evidence, several authors have conducted meta-analyses[107] that combine data from a number of studies and look for literature-wide effects. The most comprehensive of these is a 2006 meta-analysis by Charles Bond and Bella DePaulo, which combined the results of 206 studies that included over 24,000 individuals "who judge[d] deception from a brief encounter with an unfamiliar sender in real time."[108] Across this extremely large sample gathered from a variety of different labs and researchers, the authors found a 54% overall accuracy of deception

---

[101] *Id.* at 509. Confidence was measured by asking each receiver to self-rate his confidence, on a scale from one to ten, after viewing each interrogation. *Id.*

[102] Christian A. Meissner & Saul M. Kassin, *"He's Guilty!": Investigator Bias in Judgments of Truth and Deception*, 26 LAW & HUM. BEHAV. 469, 474 (2002). The investigators had an average of nearly fourteen years of experience and 68% of the investigators had undergone formal training in the detection of deception. *Id.*

[103] *Id.* at 475 tbl.2, 476.

[104] *Id.* at 476.

[105] *Id.* at 475 tbl.2.

[106] Meissner and Kassin found this "more liberal response criterion" (i.e., a tendency to assume that others are lying more frequently than telling the truth) in a number of other studies of individuals trained in using demeanor evidence to detect deception. *Id.* at 472–73 & tbl.1 (citing Paul Ekman et al., *A Few Can Catch a Liar*, 10 PSYCHOL. SCI. 263 (1999); Kassin & Fong, *supra* note 95; Stephen Porter et al., *Truth, Lies, and Videotape: An Investigation of the Ability of Federal Parole Officers to Detect Deception*, 24 LAW & HUM. BEHAV. 643 (2000)).

[107] A meta-analysis is "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." Gene V. Glass, *Primary, Secondary, and Meta-Analysis of Research*, 5 EDUC. RESEARCHER 3, 3 (1976). Typically, a meta-analysis uses statistical methods to find the size of a particular effect that has been measured by an entire field of research. *See id.*

[108] Charles F. Bond, Jr. & Bella M. DePaulo, *Accuracy of Deception Judgments*, 10 PERSONALITY & SOC. PSYCHOL. REV. 214, 216–17, 219 (2006).

judgments, much like the accuracy rates found in the Kassin and Ekman studies above.[109] This meta-analysis excluded studies that involved training, so it does not speak to the confidence and bias effects mentioned above,[110] but a number of other studies have found effects of training similar to those found by Meissner and Kassin.[111]

These studies raise serious concerns about jurors' ability to assess credibility based on demeanor evidence. If jurors are untrained, they are likely to be only slightly above chance accuracy. Moreover, brief training like that used in the 1999 Kassin and Fong study is unlikely to make jurors more accurate assessors of credibility. Perhaps extensive training and experience of the kind provided to Secret Service agents could help increase jurors' abilities slightly, but this is probably unrealistic for jurors who have only a short period of time to receive instructions from the court.[112]

Additionally, even if such training was sufficiently cost and time efficient to be feasible and was able to produce small increases in accuracy like those found by Ekman, the increased confidence associated with the training could still be highly problematic in the courtroom. When a juror hears testimony, the credibility of the witness is only one of the factors he must consider. He must also consider things such as the quality of the evidence, the weight to be given to that evidence, and the influence of the evidence on previous information. If a juror is highly confident in his ability to assess the veracity of the witness's statements, that juror may completely disregard evidence when he suspects a witness is lying, even when the witness provides reliable evidence that carries strong weight. Thus, confidence in credibility assessment is a detriment to accuracy when those assessments themselves are inaccurate.

Likewise, the bias that Meissner and Kassin found toward assuming others are deceptive as a result of the deception training could further undermine accuracy in the courtroom. In controlled experiments where there is an equal balance between truthful and deceptive statements, this bias may only have a small effect on overall accuracy, depending on the strength of the bias. But in the courtroom, where it is likely that the majority of witnesses are telling the truth, a bias toward assuming witnesses are deceptive could reduce jurors' assessments of credibility to well below chance levels.

---

[109] *Id.* at 219.

[110] *Id.* at 218.

[111] *See supra* note 106.

[112] One criticism of training paradigms like the one used by Kassin and Fong is that they are so short that they are virtually guaranteed to be ineffective. However, in the trial context, the thirty minute training session is likely very close to the maximum of what could be employed by the court, given the demands placed on a jury. Even if time constraints were not a factor, the cost of serious training would likely be prohibitive.

C. Credibility Assessment Based on Content

Though the behavioral research discussed above has not gained traction in the courts, a number of legal scholars have written on the topic.[113] The majority of these discussions have focused on research examining credibility assessment through demeanor evidence alone, and have thus concluded that jurors are likely ineffective credibility assessors.[114] Recently, however, Max Minzner noted that the literature focusing on demeanor evidence alone does not fully describe the jury's ability to assess credibility.[115] In addition to demeanor, the jury may use the context, consistency, and depth of witnesses' statements to determine their veracity.

Thus, the experiments by Kassin, Ekman, and their ilk may capture only a small amount of what juries actually do in the courtroom to assess credibility. By allowing individuals to assess only a short segment of testimony, these experiments prevent receivers from comparing the sender's story with others' stories, as a juror would be able to do while listening to multiple witnesses at trial.

Two distinct questions arise from the shortcomings of the demeanor studies. First, do jurors use context, consistency, and depth to assess credibility in the courtroom? Second, if they do use the full amount of information available at trial, are jurors accurate in making credibility determinations? Minzner cites two studies in support of the position that jurors do rely on context.[116]

In the first study, Hee Sun Park and her collaborators conducted a survey of 202 undergraduate students in which they asked participants "to recall a recent situation in which they had discovered that someone had lied to them . . . and recall as much information as they could about what happened."[117] After this free recall period, participants were asked several questions about the instance they recalled, including one question about how they detected the lie.[118] The survey found that 32% of the lies were detected through third-party information, 31% were detected through a

---

[113] *See, e.g.*, Jeremy A. Blumenthal, *A Wipe of the Hands, a Lick of the Lips: The Validity of Demeanor Evidence in Assessing Witness Credibility*, 72 NEB. L. REV. 1157 (1993); Joseph W. Rand, *The Demeanor Gap: Race, Lie Detection, and the Jury*, 33 CONN. L. REV. 1 (2000); Chris William Sanchirico, "*What Makes the Engine Go?" Cognitive Limitations and Cross-Examination*, 14 WIDENER L. REV. 507 (2009).

[114] *See supra* Part II.B.

[115] *See* Max Minzner, *Detecting Lies Using Demeanor, Bias, and Context*, 29 CARDOZO L. REV. 2557, 2564 (2008) ("Today's findings . . . are far more complex than the then-current research and give us a much greater ability to identify those situations in which we need to worry about mistaken judgments about credibility and those in which we do not.").

[116] *Id.* at 2567–68.

[117] Hee Sun Park et al., *How People Really Detect Lies*, 69 COMM. MONOGRAPHS 144, 149 (2002).

[118] *Id.* at 149–50. The question read, "Now, think about how you found out that the person lied to you. Describe in as much detail as you can the events surrounding your discovery of the lie: how exactly did you find out that the person lied to you?" *Id.* at 150.

combination of various methods, 18% were detected through physical evidence, and only 2% were detected with demeanor-based evidence.[119]

While these results are interesting, they are flawed for a number of reasons. First, because the participants were invited to freely recall a single instance in which they detected a lie, they might be biased toward recalling certain instances more than others. For example, if it is easier to detect a lie by discovering third-party information that exposes the lie than it is to detect the lie by demeanor, participants may be more likely to recall those lies detected via third-party information, skewing the results.[120] Similarly, there might be categorical differences in the extent to which people recall detecting lies based on information compared to detecting lies based on demeanor, regardless of which method is more common.

