The limited effect of electroencephalography memory recognition evidence on assessments of defendant credibility

Francis X. Shen*,†, Emily Twedell‡, Caitlin Opperman**, Jordan Dean Scott Krieg††, Mikaela Brandt-Fontaine†‡, Joshua Preston***, Jaleh McTeigue†††, Alina Yasis‡‡‡‡ and Morgan Carlson****

* Corresponding author: E-mail: fxshen@umn.edu

† Associate Professor of Law, University of Minnesota; Director, Shen Neurolaw Lab; Executive Director of Education and Outreach, MacArthur Foundation Research Network on Law and Neuroscience; Affiliated Faculty Member, Massachusetts General Hospital Center for Law, Brain & Behavior.
‡ Emily Twedell, B.A., graduated from Grinnell College with a degree in Psychology and Neuroscience, and is a postgraduate research assistant in the Shen Neurolaw Lab.
** Caitlin Opperman, B.A., is a law student at the University of Minnesota Law School and a research assistant in the Shen Neurolaw Lab. Prior to law school, she earned her undergraduate degree in psychology and worked as a clinical psychometrist.
†† Jordan Dean Scott Krieg is an undergraduate in the Department of Neuroscience at the University of Minnesota, and a research assistant in the Shen Neurolaw Lab.
‡‡ Mikaela Brandt-Fontaine is an undergraduate in the Department of Neuroscience at the University of Minnesota and a research assistant in the Shen Neurolaw Lab.
*** Joshua Preston, B.A., is a JD/MA bioethics joint-degree student at the University of Minnesota Law School, a research assistant in the Shen Neurolaw Lab, and a research fellow at the Center for Science and Law.
††† Jaleh McTeigue is an undergraduate student at Mount Holyoke College, where she has designed the college’s first degree program in Neuroscience and Law. In summer 2016 she was a research assistant in the Shen Neurolab.
‡‡‡‡ Alina Yasis, B.S., graduated from the University of Minnesota with a major in psychology and a minor in neuroscience. She is a research assistant in the Shen Neurolaw Lab.
**** Morgan Carlson, B.A., J.D., is an attorney at Adams, Rizzi, Sween in Austin, Minnesota. She graduated from Saint Mary’s University of Minnesota with a major in Psychology, and graduated from the University of Minnesota Law School, where she worked as a research assistant in the Shen Neurolaw Lab.
INTRODUCTION

Recent empirical studies confirm that the use of neuroscientific evidence is rising. This parallels a rapid increase in legal scholarship on neurolaw, and suggests a trend toward more brain evidence in courts in the coming years. As a consequence, admissibility decisions on brain science are now being made by courts, jury experts are debating the use of brain science in the courtroom in practitioner journals, and brain evidence is being used across a variety of substantive areas of law. The admissibility of neuroscientific evidence is becoming an increasingly important issue for American courts.

To be admissible, neuroscience expert testimony and exhibits must be relevant and must meet the jurisdiction’s requirements for scientific evidence (such as Daubert or Frye). But even if deemed relevant and of sufficient scientific merit, the evidence may still be excluded under Federal Rule of Evidence 403 (or its state equivalent) if the evidence’s probative value ‘is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence’.

Scholars have suggested that in the context of neuroscientific evidence generally, and brain evidence related to deception in particular, Rule 403 concerns may be particularly salient. For instance, in the first and only federal court ruling on the admissibility

---

3 Gaudet & Marchant, supra note 1 at 582. (‘As the underlying science continues to grow and evolve, so does the judiciary’s evaluations and rulings on this type of evidence. In a few jurisdictions, certain types of neuroimaging evidence have not been admitted to date, yet the same techniques and similar evidence have been admitted elsewhere.’)
5 Owen D. Jones, Jeffrey D. Schall & Francis X. Shen, Law and Neuroscience (2014).
6 Frye v. United States is a 1923 case that established ‘general acceptance’ as a litmus for admissibility. The ‘Daubert trilogy’ refers to a series of three Supreme Court cases regarding the admissibility of expert evidence: Daubert v. Merrell Dow Pharmaceuticals (1993), Gen. Elec. Co. v. Joiner (1997), and Kumho Tire Co., Ltd. v. Carmichael (1999). Daubert established the so-called gatekeeper approach, in which trial judges, at least in theory, should engage more actively in screening proffered scientific testimony. Joiner held that abuse of discretion is the appropriate standard of appellate review for district court admissibility decisions. Kumho Tire clarified that Daubert applied both to ‘scientific’ knowledge as well as other technical and specialized knowledge.
7 Fed. R. Evid. 403.
8 See discussion infra, part I, at 4.
of functional magnetic resonance imaging (fMRI) lie detection evidence, the court excluded the evidence on both Rule 702 and Rule 403 grounds.9

Yet despite concerns under Rule 403 about the prejudicial effects of neuroscientific evidence, the scholarly empirical literature on the effects of such evidence is decidedly mixed.10 Across nearly 30 previous studies, including over 50 unique experiments, the only result researchers can agree upon is that there are ‘conflicting results’.11 At present, the ‘likely effect of neuroscientific evidence in legal settings is still unclear’.12 Or, as Baker et al. (2015) describe it in a review, ‘empirical research into the neuroimage bias has produced what might appear to be a tangled mess of contradictory findings … [and a] research quagmire’.13 Recent research suggests that although neuroscience information may be persuasive under certain conditions,14 brain images themselves are not independently persuasive, and neuroscience is likely not any more persuasive than other scientific evidence.15 Amidst this quagmire of conflicting results, there is a growing appreciation that context matters.16 That is, research going forward is likely not to address ‘Does neuroscientific evidence affect outcomes?’ (inviting a binary Yes/No answer), but rather ‘How much and under what circumstances does neuroscientific evidence affect outcomes?’

This article reports on new results from a study examining the effect of neuroscientific evidence on subjects’ evaluation of a fictional criminal fact pattern, while manipulating the strength of the non-neuroscientific evidence. By manipulating the strength of the case, we are able to estimate the marginal effect of introducing neuroscientific evidence.

We do this in the context of brain-based memory recognition with electroencephalography (EEG) evidence. EEG measures electrical signals in the brain, and brain-based memory recognition technology utilizes techniques designed to ascertain whether an individual’s brain recognizes particular information at the time of presentation. As we describe in part II, researchers have developed multiple approaches, with labels such as ‘brain fingerprinting’, ‘concealed information test’, and ‘guilty knowledge test’. While there are differences across techniques, there is agreement that this technology has legitimate potential for courtroom use. Indeed, unlike fMRI lie detection technology, EEG-based memory recognition evidence has already been admitted into a US court.17

9 United States v. Semrau, No. 07-10074 ML/P, 2010 WL 6845092, at *14 (W.D. Tenn. June 1, 2010) (holding, in agreement with the government ‘that the probative value of Dr. Laken’s testimony is substantially outweighed by the danger of unfair prejudice to the government’). For further discussion of the case, see Francis X. Shen & Owen D. Jones, Brain Scans As Evidence: Truths, Proofs, Lies, and Lessons, 62 MERCER L. REV. 861 (2011).
10 See discussion infra, part I, at 4.
11 Appelbaum & Scurich, infra note 54, at 92.
12 Appelbaum et al., infra note 54, at 136.
13 Baker et al., infra note 42, at 2.
15 Hopkins et al., infra note 55.
16 Baker et al., infra note 42 at 2 (suggesting that this line of ‘research demonstrates that these understandings are often multifarious, ambiguous, and very much dependent on context and framing’).
17 Harrington v. State, 859 N.W.2d 509 (Iowa 2003).
In two experiments, one using 868 online subjects and one using 611 in-person subjects, we asked subjects to read two short, fictional vignettes describing a protagonist accused of a crime. In one vignette, the protagonist was an employee accused of stealing a diamond necklace. The employee denied ever having seen the necklace before. In the second vignette, the protagonist was a stock trader accused of insider trading. The trader denied ever having seen the memo with insider information. Subjects read the vignette, and then answered Yes or No as to whether they thought the employee stole the necklace, and whether the trader saw the memo. We manipulated (i) the expert evidence (none, incriminating brain evidence, exculpating brain evidence, incriminating polygraph evidence, and exculpating polygraph evidence), and (ii) the strength of the non-neuroscientific facts against the defendant (weak facts with an alibi, medium facts, and strong facts with a motive).

We found that although there is a statistically significant relationship between exposure to neuroscientific information and subjects’ evaluations of the fictional defendant, the neuroscientific evidence was not as powerful a predictor as the overall strength of the case in determining outcomes. Our primary conclusion is that subjects are cognizant of, but not seduced by, brain-based memory recognition evidence. Subjects consider the evidence, and it has an effect in some contexts on their evaluations, but they generally weigh it as just one of many facts on the record.

Our results, remain exploratory and limited in their forensic use. We see great value in advancing collaborations that further explore the utility and admissibility of brain-based memory recognition evidence. But in this article we do not speak to the validity or reliability of EEG memory recognition technology as it would be applied in forensic settings. And, to be abundantly clear: we are not arguing that—given the present state of the science—courts should immediately admit brain-based memory recognition evidence. We are, however, arguing that if the evidence is both relevant and reliable, then Rule 403 may not necessarily be a barrier for admissibility.

The article is organized as follows. Part I situates the article within the emerging literature on the effects of neuroscientific evidence on evaluations of scientific credibility and legal outcomes. Part II reviews the science of EEG-based memory recognition and its forensic use to date. Part III presents the design and results of experiment 1, an online experiment using Amazon Mechanical Turk subjects. Part IV presents the design and results of experiment 2, using adult subjects recruited at the Minnesota State Fair. Part V discusses the implications of these results.

I. EFFECTS OF NEUROSCIENCE INFORMATION

Concerns have been raised both about the potential of neuroscience generally (as compared with social science evidence) and neuroimaging in particular (PET and especially fMRI) to be overly persuasive. The ‘seductive allure’ hypothesis, owing its name to an oft-cited 2008 article by psychologist Deena Weisberg and colleagues, is that
neuroscientific evidence might ‘seduce’ its consumers. Yet despite widespread concern, the empirical evidence testing the seductive allure hypothesis remains quite mixed. In this section we review that evidence, with special attention to the possibility of seductive allure in legal contexts.

Although we know neuroscientific evidence is being proffered in courts, we don’t know to what effect. In four articles published in 2016 in the *Journal of Law and the Biosciences*, the authors recognized that even when it was possible to identify the introduction of brain evidence, it was difficult to track how, if at all, the evidence affected case outcomes. Anecdotal evidence suggests that the effects of the evidence may be mixed. For instance, interviews with a small set of California trial court judges suggest mixed reactions to the introduction of behavioral genetics evidence on mental disorder issues. In a death penalty case in Florida, however, two jurors reported in post-verdict interviews that brain image evidence derived from quantitative EEG had been persuasive to them in choosing life without parole instead of the death penalty.

In addition to the anecdotes, there is a growing empirical literature on the impact of neuroscientific evidence. We are aware of 29 studies that have explored whether neuroscientific information affects outcomes. Of these experiments, nearly half specifically concerned legal contexts. The remainder of the experiments generally focused on whether adding neuroscientific information (and brain images) affected credibility ratings of scientific journal articles.

The literature begins with three studies published in 2008. We review each in turn. Psychologists David McCabe and Alan Castel (2008) examined how the inclusion of a brain image affected college students’ ratings of the quality of scientific articles.