But even assuming that these individuals are correctly recalling what methods they used and that they selected the lie they described without any bias, the data still do not indicate how jurors might assess credibility. The situations that the participants described were likely vastly different from what a juror would hear in court. The majority of lies in the Park study were detected through third-party information or physical evidence,[121] both of which are unlikely to happen in the courtroom. The authors do not define what they mean by "third party information," but one can reasonably assume this means that a trusted source contradicted the liar in a way that could be confirmed. While there is contradiction aplenty in the courtroom, a witness would rarely make such a blatant lie that some piece of evidence provided at trial would clearly expose that lie, because witnesses know what evidence is going to be presented. This situation is distinct from lies outside of the courtroom, where the liar might assume that the receiver will not discover any information that would clearly expose the lie. It is precisely because lies cannot be easily detected in the courtroom via other witnesses or physical evidence that jurors might turn to demeanor-based cues.

Though the Park study tells us little about what actual jurors are doing to assess credibility in the courtroom, the second study cited by Minzner does support the notion that jurors use context in their credibility determinations. In a large and ongoing line of research, Shari Seidman Diamond examined questions asked by jurors in fifty civil cases in Arizona state courts, where jurors are allowed to present questions to witnesses in civil cases.[122] Diamond and colleagues found that 42% of all juror questions

---

[119] *Id.*

[120] This phenomenon is referred to in the psychological literature as the availability heuristic. For a more complete description, see Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCHOL. 207 (1973).

[121] Park, *supra* note 117, at 150.

[122] Shari Seidman Diamond et al., *Juror Questions During Trial: A Window into Juror Thinking*, 59 VAND. L. REV. 1927, 1937 (2006). Jurors are allowed to present questions in Arizona state civil

were what the authors labeled as "cross-checking," questions in which "jurors apply a commonsense structure to evaluating potentially unreliable sources of information."[123] Essentially, these questions reflect the jurors' attempt to find consistency across a large number of potentially inconsistent stories that they have heard from all of the witnesses. This information was often sought from a disinterested witness who might be able to provide a reliable fact against which the testimony from the interested witnesses could be compared.[124]

While this study is informative because it shows that jurors use coherence of stories to assess credibility, jurors' questions to the witnesses are unlikely to capture all of the considerations that comprise their determinations. For example, it seems highly unlikely that a juror would ask a witness a question about demeanor (e.g., "Why did you appear so nervous on the stand?"). Thus, the study does not provide a comparison point to know how much credibility assessments are influenced by demeanor as opposed to content-based cross-checking.

One other limitation of the Diamond study is that forty-seven of the fifty cases sampled were tort cases, most of which involved either a motor vehicle crash or medical malpractice.[125] It is possible that the jurors' credibility assessment varies by context. Perhaps in those cases where liability turns on the veracity of two contradictory testimonies, such as those of a single eyewitness and a criminal defendant, there will be reduced context from which jurors can draw to make content-based decisions of credibility. In such cases, jurors may rely more on demeanor in making their credibility assessments.

Despite these limitations, it seems very likely that jurors use context to determine the consistency of a witness's testimony, though it is not yet clear how this compares to the use of demeanor. Regardless, this leads to the next question in the analysis: If jurors do use context to determine credibility, does this lead to more accurate assessments of credibility than the use of demeanor evidence? Minzner argues that by placing all of the information heard at trial into context and examining the differences between the statements of various witnesses, the jury can accurately assess credibility without needing to consider demeanor evidence.[126] Minzner discusses two studies by Maria Hartwig to support this view.[127]

In the first study, Hartwig used the same crime and interrogation paradigm of the 1999 Kassin and Fong study described above. College

---

cases. The rule provides that "[j]urors shall be permitted to submit to the court written questions directed to witnesses or to the court." ARIZ. R. CIV. P. 39(b)(10) (2011).

[123] Diamond et al., *supra* note 122, at 1956–57.

[124] *Id.*

[125] *Id.* at 1937.

[126] *See* Minzner, *supra* note 115, at 2567–68.

[127] *Id.* at 2568–69.

students were instructed to either commit a mock crime by stealing a wallet out of a briefcase in a movie store, or an innocent act by going into the video store and leaving.[128] Then, subjects were interrogated as in the 1999 Kassin and Fong study, with all students denying commission of the crime.[129] However, in this experiment, the time at which the interrogator disclosed the evidence against the student was varied. In an "early disclosure" condition, the interrogator presented three items of evidence against the student at the beginning of the interrogation and then provided the student a "free recall" period to explain what happened.[130] In a complementary "late disclosure" condition, the interrogator withheld the three pieces of evidence from the student, first providing a free recall period and then presenting the evidence against the student at the end of the interrogation.[131] After the interviews, 116 students viewed the tape of one of the interrogations and judged the veracity of the statements made by the suspects.[132] Overall accuracy was 53%, but those observing the early disclosure interrogations were only 43% accurate while the late disclosure group was 62% accurate.[133]

Minzner argues that the late disclosure condition is a better simulation of the jury's experience, and thus, juries are more likely to accurately assess credibility than the previous demeanor-based experiments implied.[134] However, there are still problems in applying the Hartwig study to the courtroom. First, while late disclosure strategy could readily be employed in interrogation rooms, witnesses in the courtroom likely have a solid understanding of what questions will be asked and what evidence will be relevant to those questions. While the interrogators in the Hartwig study could easily withhold some evidence, get a story from the suspect, and then present the evidence to the suspect to probe for inconsistencies, attorneys attacking credibility on cross-examination will rarely have this opportunity.

In civil cases, where discovery provides all parties with the relevant evidence, a lying witness has plenty of time and information to craft a consistent story. Even in criminal cases, suspects may know, from their

---

[128] Maria Hartwig et al., *Detecting Deception via Strategic Disclosure of Evidence*, 29 LAW & HUM. BEHAV. 469, 473–74 (2005).

[129] *Id.* at 473.

[130] *Id.* at 475. The three pieces of evidence claimed against the student were that one witness had seen the student outside the video store, that one witness had seen the student inside the video store, and that the student's fingerprints were contained on the briefcase from which the wallet was stolen. *Id.*

[131] *Id.*

[132] *Id.* at 476.

[133] *Id.* at 477.

[134] *See* Minzner, *supra* note 115, at 2571 ("What do these results about demeanor, bias, and context indicate about lie detection in the legal system? . . . When a legal decision-maker possesses private information and uses it to probe the witness's story, lie detection accuracy will improve. . . . If the police and juries have multiple independent sources of information about an event, they are far more likely to accurately decide which witnesses to believe.").

interactions with the police, enough of the evidence against them such that they can form a coherent story.[135] Perhaps jurors could watch a video of the interrogation and assess credibility based on that video. But this would require a shift in police policy, since evidence is quickly disclosed to the suspect in the majority of interrogations.[136] This method would also only be effective in criminal cases where there is an interrogation to be shown to the jurors, and overconfidence of interrogators may influence the interrogation and bias the jury. Additionally, while the late disclosure method increased accuracy among passive viewers in credibility assessment, it still only resulted in 62% accuracy. Is this level of accuracy sufficient to declare that the jury is so good at detecting lies that it should be the sole credibility assessor?

It is also notable that the observers in the late disclosure condition were not equally adept at classifying both truth telling and lying: 68% of the deceptive interrogations were correctly classified compared to only 54% of the truthful interrogations.[137] This has implications for the normative distribution of outcomes in the trial system. For example, employing the late disclosure method would presumably reduce the perceived credibility of a defendant who chooses to testify, since the data suggest that more liars would be correctly classified than truth tellers. This could increase the number of verdicts rendered against criminal defendants—a high price to pay for a small increase in the accuracy of credibility judgments.

In a follow-up to their 2005 study, Hartwig and colleagues trained police officers in employing the late disclosure interrogation method.[138] Mock crimes and interrogations were conducted as in the 2005 Hartwig study,[139] but in this experiment, the interrogators themselves judged the veracity of the suspects' statements.[140] The interviewers trained to use the late disclosure strategy reached an accuracy level of 85%, while the

---

[135] Such knowledge could arise from several different sources. Police may reveal evidence to the suspect during interrogation. *See* Hartwig, *supra* note 128, at 470. Additionally, certain types of evidence must be revealed to the defendant prior to trial, such as documents that the prosecution plans to use in its case-in-chief or the testimony of expert witnesses, which will typically include forensic experts. *See* FED. R. CRIM. PROC. 16(a)(1)(E)–(F).

[136] Hartwig, *supra* note 128, at 470 ("In an American study, this disclosure of evidence, often together with a suggestion of guilt, was the typical way to start the interrogation, and occurred in more than 80% of the cases." (citing Richard A. Leo, *Inside the Interrogation Room*, 86 J. CRIM. L. & CRIMINOLOGY 266 (1996))).

[137] *Id.* at 478.

[138] Maria Hartwig et al., *Strategic Use of Evidence During Police Interviews: When Training to Detect Deception Works*, 30 LAW & HUM. BEHAV. 603, 608 (2006) ("[The officers] were trained in planning and asking questions concerning the evidence without disclosing it to the suspect. For example, if the case-file included information that a suspect's car had been seen close to a crime scene on the day of the crime, they were taught to plan and ask questions about the suspect's car . . . . They then practiced the strategic use of [this] technique on each other several times . . . .").

[139] *Id.* at 608–10.

[140] *Id.* at 610.

interviewers using the early disclosure strategy were accurate only 56% of the time.[141]

While these results are more striking than those of the 2005 Hartwig study, they are also much less applicable to jurors. At trial, the jury is simply the passive observer of evidence, not the interrogator. Juries cannot employ the late disclosure method themselves as in the 2006 Hartwig experiment. While the interrogators using the late disclosure strategy were very accurate in their own credibility assessments, it is not clear whether passive observers would be as accurate, and in fact they were not as accurate in the 2005 Hartwig study. Additionally, the 2006 Hartwig study used a three-hour training protocol,[142] which might be more training than the judicial system is willing to devote to teach jurors about credibility assessment, especially when that training may not even be effective.

D. Should the Jury Be the Lie Detector?

It is clear that there is a gap in the literature with regard to the jury's ability to assess credibility. While studies confining the assessment of credibility to demeanor-based evidence have shown that individuals have poor accuracy, studies in which individuals are able to use context reflect improved accuracy. However, none of these studies are directly applicable to the courtroom setting. There has not been a systematic study in which individuals hear inconsistent testimony from multiple senders and then assess the credibility of each sender, as jurors would in the courtroom. When more sophisticated research like this is conducted, researchers will have a better grasp of the extent to which jurors can assess the credibility of witnesses without the help of experts.

For now, however, the legal community is left with an incomplete understanding. Even the most optimistic studies that are remotely applicable to the courtroom suggest that individuals are just over 60% accurate in credibility assessments.[143] Yet courts continue to hold onto the shaky assumption that the jury is capable of being the sole assessor of credibility. This unsophisticated notion should be put to rest, as Justice Kennedy suggested in *United States v. Scheffer*,[144] and the only factor concerning the admissibility of expert testimony related to credibility should be its reliability under *Daubert* or *Frye*.

---

[141] *Id.* at 613.

[142] *Id.* at 608.

[143] *See* Hartwig et al., *supra* note 128, at 477.

[144] 523 U.S. 303, 318–19 (1998) (Kennedy, J., concurring in part and concurring in the judgment) ("[I]t seems the principal opinion overreaches when it rests its holding on the additional ground that the jury's role in making credibility determinations is diminished when it hears polygraph evidence. . . . [It] demeans and mistakes the role and competence of jurors . . . .").

However, the court does not lightly change its doctrine in response to social science evidence,[145] so courts will likely continue to be skeptical of neuroscience-based credibility-assessment evidence based on the notion that it "impinges on the province of the jury."[146] What does this mean for the admissibility of both the CQT and the CIT? If the jury is considered the sole assessor of credibility, it almost surely strikes the death knell for the CQT, which purports to directly identify truths and lies. As seen in *Wilson v. Corestaff Services L.P.*, courts following the standard laid out in *Scheffer* will likely continue to reject a CQT test, whether it is conducted using polygraph or fMRI, without even having to consider its reliability.

The CIT, however, may not violate the *Scheffer* standard because it is much more like substantive evidence than it is like credibility assessment. While the results of a CIT may undermine the credibility of a witness indirectly, the CIT (unlike the CQT) makes no direct claims as to whether a witness is lying or telling the truth. Instead, the CIT provides substantive evidence and leaves the credibility assessment itself to the jury. For example, if an individual defending a murder charge maintains he was not at the crime scene, but the results of a CIT indicate that he recognized the murder weapon, the CIT itself does not indicate that he is lying. Instead, it provides substantive evidence that is probative for the trier of fact's ultimate conclusion about the veracity of his testimony. Similarly, a blood sample found at the scene of the crime that contains the defendant's DNA strongly undermines the defendant's claim of innocence, but we would not consider it credibility-assessment evidence. If substantive evidence like this were to be considered credibility-assessment evidence, every piece of relevant evidence, by definition, would go to the veracity of the defendant's plea of "not guilty" or "not liable" and would therefore be inadmissible.

Despite the fact that logical analysis would group the CIT with other substantive evidence, the CIT is in danger of being grouped with the CQT as a pure credibility-assessment device for several reasons. First, because the CIT was first conducted using the polygraph,[147] it is likely to continue to be associated with the CQT, which is a credibility-assessment test at its very core. Courts have not overtly made the distinction between the CIT and the CQT, frequently just referring to both as "polygraph" evidence,[148] but not referencing each test's merits and flaws.[149]

---

[145] *See, e.g.*, McClesky v. Kemp, 481 U.S. 279, 287, 291 n.7 (1987) (accepting a social science study as demonstrating that "black defendants . . . who kill white victims have the greatest likelihood of receiving the death penalty" but reading the results narrowly and as insufficient to support a constitutional challenge).

[146] Wilson v. Corestaff Servs. L.P., 900 N.Y.S.2d 639, 642 (Sup. Ct. 2010).

[147] *See supra* note 45 and accompanying text.

[148] *See* Erikson, *supra* note 61, at 785–86.

[149] *See, e.g.*, *Scheffer*, 523 U.S. at 309 ("[T]here is simply no consensus that *polygraph* evidence is reliable." (emphasis added)).