---

20 Farah & Hook, *supra* note 18 (reviewing the literature as of 2013 and finding that beyond the studies published in 2008 ‘little additional evidence has been published in support of the disproportionate persuasiveness of brain images’). Fernandez-Duque (2015) similarly note that, up until their study, ‘after failures to replicate McCabe and Castel’s initial findings, [the Weisberg et al 2009 study is] ... currently the single piece of evidence in support of the allure of neuroscience’.
21 Chandler (2016) *supra* note 1 (noting that ‘impact of the neuroscientific evidence was often challenging to identify’).
24 There is a scholarly debate about the relative importance of experimental versus historical/archival methods. We take no position on this debate, as we see value in both lines of research. See Lisa G. Aspinwall et al., *The Double-Edged Sword: Does Biomechanism Increase or Decrease Judges’ Sentencing of Psychopaths?* 17 SCIENCE 846 (2012); Deborah W. Denno, *What Real-World Criminal Cases Tell Us About Genetics Evidence*, 64 HASTINGS L.J. 1591 (2013) (arguing that the Aspinwall et al. study was ‘significantly flawed’); Teneille R. Brown et al., *Understanding Validity in Empirical Legal Research: The Case for Methodological Pluralism in Assessing the Impact of Science in Court*, 67 HASTINGS L. J. 1067 (2016).
25 Across the 29 studies, there are 53 unique experiments (as single publications frequently report results from more than one unique experiment).
image, as compared to articles with no image and articles with a bar graph. McCabe and Castel next explored whether this effect was simply due to the visual complexity of the brain image, by running a follow-up experiment in which they found higher ratings of articles with a brain image, compared to ratings of an article with a topographical map of brain activation. They added a third experiment, finding that subjects were more likely to agree with a scientific news article’s conclusion that brain imaging could be used as a lie detector when a brain image was included with the article. This result held even when the researchers introduced a paragraph of criticism on such use of the technology. McCabe and Castel concluded from this series of experiments that their ‘results lend support to the oft mentioned notion that there is something particularly persuasive about brain images with respect to conferring credibility to cognitive neuroscience data’.28

Consistent with McCabe and Castel’s conclusions, psychologist Deena Weisberg and colleagues (2008) asked: Why should brain images be so persuasive? Weisberg et al. speculated that the brain image served as a (misleading) proxy for quality.29 To test this theory, Weisberg et al. wrote ‘good’ and ‘bad’ descriptions for each of 18 psychological phenomena (such as attentional blink). Then, they created extraneous neuroscience information which added nothing of scientific value to the explanations.30 Subjects were randomly assigned to either the ‘with neuroscience’ or ‘without neuroscience’ conditions.31 Weisberg et al. found that neuroscience information was significantly and positively related to satisfaction ratings, and in particular that it boosted ratings of the bad explanations.32

22 cognitive neuroscience students subsequently participated in a within-subjects experiment using the same stimuli as experiment 1. The researchers hypothesized that because these students were learning to be critical readers of neuroscience studies, the main effect of neuroscience information would not be present. But, counter to expectations, they found that the neuroscientific information was still a significant predictor of satisfaction ratings.33 And, just as with the first experiment, Weisberg et al. found that neuroscience information caused the bad explanation ratings to increase even more than good explanation ratings.34 The researchers then

---

27 The ratings were for the quality of the scientific reasoning in the article.
28 McCabe & Castel (2008) at 349. However, they also noted that ‘the effect of brain images in the current study was not large’. Id. They speculated that this may be because—even in the no-image group—subjects were particularly likely to believe research in cognitive neuroscience.
29 Weisberg et al. (2008), supra note 19, at 470. (“The presence of neuroscience information may be seen as a strong marker of a good explanation, regardless of the actual status of that information within the explanation. That is, something about seeing neuroscience information may encourage people to believe they have received a scientific explanation when they have not.”)
30 The neuroscience information did not include a brain image, but did specify the brain region of activation for the given psychological phenomena.
31 Eighty-one subjects read a description of each of the 18 psychological phenomena, and then a paragraph of the explanation. It was in this explanation paragraph where the 2 × 2 design entered: good/bad explanation X with/without neuroscience. Subjects rated their satisfaction with the explanation on a seven-point scale.
32 Weisberg et al. (2008), supra note 19, at 472 (finding that ‘neuroscience information seems to have the specific effect of making bad explanations look significantly more satisfying than they would without neuroscience’).
33 Id. at 473.
34 Id.
added a third experiment, using 48 neuroscience ‘experts’ as subjects. The presence of neuroscience information did not raise the experts’ ratings. Indeed, when irrelevant neuroscience information was added to an otherwise good explanation, the experts spotted the disconnect and rated those explanations significantly lower.

Psychologists Jessica Gurley and David Marcus (2008) were the first to consider a legal context, finding that neuroimages, brain injury testimony, and a combination of both all significantly increased the likelihood of a defendant receiving a not guilty by reason of insanity verdict. Because all of the studies’ published in 2008 produced results that pointed in the same direction, conventional wisdom at the time tended to agree with the conclusion of Weisberg et al. that even ‘logically irrelevant neuroscience information can be seductive—it can have much more of an impact on participants’ judgments than it ought to.’

Since 2008, however, the existence and nature of the ‘seductive allure’ phenomenon has been heavily debated. We draw several conclusions from the 26 studies published between 2011 and 2016. First, there is no consensus on whether neuroscientific information matters—and if it does, in which direction. On one hand, some studies since 2008 have found significant effects in some contexts. Psychologist Diego Fernandez-Duque and colleagues (2015) revisited the Weisberg et al. (2008) study, to improve upon the methodology and further explore the mechanisms by which neuroscience information might affect evaluations. Fernandez-Duque found that subjects provided higher satisfaction ratings when the scientific article they read contained neuroscientific information, even as compared to social science and to other scientific information. A different study by psychologists Edie Greene and Brian Cahill (2012) found that both neuropsychological test results and neuroimaging mitigated death penalty findings—though only for defendants who were at low risk of future dangerousness. And Rhodes et al. (2014) found that article quality ratings were higher when neuroscience was included, even after controlling for article length and individual differences.

‘Experts’ were defined as ‘individuals who are about to pursue, are currently pursuing, or have completed advanced degrees in cognitive neuroscience, cognitive psychology, or strongly related fields’. The researchers confirmed through additional questions that the subjects did indeed have substantive expertise in neuroscience.

Id. at 474–5 (adding that ‘questioning of several participants in this study indicated that they were indeed sensitive to the awkwardness and irrelevance of the neuroscience information in the explanations’).


Id. at 475. Yet the researchers also noted that ‘impact of superfluous neuroscience information is not unlimited’ and that there is ‘noticeable benefit of extended and specific training on the judgment of explanations’.


They also found, however, that adding a neuroimage to the textual neuroscientific explanation did not affect ratings. In short, the Fernandez-Duque study confirmed the studies finding no effect of brain images, but suggested an effect of neuroscientific explanations over and above other scientific explanations.


Rebecca E. Rhodes et al., Explaining The Alluring Influence of Neuroscience Information on Scientific Reasoning, 40 J. EXP. PSYCHOL. LEARN. MEM. & COGN. 1432 (2014).
Yet there are a wide variety of other, sometimes counterintuitive, findings. For instance, psychologists David Gruber and Jacob Dickerson (2012) attempted to replicate and extend the McCabe and Castelf finding, but they found no significant relationship between inclusion of a brain image and subject evaluation of the news article. Weisberg et al. (2015) found differences in experimental results when comparing results from MTurk subjects and college student subjects on the same task. Rhodes et al. (2014) found that irrelevant neuroscientific information added to science articles had a differential effect on different types of article evaluation metrics. In law-specific studies, Saks et al. (2014) found evidence of a ‘backfire effect’, as defendants who proffered brain images were more likely to receive the death penalty than comparable defendants who did not proffer such evidence. McCabe et al. (2011) found an effect, but it vanished when cross-examination was introduced. Cheung and Heine (2015) found that genetic explanations had little relationship to legal judgements. Where research has found an effect, the effect size has generally been small.

Second, even if there is an effect, there is growing evidence to suggest that it is not the brain image per se that is causing it. Schweitzer et al. (2011) ran a series of experiments finding that brain images had no impact on criminal verdicts or sentencing.

---

43 For a review, see D. A. Baker et al., Making Sense of Research on The Neuroimage Bias, PUB. UNDERST. SCI. (2015): 0963662515604975.
44 David Gruber & Jacob A. Dickerson, Persuasive Images in Popular Science: Testing Judgments of Scientific Reasoning and Credibility, 21 PUB. UNDERST. SCI. 938 (2012). Gruber & Dickerson, supra note 44 at 945. Thus, they cautioned ‘suggests a need to re-think the oft-repeated and unmitigated declaration that brain images hold more persuasive power than other images in terms of their ability to inflate judgments of scientific reasoning, authoritativeness, or credibility’.
45 Weisberg et al. (2015), supra note 19 at 434. (‘It is not entirely clear why undergraduates would be more attracted to explanations containing neuroscience information than MTurk workers, although currently learning about psychological and neuroscientific phenomena might have swayed the undergraduates to lend more weight to the presence of neuroscience.’)
46 Rhodes et al., supra note 42 at 7. (‘Overall, we found an effect of irrelevant neuroscience on evaluations that is small for subjective ratings such as article quality, study quality, and convincingness of the evidence. However, we did find moderate-to-large effects for ratings of scientist quality and mechanistic understanding.’)
50 See Michael et al., infra note 53. A 2015 study also found a small effect size. Rhodes et al., supra note 42 at 5. (‘It should be noted that, consistent with a recent meta-analysis … the effect sizes of neuroscience were small.’)
51 Cayce J. Hook & Martha J. Farah, Look Again: Effects of Brain Images and Mind–Brain Dualism on Lay Evaluations of Research, 25 J. COGN. NEUROSCI. 1397 (2013); Farah & Cook, supra note 18 (reviewing the literature as of 2013 and finding that beyond the studies published in 2008 ‘little additional evidence has been published in support of the disproportionate persuasiveness of brain images’); Nicholas J. Schweitzer & Michael J. Saks, Neuroimage Evidence and the Insanity Defense, 29 BEHAV. SCI. & L. 592 (2011) (finding that it was the entirety of the neuroscientific evidence package, not the neuroimage itself, that mattered for verdict and sentencing outcomes); Weisberg et al. (2015), supra note 19, at 430 (noting that the consensus ‘strongly suggest[s] that neuroscience imagery is not the source of the effect’); Gruber & Dickerson, supra note 44; Madeline Keehner et al., Different Clues from Different Views: The Role Of Image Format in Public Perceptions of Neuroimaging Results, 18 PSYCHON. BULL. & REV. 422 (2011); Adina L. Roskies et al., Neuroimages in Court: Less Biassing Than Feared, 17 TRENDS COGN. SCI. 99–101 (2013).
They concluded that ‘the best tentative conclusion appears to be that neuroimages are not the irresistible force that some have feared’. Schweitzer’s group (2013) also used eye tracking to see if brain images were visually processed differently from bar graphs, and concluded that they were not.