Second, even though the CIT does not detect any actual lying, because the CIT only provides evidence of a mental state, courts could still draw a distinction between the CIT and other forms of substantive forensic evidence, such as fingerprints or DNA samples. The court may consider any test of brain activity to be one of credibility assessment in nature, regardless of whether the test is identifying lies or simply the presence of concealed information. Third, because the history of the CIT is so intertwined with that of the CQT, the association between the tests will be difficult to countermand. The CIT was born out of frustrations with the theoretical and practical shortfalls of the CQT.[150] The CIT literature is rife with debate over which method is more accurate and reliable, and the community could be described as being split into two rival factions: CIT proponents and CQT proponents.[151] Many CIT articles are prefaced by the shortcomings of the CQT, espousing the CIT as a valid alternative.[152] This type of language implies that the two tests are performing the same credibility-assessing job, despite the fact that they actually have very different functions. And many published CIT articles are titled in ways that imply the CIT is a deception-detection tool, rather than a recognition-detection tool.[153]

To summarize, we have seen that under *Scheffer*, scientific expert testimony whose relevance is the credibility of one of the witnesses at trial will be excluded on the basis that the jury itself should assess the credibility of the witnesses. However, we have also seen that laypeople are usually at or below chance accuracy in assessing whether individuals are lying based purely on demeanor. Training is unlikely to help jurors become better credibility assessors and may actually make them worse by causing them to be overconfident in their credibility assessments to the detriment of assessments they may be more effective at making, such as the determination of the quality of the evidence presented. Jurors likely assess credibility based on the content of the message in addition to demeanor, but there is no strong evidence suggesting that jurors are effective at detecting

---

[150] *See* Iacono, *supra* note 24, at 689 ("The [CIT] . . . has been developed and promoted by Lykken as a scientifically based alternative to the CQT."); Lykken, *supra* note 45, at 385 ("Use of physiological measurements to detect not lying, but the presence of 'guilty knowledge,' requires . . . [a] more reasonable assumption [than those required by the CQT]."). Notably, the very first paragraph of the very first CIT study ever published focused on the shortcomings of more traditional "lie detector" tests. Lykken, *supra* note 45, at 385.

[151] *Compare* Iacono & Lykken, *supra* note 20, § 40:45–119 (writing a chapter titled "The Case Against Polygraph Tests," describing shortcomings of the CQT and related tests and instead favoring the CIT), *with* Honts et al., *supra* note 33, § 40:20–44 (writing a countering chapter entitled "The Case for Polygraph Tests," arguing in favor of the CQT and related tests' validity).

[152] *See, e.g.*, Ralf Mertens & John J.B. Allen, *The Role of Psychophysiology in Forensic Assessments: Deception Detection, ERPs, and Virtual Reality Mock Crime Scenarios*, 45 PSYCHOPHYSIOLOGY 286, 286 (2008) ("Conventional field polygraph examinations, based on the control-question technique (CQT), suffer from many limitations and have been widely criticized in the scientific literature.").

[153] *See, e.g.*, Farwell & Donchin, *supra* note 48; Mertens & Allen, *supra* note 152.

lies through the content of the testimony, as the research done in this area is not a good analogue for the experience of the juror at trial. Thus, if accuracy is an important motivator, the current social science data counsel against allowing the jury to function as the sole assessor of credibility.

III. THE *DAUBERT* CRITERIA AS A HURDLE TO ADMISSIBILITY

If the court does not revise the "jury [as] the lie detector" doctrine explained in *Scheffer*,[154] the court will not reach the second stage of analysis: whether the CQT and CIT are reliable enough to meet the standard laid out in Federal Rule of Evidence 702. However, as discussed in Part II, social science research may eventually push the court to consider allowing experts to assist the trier of fact in the credibility-assessment role. Even if courts do not revise their stance on experts offering credibility-assessment evidence but are willing to accept the CIT as substantive evidence rather than credibility-assessment evidence, the court will need to assess its reliability in determining admissibility. This Part examines the potential admissibility of the CIT and, to a lesser extent, the CQT, under the Federal Rules of Evidence.

A. Frye v. United States*, Federal Rule of Evidence 702, and the* Daubert *Standard*

For nearly seventy years, admissibility of scientific expert testimony was determined under the standard laid out in *Frye v. United States*, a case that itself actually decided admissibility of a type of credibility-assessment evidence.[155] In that case, the D.C. Circuit outlined a "general acceptance" test under which any scientific evidence that is to be admissible "must be sufficiently established to have gained general acceptance in the particular field in which it belongs."[156] While this was not a Supreme Court decision, federal courts generally followed this until the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[157]

In *Daubert*, the Supreme Court noted that the *Frye* standard had been the subject of frequent debate.[158] Interpreting Federal Rule of Evidence 702, the Court stated that the Rule does not indicate that "general acceptance" is

---

[154] *Scheffer*, 523 U.S. at 313 (quoting United States v. Barnard, 490 F.2d 907, 912 (9th Cir. 1973), *cert. denied*, 416 U.S. 959 (1974)).

[155] 293 F. 1013, 1013–14 (D.C. Cir. 1923) (ruling that the "systolic blood pressure deception test," a precursor to the polygraph-based CQT, had not gained "general acceptance" among physiological and psychological authorities and therefore should not be admitted as evidence).

[156] *Id.* at 1014.

[157] 509 U.S. 579 (1993).

[158] *Id.* at 586.

a requirement of reliability.[159] In fact, the advisory committee notes to the originally enacted version of Rule 702 never mention "general acceptance," and instead state that "[w]hether the situation is a proper one for the use of expert testimony is to be determined on the basis of assisting the trier."[160] The advisory committee notes to the proposed 1991 amendment of Rule 702 explicitly state that "[t]he rule does not mandate a return to the strictures of *Frye v. United States*."[161] The *Daubert* Court noted that a strict requirement of "general acceptance" limits the "liberal thrust" of the Federal Rules, which, under Rule 401, envision the acceptance of evidence that is relevant at virtually any level.[162] Thus, *Daubert* explicitly stated that the *Frye* test should not be used in federal courts.[163]

In replacing *Frye*, the *Daubert* Court provided four nonexclusive factors for the judge to consider when determining whether scientific evidence is sufficiently reliable:[164]

- "[W]hether [the theory or technique] can be (and has been) tested,"[165]
- "[W]hether the theory or technique has been subjected to peer review and publication,"[166]
- "[T]he known or potential rate of error,"[167] and
- The "general acceptance" of the technique.[168]

---

[159] *Id.* at 588 ("Nothing in the text of this Rule establishes 'general acceptance' as an absolute prerequisite to its admissibility . . . [nor] present[s] any clear indication that Rule 702 or the Rules as a whole were intended to incorporate a 'general acceptance' standard.").

[160] FED. R. EVID. 702 committee note (1975).

[161] Preliminary Draft of Proposed Amendments to the Federal Rules of Civil Procedure and the Federal Rules of Evidence, 137 F.R.D. 53, 157 (1991) (describing the proposed amendment to FED. R. EVID. 702 and the accompanying committee notes).

[162] *Daubert*, 509 U.S. at 587–88. Rule 401 is a highly lenient rule of relevancy, requiring only that evidence have "any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." FED. R. EVID. 401. This rule has been repeatedly determined to be a liberal standard. *See* JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S EVIDENCE MANUAL § 6.01[5][a] (8th ed. 2007).

[163] *Daubert*, 509 U.S. at 589 ("That austere standard, absent from, and incompatible with, the Federal Rules of Evidence, should not be applied in federal trials.").

[164] *Id.* at 593–94.

[165] *Id.* at 593.

[166] *Id.* This criterion was specifically mentioned by the Court as being "relevant, though not dispositive" because "some propositions . . . are . . . too new, or of too limited interest to be published." *Id.* at 593–94.

[167] *Id.* at 594. The court did not give any guidelines as to what would constitute an acceptable rate of error.