Psychologist Robert Michael and colleagues conducted a meta-analysis, including McCabe and Castel’s (2008) data, and 10 new experiments that used their materials. Michael et al. also concludes that there is now ‘compelling evidence that when it comes to brains, the “amazingly persistent meme of the overly influential image” has been wildly overstated’. Psychiatrist Paul Appelbaum, psychologist Nicholas Scurich, and colleagues (2015) have similarly concluded that ‘it appears that fears that neuroimages, with their graphic portrayal of brain function, would come to dominate jurors’ decisions were considerably overblown’.

Third, if it is not the brain image itself, it is not yet clear what the mechanism of action is. It could be that the neuroscience jargon is persuasive; that subjects prefer the reductive nature of neuroscience explanations; that subjects prefer the 3-D perceptual quality of some images; that subjects prefer neuroscience that affirms their pre-existing policy views; that individual subject characteristics mediate the effect; some combination of these, or something else entirely. In sum, we agree with Baker et al.’s (2015) assessment of the current state of research: ‘… an overarching theory is still out of reach[,] and this … means the use of neuroimages to convey information to laypersons should continue to be scrutinized as the context and framing in which they are presented changes’. In this article, we focus on one particular context: the use of brain-based memory recognition evidence.

---

56 See Weisberg et al. (2015) and Hopkins et al. (2016) for reviews of this question and new studies to address it. Emily J. Hopkins et al., The Seductive Allure Is A Reductive Allure: People Prefer Scientific Explanations That Contain Logically Irrelevant Reductive Information, 155 COGNITION 67 (2016); Weisberg et al. (2015), supra note 19. Earlier studies had speculated that the mechanism had to do with the image itself. For instance, McCabe & Castel (2008 p. 249–50) speculated that ‘Brain images may be more persuasive than other representations of brain activity because they provide a tangible physical explanation for cognitive processes that is easily interpreted as such. This physical evidence may appeal to people’s intuitive reductionist approach to understanding the mind as an extension of the brain.’
58 Weisberg et al. (2015), supra note 19.
59 Keehner et al., supra note 51. Keehner et al. (2011) found that not all brain images were equal in affecting the ratings of a news article. It was the 3-D images, of an inflated brain and a whole brain, that had a significant effect on the ratings.
60 Nicholas Scurich & Adam Shniderman, The Selective Allure of Neuroscientific Explanations, 9 PLOS ONE e107529 (2014).
II. BRAIN-BASED MEMORY RECOGNITION USING EEG

Many forms of neuroscientific evidence are being proffered in court. We focus here on memory recognition with EEG because it seems at least plausible—and some would argue likely—that this type of evidence might one day meet the admissibility requirements of Daubert and Rule 702. In this section, we briefly review the state of the science. While it is beyond the scope of this article to evaluate fully the EEG memory detection literature within a Rule 702 framework, our cursory review finds compelling reason to believe that brain-based memory recognition evidence may meet the Daubert standard.

EEG, discovered in the 1920s, is a method of measuring electrical activity produced by the brain. The logistics of EEG are straightforward. Electrodes are placed on the human scalp, and they are used to measure electrical activity. The exact number and placement of the electrodes can vary. The electrical activity is recorded and analysed with the aid of a computer algorithm, and the results allow for inferences about brain function. In general, as compared to fMRI, EEG has better temporal resolution (meaning you can measure activity over small time intervals), but poorer spatial resolution (meaning it is more difficult to know which areas and networks in the brain are activated). EEG has significant value for both clinical and research purposes, and the techniques employed vary widely depending on the goals of the researcher/clinician. For instance, EEG can be used to aid in the diagnosis of brain death. EEG can also be used to monitor epileptic seizures.

For our purposes here, the technique of most interest is the ‘event-related potential’ (ERP) technique. ERPs are ‘electrical potentials generated by the brain that are related to specific internal or external events (e.g. stimuli, responses, decisions)’. There is an expansive literature on the use of ERPs in many settings. One line of research,
starting as early as the 1970s,\textsuperscript{71} has found that ERP techniques can be used to distinguish whether subjects are processing novel or familiar stimuli.\textsuperscript{72} The typical and highly robust finding is that old (studied) test items elicit more positive going waveforms relative to new (unstudied) test items.\textsuperscript{73}

The forensic use of ERP techniques is being explored primarily by two research groups: one led by Lawrence Farwell and one led by Peter Rosenfeld.\textsuperscript{74} While the two groups both employ ERP techniques using EEG, they use different methods, and have debated the relative merits of these methods.\textsuperscript{75} We do not aim to resolve the debate here, but rather to simply describe for the reader each method.

A. Brain fingerprinting

In 1991, Lawrence Farwell and Emanuel Donchin published an article that adapted the ‘Guilty Knowledge Test’ (GKT), also known as the ‘Concealed Information Test’ (CIT), for use with EEG ERP techniques.\textsuperscript{76} Since the 1991 study, Farwell has published

\begin{itemize}
\item \textsuperscript{71} See eg Emanuel Donchin, The Use of Electro cortical Activity to Monitor Human Decision Making, ARPA SEMI-
\item \textsuperscript{72} Michael D. Rugg & Tim Curran, Event-Related Potentials and Recognition Memory, 11 TRENDS COGN. SCI. 251 (2007); David Friedman & Ray Johnson, Event-Related Potential (ERP) Studies of Memory Encoding And Retrieval: A Selective Review, 51 MICROs. RES. & TECH. 6 (2000). (‘The majority of ERP memory studies have investigated the neural basis of explicit recognition and recall, and have demonstrated differential ERP responses depending on the study status of the item being recognized. In such recognition memory paradigms, processing new words accesses semantic memory whereas processing old words, having been experienced previously, accesses episodic memory.’)
\item \textsuperscript{73} Holger Hill & Sabine Windmann, Examining Event-Related Potential (ERP) Correlates of Decision Bias in Recognition Memory Judgments, 9 PLoS ONE e106411 (2014).
\item \textsuperscript{74} To be sure, these are not the only researchers who have, and who are, working in this area. For reviews, see William G. Iacono, Forensic Application of Event-Related Brain Potentials to Detect Guilty Knowledge, in DETECTING DECEPTION: CURRENT CHALLENGES AND COGNITIVE APPROACHES 81 (Par Anders Granhag, Aldert Vrij & Bruno Verschueren eds., 2015), and Gershon Ben-Shakhar, Current Research and Potential Applications of The Concealed Information Test: An Overview, 3 FRONT. PSYCHOL. 342 (2012). See also John J. Allen & William G. Iacono, A Comparison of Methods for The Analysis of Event-Related Potentials in Deception Detection, 34 PSYCHOPHYSIOLOGY 234 (1997); 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); C.D. Lefebvre et al., Determining Eyewitness Identification Accuracy Using Event-Related Brain Potentials (ERPs), 44 PSYCHOPHYSIOLOGY 894 (2007); Izumi Matsuda et al., Event-Related Potentials Increase the Discrimination Performance of the Autonomic-Based Concealed Information Test, 48 PSYCHOPHYSIOLOGY 1701 (2011).
\item \textsuperscript{76} Lawrence A. Farwell & Emanuel Donchin, The Truth Will Out: Interrogative Polygraphy (‘Lie Detection’) With Event-Related Brain Potentials, 28 PSYCHOPHYSIOLOGY 531 (1991). The CIT/GKT paradigm is different than the Control Question Test used by the professional polygraph community. Iacono, supra note 74; William G. Iacono, The Forensic Application of ‘Brain Fingerprinting’: Why Scientists Should Encourage the Use of P300 Memory Detection Methods, 8 AM. J. BIOETHICS 30 (2008).
\end{itemize}
multiple studies exploring and refining what he has labeled the ‘brain fingerprinting’ technique.\textsuperscript{77}

At the core of the brain fingerprinting technique is measurement of a particular brain response: the ‘memory and encoding related multifaceted electroencephalographic response’ or P300-MERMER. P300 refers to a brain wave with a peak wave appearing 300 ms after the stimuli. Farwell’s approach measures both the P300 wave and later evoked potentials including a late negative peak.\textsuperscript{78} Farwell has described in detail 20 steps, including multiple preliminary tasks and substeps, which must be taken to employ the brain fingerprinting technique.\textsuperscript{79}

It is beyond the scope of this article to review each step in detail, but at the core of the technique is the development and use of three types of stimuli: probes, targets, and irrelevants.\textsuperscript{80} As described by Farwell:

Probes contain information that is relevant to the crime or other investigated situation. Probes have three necessary attributes (Farwell 1994, 1995a, b; Farwell and Smith 2001; Iacono 2008): 1. Probes contain features of the crime that in the judgment of the criminal investigator the perpetrators would have experienced in committing the crime; 2. Probes contain information that the subject has no way of knowing if he did not participate in the crime; and 3. Probes contain information that the subject claims not to know or to recognize as significant for any reason.

By contrast, target stimuli ‘are details about the crime that the experimenter is certain the subject knows, whether or not he committed the crime’.\textsuperscript{81} The experimenter ‘tells the subject the target stimuli and explains their significance in terms of the crime’.\textsuperscript{82}

\textsuperscript{77} Farwell et al., \textit{supra} note 62; Lawrence A. Farwell et al., \textit{Brain Fingerprinting Classification Concealed Information Test Detects Us Navy Military Medical Information With P300}, 8 \textsc{Front. Neurosci.} 410 (2014); Lawrence A. Farwell & Sharon S. Smith, \textit{Using Brain MERMER Testing to Detect Concealed Knowledge Despite Efforts to Conceal}, 46 \textsc{J. Forensic Sci.} 135 (2001). For a review, see Lawrence A. Farwell, \textit{Brain Fingerprinting: A Comprehensive Tutorial Review of Detection of Concealed Information With Event-Related Brain Potentials}, 6 \textit{Cogn. Neurodyn.} 115 (2012); Lawrence A. Farwell, \textit{Brain Fingerprinting: Detection of Concealed Information}, in \textit{WILEY ENCYCLOPEDIA OF FORENSIC SCIENCE} (A. Jamieson & A.A. Moenssens, eds. 2014). Legal commentary has been mixed on brain fingerprinting. See eg Henry T. Greely & Judy Illes, \textit{Neuroscience Based Lie Detection: The Urgent Need for Regulation}, 33 \textsc{Am. J. L. & Med.} 377 (2007) (suggesting that ‘Farwell’s claims are widely discounted in the relevant scientific community and his credibility is not helped by his inflated claims for the judicial acceptance of his technique’). Charles N. W. Keckler, \textit{Cross-Examining the Brain: A Legal Analysis of Neural Imaging for Credibility Impeachment}, 57 \textsc{Hastings L. J.} 509 (2006) (‘Despite the claims of those using ERPs to detect “information” stored in the brains of the subjects, in fact only a responsive orientation reaction is usually detected, and this is correlated only in a rather loose way with whether the information has already been encoded. Without observing the actual processes of memory encoding and memory retrieval, as they pertain to a stimulus (i.e., whether the former or the latter occurs), it remains speculative whether or not the stimulus was present in the brain prior to its presentation.’) However, other commentators have been more enthusiastic. See eg Alexandra J. Roberts, \textit{Everything New Is Old Again: Brain Fingerprinting and Evidentiary Analogy}, 9 \textsc{Yale J. L. & Tech} 234, 270 (arguing that ‘Brain Fingerprinting seems to meet every standard of admissibility under Rule 702; scrutinize it according to its merits, not the merits of the thing it most closely resembles or the category into which it purports to fit’).

\textsuperscript{78} Farwell, \textit{supra} note 77 at 119. (‘The discovery of the P300-MERMER was one more step in the ongoing progression from very short latency evoked potentials to longer and longer latency event-related potentials as the stimuli and the processing demanded by the experimental task become more rich and complex.’)

\textsuperscript{79} Farwell, \textit{supra} note 77.

\textsuperscript{80} \textit{Id.}

\textsuperscript{81} \textit{Id.} Emphasis added.

\textsuperscript{82} \textit{Id.}
Finally, the irrelevant stimuli ‘contain information that is not relevant to the crime and not relevant to the subject’.83 The irrelevants are ‘incorrect but plausible crime features’.84

To illustrate, imagine that police are investigating a homicide, in which the victim was killed by a machete knife. Assuming that the suspect had no way of knowing that a machete was used unless he committed the crime, the word machete or an image of the machete would serve as a useful probe. Assuming that the suspect has learned the name of the victim and seen the victim’s photo, then the victim’s name and photo would be useful target stimuli. Finally, irrelevant stimuli might include other plausible, but incorrect murder weapons such as a gun or rifle.

Using these three stimuli allows the researcher to see if the subject’s measured brain activity (as measured by the P300-MERMER) to the probe stimuli can be mathematically classified as being sufficiently similar to the brain activity of the target stimuli or the irrelevant stimuli. For a subject who did not encode information about the machete, the machete image will seem as irrelevant as a photo of a rifle. However, for a subject who did encode information about the machete (eg while they were using it to kill the victim), the machete image will be recognized as will other familiar details of the crime.

Using this basic approach, and following a lengthy and detailed set of protocols described in Farwell (2012), Farwell reports that ‘[i]n over 200 cases including all field and laboratory research so far, brain fingerprinting has not produced a single error, neither a false negative nor a false positive. Error rate has been 0%. 100% of determinations have been correct’.85 Moreover, and importantly for courtroom admissibility determinations, Farwell and colleagues have employed the techniques outside the laboratory in partnership with the CIA, FBI, the United States Navy, and law enforcement.86 In four studies, published in 2014, Farwell reported low error rates and high statistical confidence even when teaching subjects countermeasures and using subjects without incentives to cooperate.87 These studies demonstrated that the technique could be used to detect both ‘specific-issue information’ (such as whether the subject has information on the weapon used in the crime) and ‘specific-screening information’ (such as whether the subject has training in a particular field such as bombmaking).88

Farwell’s brain fingerprinting technique has been proffered in several legal cases.89 The most notable of these was an Iowa murder case involving Terry Harrington. On August 4, 1978, Harrington was convicted of first-degree murder.90 The victim, a security guard and retired police captain, was shot late at night in July 1977. Seventeen-year-old Harrington was charged with the murder, and his conviction rested heavily on the testimony of a juvenile named Kevin Hughes.91 Hughes, who had identified three other men as the killer in different statements, said that Harrington was at the scene and

---

83 Id.
84 Id.
85 Id. at 149.
86 Farwell et al., supra note 62.
87 Id.
88 Id.
91 Id at 515.
that Harrington admitted to shooting the security guard. There was ‘minimal’ physical evidence, but Harrington failed in both his appeal, and failed again in a postconviction relief action arguing that Hughes’ testimony was perjured. Harrington was also unsuccessful in a federal habeas corpus relief action filed in 1993.

Harrington did not file again for postconviction relief until 2000. Although the statute of limitations had run, there was an exception for an appeal on ‘a ground of fact or law that could not have been raised within the applicable time period’. An Iowa district court hearing was held to determine whether there was newly available evidence. Harrington presented evidence that (i) three key witnesses, including Hughes, had recanted their testimony; (ii) previously withheld police investigative reports had been discovered that implicated a different suspect; and (iii) new brain fingerprinting test results had been obtained.

The district court considered the brain fingerprinting evidence in a one-day hearing on November 14, 2000. Farwell had administered the test in May 2000, and submitted his report to the court in October 2000. Farwell testified that, based on his analysis, ‘there is a record of … events of that evening [of the murder] stored in Harrington’s brain’, but that brain record ‘matches the alibi; [and] doesn’t match the crime’. Reviewing this report and testimony, the court found that ‘[t]he P-300 effect has been subject to testing and peer review in the scientific community’, and that ‘[t]he consensus in the community of psycho-physiologists is that the P300 effect is valid’. The court admitted the P300 analysis, but not the MERMER analysis, finding that the MERMER analysis had not been sufficiently peer reviewed. The district court did not, however, grant a new trial based on this and the other evidence. The Iowa Supreme Court reversed, holding that ‘Harrington’s due process right to a fair trial was violated by the State’s failure to produce the police reports …’. But the Iowa Supreme Court decided on grounds other than the brain evidence.

B. Complex trial protocol

Since publishing an article in 1987 on the use of EEG for detecting concealed information, psychologist Peter Rosenfeld and colleagues have published extensively in

92 Id.
93 State v. Harrington, 284 N.W.2d 244 (Iowa 1979).
95 Harrington v. Nix, 983 F.2d 872 (8th Cir. 1993).
99 In the District court, when the defense moved to introduce Farwell’s report into evidence, the State objected on Rules 402, 403, and 702 grounds.
100 Harrington v. State, Case No. PCCV 073247 (Iowa District Court for Pottawattamie County, Mar. 5, 2001). Also testifying were Emanuel Donchin and William Iacono, both in support of admitting the scientific evidence.
102 Id.
104 Id. at 516 (‘Because the scientific testing evidence is not necessary to a resolution of this appeal, we give it no further consideration.’).
the field. Rosenfeld et al. have developed a technique—the ‘Complex Trial Protocol’ (CTP)—which has proven resistant to at least some countermeasures. While the protocol differs from Farwell’s, the underlying logic is the same: the brain will react differently to familiar versus unfamiliar stimuli, and those differing brain responses can be measured by researchers in order to discriminate between information that has, and has not, been encoded by the subject.

As with the brain fingerprinting technique, the CTP involves exposure to the same three types of stimuli. After being introduced to the probe and target words/images, the subject sits down in front of a computer monitor. The computer is attached to an EEG skull cap and electrodes, both of which are placed on the subject. The subject has two buttons: one labeled ‘Yes’ and one labeled ‘No’. The computer will randomly present one of three types of stimuli on the screen: (i) probes, i.e. the first set of information the subject was asked to learn; (ii) targets, i.e. the second set of information the subject was asked to learn; and (iii) irrelevants, i.e. words/images that have not been shown to the subject previously by the researchers. Subjects follow researcher instructions for when to use the buttons. Researchers trained in forensic application have modified this basic approach and employ a mock crime paradigm.

Much of the research remains laboratory research, but John Meixner and Peter Rosenfeld published a real-world test using body cameras. In this real-world study, 26 students wore body cameras for 2 days. The researchers used this data to develop probes based on the students’ real-life experiences. Subjects were randomly assigned to either the knowledgeable or non-knowledgeable group, with the non-knowledgeable group assigned probes that did not come from their real-life experiences. Based on the brain wave data, the researchers were able to perfectly discriminate between the knowledgeable and non-knowledgeable subjects.

C. Synthesis and summary
There are debates amongst researchers about the proper techniques to use. For instance, one distinction between methods is that Farwell’s brain fingerprinting technique employs a ‘classification’ CIT, while Rosenfeld’s CTP uses a ‘comparison’
These different approaches have produced different results, and a full Daubert analysis of the techniques would require careful consideration of these differences.

Yet for our purposes in this article, it is sufficient to note that although there are disagreements over the optimal technique, and although there are limits to all of these techniques, there is nonetheless consensus on the following four points. First, both research teams agree that ‘brain activity surely forms a substrate for deception which patient investigation may elucidate’. Second, both research teams are optimistic about the potential for (some form of) detecting concealed information in forensic settings using EEG with ERP. Third, both research teams (as well as others) continue to engage in research in this area. Fourth, in contrast to fMRI-based lie detection, at least one American court has already admitted EEG concealed information evidence. For these reasons, it seems at least plausible that the technology may in the future meet the requirements of Rule 702. If indeed the evidence has probative value, then courts will need to consider whether this probative value is substantially outweighed by unfair prejudice. We turn now to experiments to inform such Rule 403 balancing.

III. EXPERIMENT 1

We designed an experiment similar in form to many of those discussed in part 1. Most on point with our present study is psychologist Michelle West and colleagues (2014) extension of McCabe et al.’s (2011) study, in which West et al. examined the effect of EEG-based deception detection evidence in a mock juror setting. In the West et al. study, 339 college students were randomly assigned to read one of five vignettes: ‘a no-expert control, or one of four conditions that included either a behavioral or a neuroscientific expert witness for the prosecution who testified that the defendant was being deceptive when he claimed ignorance regarding the crime’. West et al.’s study included multiple components: jury instructions, opening statements from the prosecution and defense, summaries of direct and cross-examination of the victim. Their ‘results did not indicate that the admission of neuroscientific deception detection evidence automatically swayed verdicts, or that it was more prejudicial than behavioral evidence’. But one of the suggestions for future research by West and

---

111 Farwell et al., supra note 77.
112 Rosenfeld, supra note 75 at 34.
113 Supra note 102.
114 Both studies were reviewed and approved by the University of Minnesota Institutional Review Board. IRB Code Number: 150569801 (on file with author and the University of Minnesota IRB). The approval date was June 8, 2015, and the Assurance of Compliance number is FWA00000312. Research combining Amazon Mechanical Turk and Qualtrics is now routine in the social sciences See eg Research combining Amazon Mechanical Turk and Qualtrics is now routine in the social sciences. See eg Dino P. Christenson & David M. Glick, Crowdsourcing Panel Studies and Real-Time Experiments in MTurk, 20 THE POLITICAL METHODOLOGIST 27 (2013). Research using Qualtrics-based experiments has been published and presented in a number of academic fields, suggesting that it meets scholarly expectations for quality online, web-based experiments. Recent legal studies relying on Qualtrics experiments include Matthew R. Ginther et al., The Language of Mens Rea, 67 VAND. L. REV. 1327 (2014); Elizabeth Ingriselli, Mitigating Jurors’ Racial Biases: The Effects of Content and Timing of Jury Instructions, 124 YALE L.J. 1690 (2015); and Jeff Sovern et al., ‘Whimsy Little Contracts’ With Unexpected Consequences: An Empirical Analysis of Consumer Understanding of Arbitration Agreements, 75 MD. L. REV. 1 (2015).
115 Id. at 141.
116 Id. at 142.
117 Id. at 143.
colleagues was to manipulate evidence quality.\textsuperscript{118} In their study, every condition contained the same, weak evidentiary strength.\textsuperscript{119} In our studies, we manipulated strength of the case.