[168] *Id.* This factor is the very same test from *Frye* that the Court declared was no longer controlling. The Court explained the factor by stating that the *Frye* test was unnecessarily strict because a "reliability assessment does not require . . . an express determination of a particular degree of acceptance within the community." *Id.* (quoting United States v. Downing, 753 F.2d 1224, 1238 (3d Cir. 1985)). But this

In 2000, Rule 702 was amended to codify the *Daubert* factors.[169] The revised rule provides that an expert may testify as to an opinion only "if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case."[170]

Typically, when proffered scientific testimony is contested by the opposing party, the trial court will hold an evidentiary hearing, called a *Daubert* hearing, to determine the reliability of the testimony.[171] The remainder of this Part examines the likely outcome of such a hearing as applied to the P300-based CIT,[172] and to a lesser extent, the fMRI-based CQT.

B. The Daubert *Standard Applied to the fMRI-Based CQT*

In the wake of the sudden increase in fMRI-based lie-detection research over the past decade,[173] a number of scholars anticipated the question that would eventually face the court in *United States v. Semrau*: Is the fMRI-based CQT admissible under the *Daubert* standard?[174] Because

---

factor can have a bearing in the eventual calculus because a known technique that has not been accepted in its field should be viewed with skepticism. *See id.*

[169] WEINSTEIN & BERGER, *supra* note 162, § 13.02[4][a]. The rule also codifies the Supreme Court's holding in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), that *Daubert*'s principles apply to all expert testimony admissible under Rule 702. FED. R. EVID. 702 committee note (2000). While the *Daubert* standard has been codified into the Federal Rules of Evidence and is therefore used in all federal courts, some state courts that have not adopted the rules still use the *Frye* test in determinations of the admissibility of scientific evidence. *See, e.g.*, Grady v. Frito-Lay, Inc., 839 A.2d 1038, 1045 (Pa. 2003) (requiring general acceptance of an expert's methods).

[170] FED. R. EVID. 702.

[171] WEINSTEIN & BERGER, *supra* note 162, § 13.02[4][c][ii].

[172] As of the writing of this Comment, no federal court has held a formal *Daubert* hearing regarding the admissibility of the CIT.

[173] Prior to 2000, there had been no published research using fMRI to detect deception. However, from 2000 to 2008, there were fifteen papers published on the subject. *See* Kamila E. Sip et al., *Detecting Deception: The Scope and Limits*, 12 TRENDS COGNITIVE SCI. 48, 51 tbl.1 (2008).

[174] Many articles focusing on the admissibility of fMRI-based lie detection are pessimistic as to potential admissibility, but some offer countering views. *Compare* Archie Alexander, *Functional Magnetic Resonance Imaging Lie Detection: Is a "Brainstorm" Heading Toward the "Gatekeeper"?*, 7 HOUS. J. HEALTH L. & POL'Y 1, 49–55 (2007) (arguing that fMRI-based lie detection is not generally accepted within the field and that the "analytical gap" from experimental studies to practical application is too great), Joëlle Anne Moreno, *The Future of Neuroimaged Lie Detection and the Law*, 42 AKRON L. REV. 717, 734–36 (2009) (describing the potential uses and limits of cognitive neuroscience in the courtroom), Moriarty, *supra* note 31, at 758–61 (2009) (arguing that the number of fMRI lie-detection studies is too small to be considered reliable, that there has been too little replication of past research, and that these studies are not close enough to real-life situations to be applicable), *and* Cooper Ellenberg, Note, *Lie Detection: A Changing of the Guard in the Quest for Truth in Court?*, 33 LAW & PSYCHOL. REV. 139, 147 (2009) (stating that fMRI-based lie detection does not currently have the necessary accuracy rates to reach admissibility but that there is potential for future admissibility), *with* Leo Kittay, Note, *Admissibility of fMRI Lie Detection: The Cultural Bias Against "Mind Reading" Devices*, 72 BROOK. L. REV. 1351, 1376–79 (2007) (stating that the general widespread acceptance of

this question has been well discussed in the legal literature, this Comment will not examine each of the *Daubert* factors as applied to the commercially available fMRI-based CQT like the one that was offered in *Semrau*, but instead will examine general acceptance and rate of error, since these factors pose the greatest barriers to admissibility.[175]

Perhaps the most serious problem for proponents of fMRI-based CQT evidence is the lack of general acceptance in the field. This factor warrants extra consideration in the analysis because it not only influences federal courts and other jurisdictions that have adopted the *Daubert* standard, but also those states that continue to use the *Frye* test. It would be very difficult to argue that fMRI-based lie detection is "generally accepted" in the field of neuroscience. At least one neurologist has called for statutory regulation of the method, stating that the high level of discretion by the trial judge under the *Daubert* standard could lead to unwarranted admission of an fMRI-based CQT.[176] Other notable scholars in the field have taken a less extreme position, but have firmly stated that the fMRI-based CQT is not ready for courts.[177]

Some scholars have argued that the mere fact that fMRI itself is generally accepted in the field of neuroscience, a proposition that is unquestionably true,[178] means that fMRI-based lie detection should also be considered generally accepted.[179] This view misconstrues the meaning of the general acceptance test, which analyzes the general acceptance of the specific method being applied, not the underlying framework.[180] For example, general acceptance in the field of biology that humans are made up of DNA that is unique on an individual level would not allow for the admissibility of forensic DNA evidence until the process of matching an

---

fMRI research meets the general acceptance test and that the error rate for fMRI-based lie detection is lower than that of other forensic evidence that is regularly admitted).

[175] I should also note that some of the studies described here do not follow the traditional CQT format, but are instead hybrids between the CQT and other types of lie-detection protocols. While a more nuanced approach would be necessary for a complete assessment of the admissibility of these tests, the critical distinction here is that these fMRI-based tests all seek to determine whether a person is being truthful, as the CQT does, instead of determining whether a person recognizes information related to a crime, as the CIT does. Thus, I loosely use the term CQT to refer to these fMRI-based CQT-like tests.

[176] Henry T. Greely & Judy Illes, *Neuroscience-Based Lie Detection: The Urgent Need for Regulation*, 33 AM. J.L. & MED. 377, 413 (2007) ("The federal government—or, barring that, state governments—should ban any non-research use of new methods of lie detection, including specifically fMRI-based lie detection, unless or until the method has been proven safe and effective to the satisfaction of a regulatory agency and has been vetted through the peer-reviewed scientific literature.").

[177] *See* sources cited *supra* note 174.

[178] A Google Scholar search for "fMRI" yields thousands of citations in scholarly journals, and fMRI is regularly used in neuroscience research.

[179] *See, e.g.*, Kittay, *supra* note 174, at 1383.

[180] For example, *Frye* held that a systolic blood pressure test to measure deception was not generally accepted despite the fact that the measurement of blood pressure was undoubtedly widely accepted in medicine at the time. *See* Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

individual with a DNA sample from a crime scene has been generally accepted in the field. While scholars agree that fMRI is a valid method, there is not broad agreement that its use for lie detection is valid.