A. Methods

1. Participants

Subjects were recruited to participate, via Amazon Mechanical Turk, in a web-based experiment hosted on the online platform Qualtrics. All subjects received modest payments of 50 cents made available through Amazon Mechanical Turk’s payment service. No personally identifying information was collected. Studies assessing the quality of Turk subjects have found participants to be engaged in the online experimental stimuli, and to be significantly more representative than the convenience samples that would otherwise be used.\textsuperscript{120} Research by Berinsky, Huber and Lenz (2012) demonstrates that replicating experiments on samples recruited in this way yields very similar results to previously published studies with nationally representative samples.\textsuperscript{121}

Concerns about subjects’ compliance with task instructions are of special concern with online experiments because subjects cannot be monitored while engaged in the experimental tasks. To address this issue, psychologists have developed ‘attention filters’ designed to ascertain whether subjects are following instructions and paying attention to the material being presented to them online. We employed a modified version of the filter developed by psychologist Daniel Oppenheimer and his colleagues.\textsuperscript{122} Specifically, users who did not read carefully would only read, in large font, a headline reading ‘Background Questions on Sources for News’ as well as another large, bold question: ‘From which of these sources have you received information in the past month?’ A series of check-box options were provided (eg local newspaper, local TV news). Subjects reading carefully, however, were instructed, in a smaller font, \textit{not} to check any of the boxes, but instead to type ‘123’ into the text box provided.

\textsuperscript{118} Id. at 142.

\textsuperscript{119} Id. at 142. (‘Only weak, circumstantial evidence connected the defendant to the crime: The victim’s neighbor said the defendant had a similar build to the person she saw fleeing the scene, the victim’s wallet and a black ski mask were found in the trash can outside the defendant’s building, and the defendant tried to evade police when they initially came to question him.’)

\textsuperscript{120} Multiple studies have validated results using Amazon’s Mechanical Turk on a variety of assessments, especially when compared to samples of convenience. See eg Tara S. Behrend et al., \textit{The Viability of Crowdsourcing for Survey Research}, 43 BEHAV. METHODS 800 (2011); Michael D. Buhrmester et al., \textit{Amazon’s Mechanical Turk: A New Source of Inexpensive, yet High-Quality, Data?} 6 PERSP. ON PSYCH. SCI. 3 (2011); Joseph K. Goodman et al., \textit{Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples} 26 J. BEHAV. DECISION MAKING 213 (2012); Daniel J. Hauser & Norbert Schwarz, \textit{Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks Than Do Subject Pool Participants}, 48 BEHAV. RES. METHODS 400 (2016); Gabriele Paolacci & Jesse Chandler, \textit{Inside the Turk: Understanding Mechanical Turk as A Participant Pool}, 23 CURR. DIRECTIONS PSYCHOL. SCI. 184 (2014); Jon Sprouse, \textit{A Validation of Amazon Mechanical Turk for the Collection of Acceptability Judgments in Linguistic Theory}, 43 BEHAV. RES. 155 (2011).

\textsuperscript{121} Adam J. Berinsky et al., \textit{Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk}, 20 POL. ANALYSIS 351 (2012).

\textsuperscript{122} See Daniel M. Oppenheimer et al., \textit{Instructional Manipulation Checks: Detecting Satisficing To Increase Statistical Power}, 45 J. EXP. SOC. PSYCHOL. 867, 868 (2009) (describing a filter in which subjects must carefully read instructions which, counter to the boldface headline above the instructions, tell subjects not to actually click on an answer to the question).
Nine hundred and fifty-one subjects completed the experiment, but only 868 ‘attentive’ subjects passed the attention filter, and supplied full demographic information. The results presented in this article are based only on the ‘attentive’ subjects, i.e., those subjects who were paying attention and correctly responded to the attention filter. Demographics were collected at the end of the experiment. Summary demographics for the participants are provided in Appendix Table A1. Compared to national averages, our sample was skewed right on education (with 50% holding either a college or graduate degree); skewed left on income (with 54% reporting less than $40k in annual income); 57% female; 54% under age 34; and 79% white. While not fully nationally representative, the sample provides a mix of demographic and geographic backgrounds (with at least one subject hailing from each of the 50 states).

2. Experimental design

The first experiment was a 3 (strength of case: strong, medium, weak) × 5 (expert evidence: none; polygraph-information present; polygraph-information not present; P300 information present; P300 information not present) between-subjects design. These 15 cells are summarized in Figure 1a. Two short, fictional fact patterns were employed: (i) a ‘stolen diamond necklace’ fact pattern, in which an employee is alleged to have stolen a blue diamond necklace from the company safe, and (ii) an ‘insider trading’ fact pattern, in which an individual is alleged to have engaged in insider trading based on viewing confidential memos. These fact patterns were chosen because they present the type of fact pattern in which brain-based memory recognition evidence would be likely to be of forensic use.

After providing informed consent, subjects were randomly assigned to the first of two short, written scenarios. As described above, there were 15 cells for each of these two fact patterns, and presentation order was counterbalanced. For each fact pattern, subjects were randomly assigned to one (and only one) of the 15 possible scenarios within the fact pattern. After reading and answering questions about one of the 15 scenarios from the first fact pattern, subjects were then presented with one of the 15 scenarios from the second fact pattern.

3. Scenario materials and outcome measure

The stolen diamond scenarios were constructed as follows. First, all subjects read the same introductory text:

The owner of a small hardware store has an office in the back of the store that contains a safe. The owner uses the office safe to store his valuables. One of the items stored in the safe is a very unique looking, blue diamond necklace of great value. No one other than the owner has ever seen the necklace because it is in a sealed box in the back of the safe. The owner checks

123 By way of comparison, this attention rate of 91% is better than the 80% rate in Douglas L. Kriner & Francis X. Shen, How Citizens Respond to Combat Casualties The Differential Impact of Local Casualties on Support for the War in Afghanistan, 76 PUB. OPIN. QUARTERLY 761 (2012).
124 The results obtained when re-running the statistical models with the full sample (attentive and non-attentive) do not alter our substantive conclusions.
the safe every night when he puts away the store’s extra cash for the evening. One night, he finds that the blue diamond necklace is missing.

If the necklace was stolen by one of his employees, the owner wants to terminate the employee. But the owner does not want to fire someone unless he is very sure that the accused employee took it.

Figure 1  (a) shows the design of the online experiment used in this study. Each subject read two fact patterns: (1) insider trading and (2) stolen diamond necklace. The order of fact pattern presentation was counterbalanced. Within each fact pattern, each subject was randomly assigned to one of 15 experimental conditions, as described in the text. After reading their assigned scenario, participants then answered a question about whether the protagonist had seen the diamond necklace/trading memo, and rated their confidence in their decision. Figure 1(b) shows the design of the in-person experiment. It was similar to the online experiment, except that it involved only 9 instead of 15 conditions (eliminating the polygraph conditions). At the end of the in-person experiment tasks, subjects were asked, by a member of the research team, to provide an oral explanation for their results.
Second, subjects were randomly assigned to read one of these three variations of strength of the case (strong, medium, weak):

**Strong:** One of the store’s employees, Greg, is a potential suspect. Greg was the only employee in the store the night the necklace went missing, Greg had the keys to the office and was the last one to leave, and Greg has previously been convicted of a theft at another place of employment. Greg’s cell phone text message log from the night of the theft includes a message that reads ‘Big time score’.

**Medium:** One of the store’s employees, Greg, is a potential suspect. Greg worked in the store the day of the theft, and Greg has keys to the office. But surveillance video footage does not show him entering the office. Greg does have a juvenile charge for shoplifting candy, but he has never been fired from a job, and he has been working at the hardware store for 18 years without incident. Greg’s cell phone text message log from the night of the theft shows no unusual messages.

**Weak:** One of the store’s employees, Greg, is a potential suspect. Greg worked in the store the day of the theft, but Greg does not have keys to the back office and surveillance video footage does not show him entering the office. Greg has never had any criminal charges filed against him, has never been fired from a job, and has been working at the hardware store for 18 years without incident. Greg’s cell phone text message log from the night of the theft shows no unusual messages.

Third, subjects were randomly assigned to read one of the following five variations of expert evidence:

**No expert evidence:** Greg denies that he took the necklace and says, ‘I’ve never seen it before in my life’. A search of his apartment turns up nothing.

**Polygraph dishonest:** Greg denies that he took the necklace and says, ‘I’ve never seen it before in my life’. A search of his apartment turns up nothing. Greg takes a polygraph test, which is used in this instance to measure indicators such as heart rate and blood pressure to assess whether or not Greg recognizes the necklace he claims never to have seen. The results of the polygraph suggest that Greg is not being honest when he says that he has never seen the necklace before in his life.

**Polygraph honest:** Greg denies that he took the necklace and says, ‘I’ve never seen it before in my life’. A search of his apartment turns up nothing. Greg takes a polygraph test, which measures indicators such as heart rate and blood pressure to assess whether or not Greg recognizes the necklace he claims never to have seen. The results of the polygraph suggest that Greg is...
being honest when he says that he has never seen the necklace before in his life.

*EEG P300 with information present:* Greg denies that he took the necklace and says, ‘I’ve never seen it before in my life’. A search of his apartment turns up nothing. Greg takes a brain-based memory recognition test, which uses a particular electrical signature (the ‘P300 wave’) in Greg’s brain to assess whether or not Greg’s brain recognizes the necklace he claims never to have seen. The results of this P300 wave test suggest that Greg is not being honest when he says that he has never seen the necklace before in his life.

*EEG P300 with information not present:* Greg denies that he took the necklace and says, ‘I’ve never seen it before in my life’. A search of his apartment turns up nothing. Greg takes a brain-based memory recognition test, which uses a particular electrical signature (the “P300 wave”) in Greg’s brain to assess whether or not Greg’s brain recognizes the necklace he claims never to have seen. The results of this P300 wave test suggest that Greg is being honest when he says that he has never seen the necklace before in his life.

**Outcome variables:** Fourth and finally, all subjects had to answer the question ‘Did Greg steal the necklace?’ with either a ‘Yes’ or ‘No’ response. Subjects were required to provide their confidence, 0%-100% in the statement that Greg stole the necklace.

An identical approach was used for the *insider trading scenario*. First, all subjects read the same introductory text:

Government prosecutors have learned that the Insight Investment Firm has recently made a very substantial amount of money by trading stocks of a start-up automobile company called Starbright. After an extended investigation, the government concludes that a manager at Insight Investment must have gained early access to a confidential Starbright memo talking about the release of a new product. The memo included a photo of the new product, which was a unique orange and red electric car with double the battery strength of the leading competitor.

Second, subjects were randomly assigned to read one of these three variations of strength of case (strong, medium, weak):

*Strong:* One of the managers at Insight Investment named Mark is charged by the government with insider trading. A person is guilty of committing insider trading if they use private, nonpublic information to trade stock in a
company for profit. Mark had previously been charged with insider trading when he worked at a different firm, and Mark had emailed with Starbright executives the day before the big trade. Mark had been facing some financial troubles at home prior to the trade, but records indicate that he bought a new car just a few hours before the big trade went through.

*Medium:* One of the managers at Insight Investment named Mark is charged by the government with insider trading. A person is guilty of committing insider trading if they use private, nonpublic information to trade stock in a company for profit. Mark has never been previously charged with insider trading, and Mark has a track record of making good guesses about which start-up stocks will rise. But Mark had emailed with Starbright executives the day before the big trade, and was also facing some financial troubles at home.

*Weak:* One of the managers at Insight Investment named Mark is charged by the government with insider trading. A person is guilty of committing insider trading if they use private, nonpublic information to trade stock in a company for profit. Mark has never been previously charged with insider trading, and Mark has a track record of making good guesses about which start-up stocks will rise. There was no direct evidence that Mark had ever spoken with or emailed with Starbright executives. Mark was not facing any financial troubles at home, since he had recently inherited a large sum of money from his deceased father.