The fMRI-based CQT may also have difficulty meeting the accuracy requirement of *Daubert*. Across the small number of published fMRI-based CQT studies, accuracy rates have varied between 78% and 90%.[181] While it is unclear exactly what error rates are necessary to survive the *Daubert* analysis,[182] Leo Kittay has argued that because some types of forensic evidence that are commonly admitted, such as fingerprint analysis, have lower accuracy rates, fMRI-based lie detection should pass the error rate test.[183] This argument, however, assumes that these types of forensic evidence are being admitted due to their reliability rather than their strong tradition of admissibility, which may not be true.[184] Just as tradition may keep unreliable forensic sciences in the court, the past history of the inadmissibility of the CQT conducted using the polygraph may make it difficult for the fMRI-based CQT to be admitted. Even if the *Daubert* criteria are met, because courts have rejected lie-detection evidence for years, they are likely to remain skeptical of sudden increases in published accuracy rates because of a change in the machine used to conduct the test.[185] Of course, such reasoning is not grounded in the reliability-seeking considerations of *Daubert*, but it is a factor that will likely influence admissibility decisions.

While many of the fMRI-based CQT reliability issues are theoretically correctable through further experimentation and refinement, it seems likely that the field will take time to address these problems. Combined with the

---

[181] *See* Kozel et al., *supra* note 35, at 610 (describing a 90% accuracy rate); Langleben et al., *supra* note 35, at 269 (reporting a 78% accuracy rate). Additionally, these studies have only occurred in lab settings and have never tested real-world deception.

[182] The *Daubert* opinion only mentions that courts "should consider the known or potential rate of error." Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 594 (1993). The case is silent, however, as to what rate of error is sufficient to meet the test.

[183] *See* Kittay, *supra* note 174, at 1382. A number of types of commonly admitted forensic evidence have recently come under scrutiny for being inaccurate. A 2009 report from the National Research Council called for a major overhaul of the forensic sciences, stating that: "In a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions . . . ." NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 53 (2009).

[184] *Cf.* United States v. Llera Plaza, 188 F. Supp. 2d 549, 571–72 (E.D. Pa. 2002) (recognizing the potential for inaccurate fingerprint evidence, but still allowing its use by an expert at trial).

[185] One related problem may be that the probative accuracy may not be enough to overcome the prejudicial influence of the test. Rule 403 provides that "[a]lthough relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice." FED. R. EVID. 403. Particularly in unilaterally obtained lie-detection tests, where the test is sought by one of the parties and will not be used if the outcome is not to that party's liking, the probative value is small because the party has nothing to lose by taking the test.

courts' long-standing policy against admitting true lie-detection evidence, it is unlikely that the fMRI-based CQT will be admitted in the near future.

C. The Daubert *Standard Applied to the CIT*

Unlike the fMRI-based CQT, the P300-based CIT has over twenty years of published research supporting it.[186] Despite this history, the test has rarely been proffered as evidence in court and has been largely ignored in both case law and the legal literature.[187] There are currently no published cases that have conducted a formal *Daubert* hearing regarding the admissibility of either a P300-based or polygraph-based CIT. However, a form of the P300-based CIT has been proffered in two state court post-conviction murder exoneration cases: *Harrington v. State*, an Iowa case in which the P300-based CIT was partially admitted into evidence,[188] and *Slaughter v. State*, an Oklahoma case in which the P300-based CIT was rejected.[189] Both cases considered the admissibility of a P300-based CIT conducted by Brain Fingerprinting Laboratories, started by Dr. Lawrence Farwell, which is the only company that currently offers a commercially available CIT.[190] Despite these cases and the increasing prevalence of neuroimaging evidence in court, no article has thoroughly examined the P300-based CIT's admissibility under *Daubert*.[191] The remainder of this Part discusses how the P300-based CIT would fare under each of these factors and concludes that while the P300-based CIT should not currently

---

[186] *See, e.g.*, Rosenfeld et al., *supra* note 48.

[187] A Westlaw search of the "All State and Federal Cases" database conducted on July 6, 2012, for "concealed information test" yielded only one result. A search in the same database for "guilty knowledge test" (a synonymous term for the CIT) yielded seven additional responses. A search for "polygraph" on the same database returned over 10,000 cases.

[188] No. PCCV 073247, at *5–10 (Iowa Dist. Ct. Mar. 5, 2001).

[189] 105 P.3d 832, 835–36 (Okla. Crim. App. 2005). The *Slaughter* court noted that "Dr. Farwell makes certain claims about the Brain Fingerprinting test that are not supported by anything other than his bare affidavit." *Id.* at 834–35. The court also opined that there was no evidence that Brain Fingerprinting had been tested, subjected to peer-review, or otherwise met the *Daubert* factors and thus would fail such an analysis. *Id.* at 836.

[190] BRAIN FINGERPRINTING, http://www.governmentworks.com/bws/ (last visited Aug. 15, 2012). A brain fingerprinting test is similar to other P300-based CITs. A subject is shown probe items that relate to a particular crime or event and irrelevant items of the same class that do not relate to the crime or event. A larger P300 is expected to occur to probe items if the subject has knowledge of the crime or event in question. *See* Lawrence A. Farwell & Sharon S. Smith, *Using Brain MERMER Testing to Detect Knowledge Despite Efforts to Conceal*, 46 J. FORENSIC SCI. 135, 136 (2001). However, in addition to the P300, brain fingerprinting uses a secondary brain response, termed by Farwell as the MERMER effect. *See id.* at 135. The basis for the MERMER effect is never explained by Farwell as it is the proprietary secret of the Brain Fingerprinting company. *See* J. Peter Rosenfeld, *'Brain Fingerprinting': A Critical Analysis*, 4 SCI. REV. MENTAL HEALTH PRAC. 20, 23 (2005).

[191] Some articles have discussed the admissibility of the P300-based CIT under *Daubert*, but have not discussed each factor in detail. *See, e.g.*, Erikson, *supra* note 61; Eric K. Gerard, *Waiting in the Wings? The Admissibility of Neuroimagery for Lie Detection*, 27 DEV. MENTAL HEALTH L., July 2008, at 1.

be admissible under *Daubert*, it should be admissible in the future once more critical studies are conducted to address its accuracy in real-world situations.

1. Testability.—Testability as described by the *Daubert* Court is likely the easiest factor for the P300-based CIT to meet. Testability, or falsifiability, of a scientific theory is typically considered the chief criterion of scientific status.[192] If a particular test or theory is not falsifiable, it will generally not be considered a science and thus will struggle to meet the other *Daubert* criteria.[193] In this manner, the testability factor could be considered a threshold standard, though it was not labeled as such by the *Daubert* Court.

Courts would very likely consider the P300-based CIT testable. Like any other diagnostic test, the P300-based CIT's accuracy can be determined by conducting the test on an individual for whom ground truth is known.[194] For example, many tests of the P300-based CIT compare CIT results from participants who commit a mock crime with results from participants who have not committed any crime but are given the same CIT, thereby directly testing diagnostic accuracy.[195]

One criticism of the CIT's testability is that it is very difficult to assess the CIT's accuracy in real-world conditions because it would be nearly impossible to know whether the suspect had actually committed the crime. While this complaint is a valid one for the rate of error factor in the sense that experiments of this type have not been conducted and so the error rate of the P300-based CIT in the field is not known, it is not true that these types of tests could not be conducted. If the CIT were given to a number of criminal suspects and incontrovertible independent evidence (such as DNA testing) were later revealed that determined the suspect's guilt or innocence, the effectiveness of the CIT in the field could be assessed. Thus, the CIT is testable, and though one could argue that it has not yet been fully tested in the most realistic scenarios, its basic reliability has been repeatedly tested.[196]

---

[192] Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593 (1993) ("[T]he criterion of the scientific status of a theory is its falsifiability, or refutability, or testability." (quoting KARL R. POPPER, CONJECTURES AND REFUTATIONS: THE GROWTH OF SCIENTIFIC KNOWLEDGE 37 (5th ed. 1989) (emphasis omitted))). A science is generally understood to be falsifiable when it is "capable of empirical test" such that the theories on which the method is based could be objectively proven wrong. *See id.* (quoting CARL GUSTAV HEMPEL, PHILOSOPHY OF NATURAL SCIENCE 49 (1966)).

[193] *See* David L. Faigman et al., *Admissibility of Scientific Evidence*, 1 MODERN SCIENTIFIC EVIDENCE, *supra* note 20, § 1:16.

[194] Ground truth is known in a CIT where the researchers know, a priori, whether a participant possesses the concealed information they are attempting to detect. Thus, the researchers can determine whether their test came to the correct result, unlike in cases where ground truth is not known.

[195] *See, e.g.*, Meixner & Rosenfeld, *supra* note 44, at 2.

[196] *See, e.g.*, Farwell & Donchin, *supra* note 48; Meixner & Rosenfeld, *supra* note 44; Rosenfeld et al., *supra* note 48.

2. Peer Review, Publication, and General Acceptance in the Field.—The P300-based CIT is also likely to pass *Daubert*'s "peer review" factor. The *Daubert* Court specifically noted that this factor "does not necessarily correlate with reliability" but instead reflects "the likelihood that substantive flaws in methodology will be detected."[197] The CIT has undergone substantial peer review.[198] As a group of researchers repeatedly publish papers using similar methods, these methods become exposed to multiple phases of peer review designed to eliminate methodological errors. For example, the Rosenfeld lab at Northwestern University has produced results of a highly accurate P300-based CIT protocol across three separate publications.[199] With the repetition of similar studies, it is likely that the peer review process will weed out major methodological flaws.

When a court is considering whether a particular method has been subjected to peer review, it must examine the specific method being presented. In *Harrington*, the court noted that while the P300 component itself had been "extensively tested and reviewed," Brain Fingerprinting, which uses a specific proprietary "MERMER" algorithm, had not been tested.[200] Thus, a CIT relying only on the well-published P300 brain wave will be considered to be peer reviewed, while modified versions that have not been published, like Brain Fingerprinting, will not pass muster.

Many of the same considerations apply to the question of whether the P300-based CIT is "generally accepted" in the field. It is not sufficient that the P300 brain wave itself is well-established in the field; its use as an indicator of concealed information must be generally accepted as well. The *Harrington* court accepted that the P300 is a reliable marker of the recognition of salient information, but was unclear as to whether it accepted

---

[197] *Daubert*, 509 U.S. at 593.

[198] A search on Google Scholar for "P300-based concealed information test" and "P300-based guilty knowledge test" conducted on July 6, 2012, revealed more than fifty peer-reviewed publications of empirical tests of the CIT on only the first five pages of results, and many more papers have undoubtedly been published. These publications have come from a number of labs both within the United States and abroad, indicating peer-reviewed replication of the validity of the CIT.

[199] *See* Meixner & Rosenfeld, *supra* note 44; Rosenfeld et al., *supra* note 49; Michael R. Winograd & J. Peter Rosenfeld, *Mock Crime Application of the Complex Trial Protocol (CTP) P300-Based Concealed Information Test*, 48 PSYCHOPHYSIOLOGY 155 (2010). It should be noted that the research coming from this lab is, in part, my own. However, other labs have also tested the P300-based CIT. *See, e.g.*, John J. Allen et al., *The Identification of Concealed Memories Using the Event-Related Potential and Implicit Behavioral Measures: A Methodology for Prediction in the Face of Individual Differences*, 29 PSYCHOPHYSIOLOGY 504 (1992); Farwell & Donchin, *supra* note 48; Kenta Kubo & Hiroshi Nittono, *The Role of Intention to Conceal in the P300-Based Concealed Information Test*, 34 APPLIED PSYCHOPHYSIOLOGY & BIOFEEDBACK 227 (2009).

[200] *See* Harrington v. State, No. PCCV 073247, at *8–9 (Iowa Dist. Ct. Mar. 5, 2001). A full explanation of the basis of the MERMER algorithm and the difference between Brain Fingerprinting and the more mainstream P300-based CIT is beyond the scope of this Comment. For an excellent discussion of this topic, see Rosenfeld, *supra* note 190.

the applied use of the P300 in a CIT.[201] However, the court clearly rejected the Brain Fingerprinting MERMER effect for lack of general acceptance in the field,[202] a view that has been supported by a recent publication discussing the many problems with Brain Fingerprinting.[203]

Unlike the specific Brain Fingerprinting MERMER method, individual surveys have shown that the CIT itself is generally accepted.[204] William Iacono & David Lykken surveyed members of the Society for Psychophysiological Research, asking them whether the CQT and CIT were "based on scientifically sound psychological principles or theory."[205] Only 36% of the respondents stated that the CQT was based on scientifically sound principles, but 77% agreed that the CIT was scientifically sound.[206] Among members of the American Psychological Association, the results were very similar, with 30% and 72% agreeing that the CQT and CIT were scientifically sound, respectively.[207] This survey was conducted in 1997, and with the increasing existence of publications regarding the CIT, it is likely that this acceptance rate has increased. While these numbers do not indicate complete confidence among the scientific community in the CIT, it is unlikely that the general acceptance factor would keep the CIT out of the courtroom.

3. Rate of Error.—The *Daubert* Court did not specify what error rate is tolerable for admissibility. This has left courts confused in interpreting error rates of various types of scientific testimony.[208] However, while this may seem like simple oversight or reluctance to form an easily applied bright-line rule on the part of the *Daubert* Court, the failure to specify a maximum error rate is defensible on the grounds that differing error rates are acceptable in different contexts. Where the cost of making a mistake is higher, a lower maximum error rate should be required.[209] For example, "a judge might require a relatively low error rate before admitting predictions of violence in a capital case, but permit higher error rates in a probation matter."[210] This distinction does not run merely across the potential level of punishment either. Thus, a Federal Rule of Evidence 403 prejudice determination is naturally intertwined with determining the acceptable level of error—in cases where the expert's testimony is highly likely to prejudice

---

[201] *Harrington*, No. PCCV 073247, at *9.

[202] *Id.*

[203] *See* Rosenfeld, *supra* note 190.

[204] *See, e.g.*, W.G. Iacono & D.T. Lykken, *The Validity of the Lie Detector: Two Surveys of Scientific Opinion*, 82 J. APPLIED PSYCHOL. 426, 426–28 (1997).

[205] *Id.* at 430 tbl.2.

[206] *Id.*

[207] *Id.*

[208] *See* Faigman, *supra* note 193, § 1:20.

[209] *See id.* § 1:21.

[210] *Id.*

the jury and become a critical piece of evidence in the case, the error rate will likely be viewed more strictly.

Under this rationale, one can envision the error rate of the CIT being judged very differently depending on the context of the case. Research on the CIT has been predominantly in the criminal context,[211] making the evidence likely to be considered more prejudicial and thus requiring a higher standard of accuracy. In a murder trial, for example, the result of the CIT could be a critical piece of information that determines whether the defendant is convicted, and thus likely to be highly prejudicial. In this case, a trial judge would be wise to require a high accuracy rate. However, in other applications of the CIT where the potential prejudice is lower, such as showing that an employee had knowledge of his employer's discriminatory hiring practices, a higher rate of error may be acceptable.