Third, subjects were randomly assigned to read one of the following five variations with regard to expert evidence:

*No expert evidence:* Mark denies that he ever saw the memo from Starbright and says, ‘I never saw it before, and I didn’t engage in insider trading’.

*Polygraph not honest:* Mark denies that he ever saw the memo from Starbright and says, ‘I never saw it before, and I didn’t engage in insider trading’. Mark takes a polygraph test, which is used in this instance to measure indicators such as heart rate and blood pressure to assess whether or not Mark recognizes the memo he claims never to have seen. The results of the polygraph suggest that Mark is not being honest when he says that he has never seen the memo.

*Polygraph honest:* Mark denies that he ever saw the memo from Starbright and says, ‘I never saw it before, and I didn’t engage in insider trading’. Mark takes a polygraph test, which is used in this instance to measure indicators such as heart rate and blood pressure to assess whether or not Mark recognizes the memo he claims never to have seen. The results of the polygraph
suggest that Mark is being honest when he says that he has never seen the memo.

EEG P300 with information present: Mark denies that he ever saw the memo from Starbright and says, ‘I never saw it before, and I didn’t engage in insider trading’. Mark takes a brain-based memory recognition test, which uses a particular electrical signature (the ‘P300 wave’) in Mark’s brain to assess whether or not Mark’s brain recognizes the memo he claims never to have seen. The results of this P300 wave test suggest that Mark is not being honest when he says that he has never seen the memo.

EEG P300 with information not present: Mark denies that he ever saw the memo from Starbright and says, ‘I never saw it before, and I didn’t engage in insider trading’. Mark takes a brain-based memory recognition test, which uses a particular electrical signature (the ‘P300 wave’) in Mark’s brain to assess whether or not Mark’s brain recognizes the necklace he claims never to have seen. The results of this P300 wave test suggest that Mark is being honest when he says that he has never seen the memo.

Outcome variables – Assessment of Guilt and Reported Confidence in that Assessment: Fourth and finally, all subjects had to answer the question ‘Did Mark see the memo?’ with either a ‘Yes’ or ‘No’ response. Subjects were also required to provide their confidence, 0%-100% in the statement that Mark saw the memo.

B. Analysis and results

What factors make it more (or less) likely that our subjects chose ‘Yes’ in response to the questions, ‘Did Greg steal the necklace?’ and ‘Did Mark see the memo?’ Because our outcome variable is dichotomous (Yes/No), we employed logistic regression to examine the effects of expert evidence, strength of the case, and subject demographics (Table 1, column 1). To estimate an average effect over both fact patterns, we ran a single regression with responses to both fact patterns included in the analysis. We added a dummy variable to control for vignette-specific effects. To account for possible order effects, we included a dichotomous variable measuring whether the subject saw the diamond fact pattern before the insider trader fact pattern. Because our subjects gave us repeated measures, we also employed robust standard errors clustered on subject.

The results find a significant and very large effect of strength of the case (Table 1). Subjects presented with a strong case against the defendant were more likely to respond that he had seen the object in question. The results also find that—after controlling for strength of case as well as demographics—there is a significant effect of the neuroscientific evidence. Subjects were more likely to answer yes when exposed to the EEG

125 In the stolen diamond necklace fact pattern, 0% of the subjects responded that Greg stole the necklace in our ‘medium’ strength of case condition without expert evidence. Post hoc, it is clear that we did not sufficiently delineate between the weak and medium fact patterns. Thus, for purposes of analysis we collapsed the weak strength of case and medium strength of case conditions into a single ‘weak’ strength of case variable.

126 \( \chi^2(1) = 283.43, p < 0.001. \)
Table 1: Experiment 1 (online subjects): effect of strength of case, EEG and polygraph treatments on outcome measures

<table>
<thead>
<tr>
<th></th>
<th>Outcome variable: Did protagonist see/steal the object? (Yes = 1/No = 0)</th>
<th>Outcome variable: Did protagonist NOT see/steal the object? (Yes = 0/No = 1)</th>
<th>Outcome variable: confidence rating (0–100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B odds ratio</td>
<td>B odds ratio</td>
<td>OLS Reg. Coeff.</td>
</tr>
<tr>
<td>Decided that D saw object</td>
<td>–</td>
<td>–</td>
<td>29.73** (1.508)</td>
</tr>
<tr>
<td>Strong case against D</td>
<td>14.33** (2.267)</td>
<td>0.0698** (0.0110)</td>
<td>6.195** (1.322)</td>
</tr>
<tr>
<td>EEG information present</td>
<td>2.655** (0.467)</td>
<td>0.377** (0.0662)</td>
<td>5.199** (1.938)</td>
</tr>
<tr>
<td>EEG information not present</td>
<td>0.333** (0.0706)</td>
<td>3.004** (0.637)</td>
<td>–0.117 (2.276)</td>
</tr>
<tr>
<td>Polygraph dishonest</td>
<td>2.480** (0.439)</td>
<td>0.403** (0.0714)</td>
<td>4.270* (1.967)</td>
</tr>
<tr>
<td>Polygraph honest</td>
<td>0.282** (0.0594)</td>
<td>3.551** (0.749)</td>
<td>1.448 (2.260)</td>
</tr>
<tr>
<td>Subject age</td>
<td>1.002 (0.00486)</td>
<td>0.998 (0.00485)</td>
<td>–0.0273 (0.0624)</td>
</tr>
<tr>
<td>Subject college grad</td>
<td>1.177 (0.144)</td>
<td>0.850 (0.104)</td>
<td>–1.775 (1.513)</td>
</tr>
<tr>
<td>Subject conservative ideology</td>
<td>1.083* (0.0408)</td>
<td>0.924* (0.0348)</td>
<td>0.345 (0.467)</td>
</tr>
<tr>
<td>Subject non-white</td>
<td>1.263 (0.205)</td>
<td>0.792 (0.129)</td>
<td>2.355 (1.861)</td>
</tr>
<tr>
<td>Diamond necklace scenario</td>
<td>0.224** (0.0292)</td>
<td>4.463** (0.582)</td>
<td>–3.183** (1.129)</td>
</tr>
<tr>
<td>Read diamond necklace first</td>
<td>0.934 (0.115)</td>
<td>1.070 (0.131)</td>
<td>–0.928 (1.508)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.369** (0.0982)</td>
<td>2.709** (0.721)</td>
<td>37.32** (3.354)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,736</td>
<td>1,736</td>
<td>1,734</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.3083</td>
<td>0.3803</td>
<td>0.288</td>
</tr>
</tbody>
</table>

The dependent variable for models reported in columns 1 and 2 is a dichotomous (0,1) variable indicating whether the protagonist in the scenario was judged to have seen/stolen the object in question. The dependent variable in the model reported in column 3 is a continuous variable (spanning 0–100) measuring the protagonist’s self-reported confidence in his/her assessment. All models employed clustered robust standard errors. Two-tailed statistical significance is denoted as follows: **p < .01, *p < .05.

*The wording was different in the two fact patterns. For the insider trader fact pattern, subjects were asked: ‘Did Mark see the memo?’ For the diamond necklace fact pattern, subjects were asked: ‘Did Greg steal the necklace?’

bWe flip the dependent variable in column 2 in order to more intuitively interpret the odds ratio for the EEG information not present condition.

P300 condition with information present. By contrast, subjects were more likely to answer no when exposed to the EEG P300 condition without information present. Exposure to the polygraph evidence was also a significant explanatory factor. Subjects were more likely to answer yes when exposed to the polygraph dishonest condition.
Figure 2: Experiment 1 (online subjects): predicted change in likelihood of subject deciding that the protagonist saw the object. The graph plots the change in predicted probability associated with each independent variable changing from 0 to 1. Predictions are based on the logistic regression model as described in the text, with results reported in Table 1. All other variables are held constant at their mean or medians. The dots are a point estimate, and the bars indicate 95% confidence intervals.

But subjects were more likely to say no when exposed to the polygraph honest condition.\footnote{\chi^2(1) = 36.14, \( p < 0.001 \).}

In Figure 2 we graph the estimated change in predicted probability for selecting yes, if the independent variable (e.g., strength of case) changes from 0 to 1, with all other variables in the model held constant at their means or medians. As seen graphically, we find that neuroscientific evidence affects legal decision-making, but that the magnitude of the effect is not nearly as large as the overall strength of the case against the defendant. A change from a non-strong to a strong case increases the probability of a yes response by .58, while introducing EEG evidence suggesting he saw it increases the probability only .23. This is roughly the same effect as introducing a dishonest polygraph, which raises the probability .22.

We also analysed subjects’ self-reported confidence in their decision. The outcome variable is a continuous variable, which can take integer values from 0 to 100. We thus employed ordinary least squares (OLS) regression. We specified the model as discussed above, only we added one additional explanatory variable: whether the subject determined that the defendant had seen/stole the item. The results (Table 1, column 3) of the analysis find that confidence is significantly affected by whether or not
the subject thought the defendant saw/stole the item,\textsuperscript{131} by the strength of the case,\textsuperscript{132} by an EEG P300 result suggesting that the defendant possessed the information,\textsuperscript{133} and by a polygraph result suggesting dishonesty.\textsuperscript{134} In sum, we find that neuroscientific evidence may affect both legal outcomes and confidence in those judgements.

**IV. EXPERIMENT 2**

While online experiments have the virtue of allowing researchers to recruit large numbers of subjects at a low cost, it is important to check the robustness of such results with similar in-person experiments. To accomplish this, we ran an in-person version of experiment 1 at the 2015 Minnesota State Fair.

**A. Methods**

The University of Minnesota maintains a dedicated research building (the ‘Driven to Discover’ or D2D building), allowing researchers to directly interact with some of the nearly 2 million Fairgoers who attend the 2-week Minnesota State Fair.\textsuperscript{135} Our lab participated in 2015 to carry out this experiment.

We worked a total of 24 hours at the Fair, in four separate 6-hour shifts. Across this period, 611 adult subjects (18+ years old) were recruited in person outside the research building. Subjects received a free tote bag for their participation, which on average took about 10 minutes. Subjects used either an iPad to complete the experiment via Qualtrics, or a pen-and-paper hard copy version. Subject demographics are reported in Table A1. The State Fair subjects were skewed right on education (59\% with a college or graduate degree); 64\% female, and 90\% white. Compared to our online sample, the State Fair subjects were older.

Because of uncertainty about the number of subjects who would participate in person and the associated concerns about requisite statistical power, we reduced the number of experimental conditions from 15 to 9 by eliminating the ‘polygraph honest’ and ‘polygraph dishonest’ conditions. We eliminated these two conditions because our primary focus was on the effect of the EEG memory recognition evidence. Other than these changes, the State Fair experimental task was identical to the online study task.\textsuperscript{136}

\textsuperscript{131} (F(1, 867) = 388.61, p < 0.01).

\textsuperscript{132} (F(1, 867) = 21.98, p < 0.01).

\textsuperscript{133} (F(1, 867) = 7.19, p < 0.01).

\textsuperscript{134} (F(1, 867) = 4.71, p < 0.01).

\textsuperscript{135} The Minnesota State Fair is the biggest per-capita state fair in the country. It is called the ‘Great Minnesota Get Together’ for good reason.

\textsuperscript{136} To better understand how subjects assessed these various types of evidence, we collected qualitative data from our in-person State Fair sample of subjects. At the completion of the experiment, every subject was prompted by a researcher to provide a brief, oral explanation for why they chose guilty or not guilty. Researchers used varieties of phrases such as, ‘Can you please explain your decision today?’ or ‘What did you decide and why?’ Researchers did not prompt particular evidence. For instance, we did not directly ask whether the EEG evidence affected decisions. Responses were recorded from 563 subjects. All responses were transcribed, and then coded by the research team to determine which pieces of evidence were mentioned most frequently. Coding was carried out by a team of four undergraduate students at the University of Minnesota. A subsample of 100 subjects were used to establish intercoder reliability, which was over 85\% across the four coders. But analysis of the qualitative responses found no significant, noteworthy patterns. Data available on request from authors.
Table 2: Experiment 2 (in-person subjects): effect of strength of case and EEG treatments on outcome measures

<table>
<thead>
<tr>
<th>Outcome variable:</th>
<th>Did protagonist see/steal the object?</th>
<th>Did protagonist NOT see/steal the object</th>
<th>Confidence rating (0–100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Yes = 1/No = 0)</td>
<td>B odds ratio</td>
<td>B odds ratio</td>
<td>OLS Reg. Coeff.</td>
</tr>
<tr>
<td>Decided that D saw object</td>
<td>–</td>
<td>–</td>
<td>31.63**</td>
</tr>
<tr>
<td>Strong case against D</td>
<td>4.019** (0.578)</td>
<td>0.249** (0.0358)</td>
<td>2.437</td>
</tr>
<tr>
<td>EEG information present</td>
<td>1.763** (0.279)</td>
<td>0.567** (0.0899)</td>
<td>–1.496</td>
</tr>
<tr>
<td>EEG information not present</td>
<td>0.405** (0.0703)</td>
<td>2.467** (0.428)</td>
<td>–1.070</td>
</tr>
<tr>
<td>Subject age</td>
<td>0.995 (0.00434)</td>
<td>1.005 (0.00438)</td>
<td>–0.0835</td>
</tr>
<tr>
<td>Subject college grad</td>
<td>0.653** (0.0890)</td>
<td>1.532** (0.209)</td>
<td>–2.651</td>
</tr>
<tr>
<td>Subject conservative ideology</td>
<td>1.099* (0.0470)</td>
<td>0.910* (0.0389)</td>
<td>0.653</td>
</tr>
<tr>
<td>Subject non-white</td>
<td>1.157 (0.278)</td>
<td>0.865 (0.208)</td>
<td>4.008</td>
</tr>
<tr>
<td>Diamond necklace scenario</td>
<td>0.166** (0.0245)</td>
<td>6.033** (0.890)</td>
<td>–4.822**</td>
</tr>
<tr>
<td>Read diamond necklace first</td>
<td>1.198 (0.160)</td>
<td>0.835 (0.112)</td>
<td>1.650</td>
</tr>
<tr>
<td>Constant</td>
<td>0.990 (0.294)</td>
<td>1.010 (0.300)</td>
<td>38.54**</td>
</tr>
<tr>
<td>Observations</td>
<td>1,222</td>
<td>1,222</td>
<td>1,206</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.2108</td>
<td>0.2108</td>
<td>0.297</td>
</tr>
</tbody>
</table>

The dependent variable for models reported in columns 1 and 2 is a dichotomous (0,1) variable indicating whether the protagonist in the scenario was judged to have seen/stolen the object in question. The dependent variable in the model reported in column 3 is a continuous variable (spanning 0–100) measuring the protagonist’s self-reported confidence in his/her assessment. All models employed clustered robust standard errors. Two-tailed statistical significance is denoted as follows: **p < .01, *p < .05.

The wording was different in the two fact patterns. For the insider trader fact pattern, subjects were asked: ‘Did Mark see the memo?’ For the diamond necklace fact pattern, subjects were asked: ‘Did Greg steal the necklace?’

We flip the dependent variable in column 2 in order to more intuitively interpret the odds ratio for the EEG information not present condition.

B. Results

We conducted the same type of statistical analysis as we did for experiment 1, only we did not include the polygraph variables as predictors. The results find a significant effect of strength of the case (Table 2). Subjects presented with a strong case against the defendant were more likely to respond that he had seen the object in question. The results once again find that—after controlling for strength of case as well as demographics—there is a significant effect of the neuroscientific evidence. Subjects were more likely to answer yes when exposed to the EEG P300 condition with

---

137 In the stolen diamond necklace fact pattern, 0% of the subjects responded that Greg stole the necklace in our ‘medium’ strength of case condition without expert evidence. Post hoc, it is clear that we did not sufficiently delineate between the weak and medium fact patterns. Thus, for purposes of analysis we collapsed the weak strength of case and medium strength of case conditions into a single ‘weak’ strength of case variable.

138 $\chi^2(1) = 93.48, p < .001$. 

---
Figure 3: Experiment 2 (in-person subjects): predicted change in likelihood of subject deciding that the protagonist saw/stole the object. The graph plots the change in predicted probability associated with each independent variable changing from 0 to 1. Predictions are based on the logistic regression model as described in the text, with results reported in Table 2. All other variables are held constant at their mean or medians. The dots are a point estimate, and the bars indicate 95% confidence intervals.

By contrast, subjects were more likely to answer no when exposed to the EEG P300 condition without information present.  

In Figure 3, we graph the estimated change in predicted probability for selecting yes, if the independent variable (e.g., strength of case) changes from 0 to 1, with all other variables in the model held constant at their means or medians. We again find that neuroscientific evidence affects legal decision making, but that the magnitude of the effect is not as large as the overall strength of the case against the defendant. Moving from a non-strong case to a strong case increases the likelihood of a yes response by .32, while the effect of an EEG P300 finding information in the brain increases the likelihood by only .13. An EEG P300 finding no memory recognition reduces the likelihood by .19.

As with the online data, we also analysed the in-person subjects’ self-reported confidence in their decision. We specified the model as we did for experiment 1. The results (Table 2, column 3) of the analysis find that confidence is significantly affected by whether or not the subject thought the defendant saw/stole the item, but none of the other factors that were significant in experiment 1 (strength of the case, and EEG P300 information) were significant in experiment 2.

---

139 \( \chi^2(1) = 12.80, p < 0.001. \)

140 \( \chi^2(1) = 27.09, p < 0.001. \)

141 \( F(1, 609) = 272.94, p < 0.01. \)
V. DISCUSSION

This study explored the contributions of strength of case and EEG memory recognition on lay subjects’ determinations of whether fictional defendants had seen legally salient items. In two experiments, one online and one in-person, we found evidence that both strength of the case and the neuroscientific evidence affected outcomes. We also found in both experiments that the magnitude of the strength of case effect was greater than the neuroscience effect. Findings were not consistent with regard to the relationship between neuroscientific evidence and self-reported confidence in outcomes.

As reviewed in part I, the experimental literature examining the effects of neuroscience information on legal outcomes remains mixed. To this literature, our results add several new insights. Our evidence suggests that the effect of neuroscientific evidence is not nearly enough to outweigh the overall strength of the case. That is, consistent with much of the recent scholarship, our findings suggest that there is not a seductive allure to neuroscientific evidence. At the same time, just because neuroscience does not have a large effect, it does not follow that neuroscience will never be of use to the court. For instance, as Shen (2016) has argued, neuroscience may play a role in courts akin to how instant replay is used in professional sports. In a few, choice trials, where the evidence on both sides seems to be even, perhaps neuroscience could tip the scales.

For the research literature, we also present in this study a new way to approach the question of when there is ‘seduction’, or in the language of Rule 403, when the probative value is ‘substantially outweighed by ... unfair prejudice ...’. To date, most studies have followed Weisberg (2008) by designing the experiment such that the neuroscience information is superfluous. As described by Schweitzer et al. (2011), ‘the best way of addressing our research question is to remove the objective value of the neuroimage completely’. In these designs, any influence of the neuroscience in the experiments can be deemed overinfluence. But these approaches leave open the question: when the evidence does have at least some probative value, how can we know if that probative value is substantially outweighed by unfair prejudice? Our approach cannot answer this question, but it can inform it because it allows us to compare the relative influence of the neuroscience vis-à-vis the rest of the evidentiary record. As Schweitzer et al. (2013) have previously noted, ‘in legal contexts a jury is quite likely to be exposed to multiple sources of information and the situation would thus be more akin to a relative judgment task than to an absolute one’. In short, the effect of neuroscience should be understood as its effect relative to the rest of the evidentiary record.

If our study can be generalized to neuroscientific evidence generally, the implication is that neuroscience is not as substantial a factor as strength of the case. This may seem obvious, since neuroscience is just one piece of information in the array of an evidentiary record. But it is an important point because it suggests that neuroscience is unlikely to be an outcome-determinative silver bullet. We agree with Mowle et al. (2016), whose ‘results are generally consistent with recent studies suggesting that fears of neuroscience evidence being overly influential with jurors may be somewhat overblown’. Still, it is worth emphasizing that we do find an effect. And our study’s finding that neuroscience can affect outcomes, even when not presenting a brain image as part of the

142 Shen, supra note 1.
A clear story to emerge from our study is that neuroscience does not overcome strength of the case. But our results also raise some puzzling questions. For instance, confidence ratings were significantly related to neuroscientific information in the online version, but not in the in-person version. We do not have a satisfactory answer for this difference, as we controlled for subject characteristics such as age and education level. In addition, although it was not the focus of our study, subjects in both experiments were more likely to believe the employee accused of stealing a diamond necklace than they were a stock broker accused of insider trading. We selected the fact pattern subject matter based on the potential applicability of EEG memory recognition technology. Going forward, more careful attention is called for in considering what details in those vignettes may heighten, or dampen, the effect of the brain evidence. In short, and in keeping with the recent wave of studies in this area, it is clear that context matters.

A. Limitations
This study is limited in many ways. Most notably, we do not know the extent to which the experimental results are indicative of how real jurors would behave in a real trial setting. In our short vignettes, jurors read about the evidence. In an actual trial, jurors would be presented with the evidence through an expert witness, guided by a skilled attorney. They would also hear testimony from a competing expert. Moreover, jurors would deliberate, and decide collectively, about the fate of the defendant. Additional differences are readily apparent between subjects participating in a 10-minute task and jurors participating in a trial stretching multiple days. Such differences might well affect the way in which the memory recognition evidence is received and processed. To better understand the reasons for these differences, future research is warranted to more closely mimic the presentation of evidence in a courtroom setting.

Researchers should pay careful attention to exactly how such information would be presented in a trial setting. We focused on just one type of neuroscience. And our treatment was blunt: we told subjects in writing that a particular result had been obtained. We did not present subjects with any of the scientific debate we highlighted in part II, and we do not know how subjects would respond to testimony parsing the details of study design. For instance, one difference between the methods described in part II is the difference between a ‘classification’ CIT and a ‘comparison’ CIT. In real life, competing experts might disagree about such methods. But would a jury be able to effectively process such technical differences? In answering this question, the literature on neuroscientific evidence and the law may benefit from revisiting the more extensive literature on scientific evidence and jury decision making.