While this determination will always be a case-by-case decision for the trial judge, recent P300-based CITs show promise to meet even the high standard of accuracy that might be required at a criminal trial. One recent study achieved a 100% accuracy rate among twenty-four participants being tested for knowledge of a mock terrorist attack.[212] Another paper from the same laboratory reported a 92% accuracy rate, with one false negative and one false positive among twenty-four participants.[213] Most other studies have typically found an 80% to 90% overall accuracy rate.[214]

However, the overall accuracy rate of a method is not the only consideration. The error rate is comprised of two types of errors: "false positives," which are cases in which a person does not have concealed knowledge but is detected as having such knowledge, and "false negatives," which are cases in which a person has concealed knowledge but it is not detected by the test. These two types of errors cannot be viewed as equal: in the criminal trial, for example, where the burden of proof is designed to ensure that innocent parties are not wrongfully convicted, a low false-positive rate is extremely important in any type of forensic test.

One of the CIT's strong points is that the false-positive error rate is extremely low.[215] In a CIT where the participant is shown one probe item and four irrelevant items, an innocent participant would be expected to show the largest P300 response to the probe item entirely by chance 20% of the time. So, for example, if an individual accused of murder were shown

---

[211] *See, e.g.*, Meixner & Rosenfeld, *supra* note 44.

[212] *Id.* at 150–51 & tbl.1.

[213] Rosenfeld et al., *supra* note 49, at 913 tbl.3b.

[214] *See, e.g.*, Allen et al., *supra* note 199; Farwell & Donchin, *supra* note 48, at 539 tbl.2; Rosenfeld et al., *supra* note 48, at 161; J. Peter Rosenfeld et al., *Simple, Effective Countermeasures to P300-Based Tests of Detection of Concealed Information*, 41 PSYCHOPHYSIOLOGY 205, 209 & tbl.1 (2004). One recent test reported an accuracy rate as low as 50%. Mertens & Allen, *supra* note 152, at 293 tbl.3.

[215] *See* Iacono, *supra* note 24, at 689 ("A properly administered [CIT] with a sufficient number of items has almost no chance of producing a false positive outcome.").

different weapons, with a gun being the weapon used to commit the crime and four other weapons as irrelevants, an innocent person would have the largest response to the gun one-fifth of the time. However, as more classes of questions are presented, this false-positive rate becomes exponentially smaller. Thus, if the same individual were shown five weapons in one block of testing and five addresses where the crime might have been committed in a separate block, an innocent individual would only be expected to show the largest P300 response to both probe items in one out of 25 cases, or about 4% of the time. If a third block is added, this number shrinks to one out of 125, and so on. A recent study that used this approach of testing multiple blocks of items found a 0% false-positive rate,[216] and most other P300-based CITs have reported false-positive error rates below 10%.[217]

Even if high rates of false negative errors are found in a P300-based CIT conducted on real crime suspects, there is still hope for admissibility. If the false-positive error rate approaches zero, then even with a high false-negative rate, the test still has probative value because in cases where the test produces a positive result, jurors can be more certain that the defendant has concealed knowledge.

Looming over all of these considerations, however, is one very large problem for proponents of the P300-based CIT: there have not been any studies of the P300-based CIT conducted on actual criminal suspects and seeking information about their crimes. Thus, it remains unknown whether the reported high accuracy rates from laboratory tests will translate to the field. There is some reason for concern because one polygraph-based CIT study conducted on actual suspects found high rates of false negative errors.[218] Such a test has not been done using the P300-based CIT. This problem was specifically noted by the *Semrau* court with regard to the fMRI-based CQT and is likely to be fatal to attempts to admit a P300-based CIT until such studies are conducted.[219] Interestingly, accuracy rates in the field may actually exceed those in the laboratory because criminals who planned crimes might be intimately familiar with the details of the crime, leading to better recognition of those details and thereby larger P300 responses as compared to test participants who are only briefly exposed to the crime-related details.[220] However, this remains conjecture; it should not be relied upon until it has been empirically tested.

---

[216] Meixner & Rosenfeld, *supra* note 44, at 151 tbl.1. There is no reason that other CITs could not follow the same approach; one could employ a block for every crime-related detail that a suspect is likely to recognize.

[217] *See, e.g.*, Rosenfeld et al., *supra* note 48.

[218] Eitan Elaad, *Detection of Guilty Knowledge in Real-Life Criminal Investigations*, 75 J. APPLIED PSYCHOL. 521, 526 (1990) (correctly detecting only 50% of guilty individuals).

[219] United States v. Semrau, No. 07-10074 Ml/P, 2010 WL 6845092, at *11 (W.D. Tenn. June 1, 2010) ("[T]here are no known error rates for fMRI-based lie detection outside the laboratory setting, i.e. in the 'real-world' or 'real-life' setting.").

[220] *See* Meixner & Rosenfeld, *supra* note 44, at 154.

Thus, the CIT likely meets the testability, peer review, and general acceptance prongs of *Daubert*. However, until field studies are conducted using the P300-based CIT, courts are unlikely to find error rates generated through laboratory tests to be sufficient to admit the CIT into evidence.

CONCLUSION

The law is at a crossroads regarding the admissibility of expert testimony that assists the trier of fact in assessing credibility. As the accuracy of fMRI and P300-based methods improve and become more available to litigants both in the civil and criminal realms, courts must make moral decisions regarding the admissibility of evidence in addition to decisions based solely on accurate trial outcomes.

In addition to the factors discussed above that are relevant to admissibility, courts will have to decide whether they truly *want* machines to aid jurors in determining whether a witness is lying or whether a witness had knowledge of information relating to a crime. Even a test that is accurate enough to meet the *Daubert* standard will have serious implications for perceived systemic legitimacy if it is persuasive enough to yield a conviction without other strong supporting evidence but is not accurate enough to ensure that an innocent person is never misdiagnosed.[221] It is not clear whether this reduction in procedural justice would be worth the gain in trial accuracy, though one could argue that modern forensic science has the same problem yet continues to be admitted. Additionally, there is a certain unease with using a mental state as evidence of a crime, and an undesirable invasion of privacy could result from broad use of these tools. These philosophical and policy questions, while beyond the scope of this Comment, are worthy of further consideration, and must be approached before we hand credibility assessment off to scientific experts.

However, the case against the trier of fact's status as the most accurate assessor of credibility continues to mount. The American common law has consistently restricted decisions of credibility to the trier of fact, but recent social science indicates that judges and jurors do not accurately use demeanor cues offered by a witness to assess credibility. While studies that put the testimony of a liar or truth teller in context show more promise for the ability of laypeople to assess the veracity of witnesses, they still do not approach the accuracy rates boasted by the modern P300-based CIT. However, the jury is still out on the jury's ability to detect lies in the courtroom, as studies that provide mock jurors with the rich context of a full trial have not been conducted.

For now, the science behind modern neuroscience-based credibility-assessment tools has not yet reached the standards that *Daubert* requires for

[221] *Cf. supra* note 79 (discussing potential goals of the trial system that are not related to outcome accuracy).

admissibility at trial. The fMRI-based CQT remains inconsistent in accuracy rates and has the same theoretical issues that have plagued the polygraph-based CQT for years. The P300-based CIT, however, meets many of the *Daubert* factors—it has been well tested throughout the psychological literature and has consistently achieved high levels of accuracy. Still, admissibility is not appropriate until realistic field tests of the P300-based CIT are conducted.

However, the science behind these methods should progress quickly. As it does, courts should not preclude these devices, which could help the trier of fact make more accurate credibility decisions, leading to more accurate trial outcomes. As it becomes more apparent that the jury alone may not be the most accurate assessor of credibility, courts should consider abandoning the tradition of the trier of fact as the sole assessor of credibility. Though these methods are not yet ready for use in trials, the court should not render its verdict against neuroscience-based credibility assessment before it has heard all of the evidence.