Our results also do not extend to all Rule 403 prongs, as Rule 403 invokes considerations not only of unfair prejudice, but also juror confusion, waste of time, misleading the jury, and presenting cumulative evidence. An attorney could argue, for instance, that if there is little probative value from the evidence, then the time delay for multiple experts

---

144 Discussed in Farwell et al., supra note 62.
and cross-examination is a waste of the court’s time. The scientific details, which we have not delved into in this article, might also be unduly confusing to jurors. Nothing in our study addresses these concerns. Moreover, the court in a criminal case may need to be particularly cognizant of any potential unfair prejudice. Criminal defendants might require more protection than a civil litigant due to the harsher consequences faced in a criminal trial. Civil litigants typically risk money damages, while criminal litigants are facing a loss of liberty.

B. Rule 403, the polygraph, and EEG memory recognition evidence
Although the limitations just described limit the real-world inferences we can make from the experiments, the results at least suggest a path forward with regard to Rule 403 and EEG memory recognition evidence. Because of the inevitable comparisons that will be made between EEG memory recognition evidence and the polygraph, analyzing the admissibility of memory recognition evidence under Rule 403 can be meaningfully informed by examining how courts apply Rule 403 to polygraph.

For many decades, since the influential Frye decision in 1923, American courts have generally excluded polygraph evidence.146 The leading U.S. Supreme Court case on the issue is U.S. v. Scheffer, a 1998 case in which a military defendant challenged the constitutionality of the per se ban on the polygraph contained in the Military Rules of Evidence. The court, with Justice Thomas writing the opinion, ruled that the military was justified in this per se exclusion.147

Why are courts adamant about excluding the polygraph? The most frequent reason announced by courts is a failure of the polygraph to meet the requirements of Rule 702. But courts also routinely point to Rule 403 concerns. An Eighth Circuit case illustrates this concern:148

Aside from doubts about the reliability of polygraphy at this stage of the art, there are some countervailing policy considerations which militate against the admissibility of unstipulated polygraph evidence at trial. A foremost concern is the effect that such evidence

146 Frye v. United States, 293 F. 1013 (D.C. Cir. 1923); 31A C.J.S. EVIDENCE § 317; 5 MOD. SCI. EVIDENCE § 40:2 (2014–2015 ed.); Robert L. Arendell & Stephen C. Peters, Revisiting the Admissibility of Polygraph Evidence After Daubert, COLO. L., Feb. 1996, at 65; Greely & Illes, supra note 85, at 385. Exceptions are sometimes made when both parties stipulate to admissibility. See James L. Buchwalter, Stipulation to Evidence, 83 C.J.S. STIPULATIONS § 71. (‘According to some authorities, where both parties stipulate to the admission of polygraph evidence, it may be admitted. According to other authorities, however, a stipulated polygraph test, like any other polygraph test, is not admissible into evidence at trial. According to still other authorities, stipulating to the use of a polygraph examination waives any evidentiary objection to admissibility grounded in scientific unreliability, but does not preclude counsel from adducing evidence on the issue of scientific unreliability and commenting on it during argument.’)

147 United States v. Scheffer, 523 U.S. 303, 310–11 (1998). Summarizing the state of the law, Thomas wrote: ‘... Although some Federal Courts of Appeals have abandoned the per se rule excluding polygraph evidence, leaving its admission or exclusion to the discretion of district courts under Daubert, at least one Federal Circuit has recently reaffirmed its per se ban, and another recently noted that it has ‘not decided whether polygraphy has reached a sufficient state of reliability to be admissible’. Most States maintain per se rules excluding polygraph evidence.

148 United States v. Alexander, 526 F.2d 161, 164 (8th Cir. 1975). First, the court discusses the inadequacy of the polygraph science: in applying the scientific acceptability standard to polygraph tests, all United States courts of appeals addressing the issue have excluded the results of unstipulated polygraph tests. These courts reason that the polygraph does not command scientific acceptability and that it is not generally believed to be sufficiently reliable in ascertaining truth and deception to justify its utilization in the trial process.
will have on juries provided that the polygraph attains the degree of scientific acceptability and reliability which will allow its admission in evidence. 149

The court went on to observe that ‘[w]hen polygraph evidence is offered in evidence at trial, it is likely to be shrouded with an aura of near infallibility, akin to the ancient oracle of Delphi’. 150

Yet the experimental literature on the polygraph suggests that the polygraph would not likely function as such an oracle. A small body of psychology research has explored the effect of the polygraph on mock juror decision-making. 151 Reviews have concluded that ‘the research on polygraph evidence suggests it has relatively little impact on jurors’. 152 There is thus a disconnect between the experimental research consensus and the old conventional wisdom in courts that the polygraph would be unduly persuasive.

A similar disconnect may soon emerge in the context of neuroscientific evidence. The growing scholarship in this area suggests that there may be no effect at all, and if there is one, it is not unduly persuasive. The history of polygraph admissibility presents an important lesson for those who wish to explore expanded use of EEG memory recognition evidence in courtrooms: simply amassing experimental evidence is not enough to dissuade Rule 403 concerns.

What is called for is more dialogue between science and law to better inform courts about what neuroscientific evidence (including but not limited to EEG memory detection evidence) is and is not. For instance, all researchers in the field agree that EEG memory recognition testimony is not lie detection evidence. As Farwell has stressed, ‘[b]rain fingerprinting does not detect whether or not the subject committed the crime. It only detects whether or not the subject knows the relevant knowledge contained in the probes’. 153 Memory recognition evidence is simply one additional piece of evidence that may be weighted (more or less) by the fact finder. Given the inevitable, but problematic, analogies that have been made between EEG memory detection and the polygraph, the distinctions should be stressed so that Rule 403 evaluations of memory detection do not simply invoke Rule 403 evaluations of the polygraph.

C. The future forensic use of EEG memory detection

Shen (2016) has argued that the future of neuroscience and law requires balance: ‘imagination, but not too much … excitement, but not over-exuberance. … We must add to

149 Id.
150 Id.
152 Devine & Macken, supra note 145, at 111. Myers et al. (2006), supra note 151. See also Myers et al. (2006), supra note 151, at 521 (citing Ann Cavoukian & Ronald J. Heselgrave, The Admissibility of Polygraph Evidence in Court: Some Empirical Findings, 4 L. & HUM. BEHAV. 117 (1980); Bryan Myers & Jack Arbuthnot, Polygraph Testimony and Juror Judgments: A Comparison of The Guilty Knowledge Test and the Control Question Test, 27 J. APPL. SOC. PSYCHOL. 1421 (1997); Myers et al. (2003), supra note 151; Nicholas P. Spanos et al., The Effects of Polygraph Evidence and Eyewitness Testimony on The Beliefs and Decisions of Mock Jurors, 12 IMAGINATION COGNITION & PERSONALITY 103 (1992)). (Little empirical research in this area supports the notion that jurors are likely to uncritically accept polygraph test evidence. On the contrary, based on a series of studies that have addressed this issue, jurors may be quite skeptical of the polygraph.)
153 Farwell, supra note 77, at 128.
our important, futuristic “What if?” queries an entrepreneurial “What now?” mentality. EEG memory recognition technology is a possible ‘what now’ (or at least in the near future) technology. On a practical level, it is affordable, mobile, and easy to operate. On a research level, we now have three decades of studies from multiple labs. Research has suggested high sensitivity and specificity with P300-GKT approaches.

Additional research will be required before courts are likely to admit the evidence as the Iowa District Court did, but regardless of if/when EEG memory recognition becomes admissible in courtrooms, it may be utilized in legal contexts outside the courtroom. Although the polygraph is generally excluded from courtrooms, the law allows for a multitude of other ways in which the polygraph (and therefore potentially EEG memory recognition) may be used. For instance, police departments and federal authorities routinely use the polygraph to aid their investigations. It seems at least plausible that EEG memory recognition could be utilized in such contexts in the not so distant future.

CONCLUSION

This article presented results from a study of the effect of brain-based EEG memory recognition evidence on legal assessments made by adult American subjects. We found that neuroscientific evidence does affect outcomes, but it has a weaker effect than the strength of the case. The results may have implications for Rule 403 concerns that the probative value of neuroscientific evidence will be substantially outweighed by unfair prejudice.

Momentum is building for the integration of neuroscience and law. There are a growing number of cases with neuroscientific evidence, publications on neurolaw topics, and courses in law and neuroscience. Despite this momentum, however,

155 Allen & Iacono, supra note 74; Allen et al., supra note 74; John J. Allen et al., An Event-Related Potential Investigation of Posthypnotic Recognition Amnesia, 104 J. Abnormal Psychol. 421 (1995); Gershon Ben-Shakhar, Current Research and Potential Applications of the Concealed Information Test: An Overview, 3 Front. Psychol. 342 (2012); Iacono, supra note 74; Donald J. Krapohl, Limitations of the Concealed Information Test In Criminal Cases, in MEMORY DETECTION: THEORY AND APPLICATION OF THE CONCEALED INFORMATION TEST 151 (Bruno Verschuer et al. eds., 2011); Lefebvre et al., supra note 74; Izumi Matsuda et al., supra note 74; John B. Meixner et al., P900: A Putative Novel Erp Component that Indexes Countermeasure Use in the P300-Based Concealed Information Test, 38 Appl. Psychophysiol. & Biofeedback 121 (2013); J. Peter Rosenfeld, P300 in Detecting Concealed Information, in MEMORY DETECTION: THEORY AND APPLICATION OF THE CONCEALED INFORMATION TEST 63 (Bruno Verschuer et al. eds., 2011); Rosenfeld, supra note 75; Rosenfeld et al., supra note 105; Johanna C. van Hooff et al., Event-Related Potentials As Indirect Measures Of Recognition Memory, 21 Int'l. J. Psychophysiol. 15 (1996).
158 Jones et al., supra note 5; Shen, supra note 2.
interdisciplinary empirical work in neurolaw remains underdeveloped.\textsuperscript{160} That is, scientists are pursuing legally relevant studies, and legal scholars are writing commentaries on that science.\textsuperscript{161} But still too rarely are lawyers and scientists in the same room, collaborating on the same article, and jointly interfacing with courts and legal actors.\textsuperscript{162} EEG memory recognition technology is an area ripe for joint investigation.

**ACKNOWLEDGEMENTS**

We thank participants in the William and Mary Law School conference on Neuroscience and Criminal Responsibility for very helpful feedback. We also thank Matthew Ginther, William Iacono, John Meixner, and Tom Pryor for helpful feedback on this project. Ryan Pesch, Geordin Crist, Jordan Krieg, Aubrey Strenger, Rebecca Carlson, and Dara Johnson-Ayodele provided excellent research assistance. We thank Logan Spector, Ellen Demerath, and Amanda Kabage for facilitating use of the University of Minnesota Driver to Discover Building at the Minnesota State Fair. Support for this research was provided by the University of Minnesota.

\textsuperscript{160} Exceptions include work supported by the MacArthur Foundation Research Network on Law and Neuroscience http://lawneuro.org/publications.php (accessed March 1, 2017) and work conducted by neuroscientist Kent Kiehl on inmates.


\textsuperscript{162} Peter Rosenfeld et al. (2013) ‘While we are nearing 25 years of applied and theoretical research regarding the P300-based CIT [a brain-based memory recognition technique], there has been relatively little discussion of the legal relevance of the test, and what little discussion there has been has occurred almost exclusively in the legal literature as opposed to the psychological literature, with little interaction between the two.’
## APPENDIX

### Table A1: Participant demographics

<table>
<thead>
<tr>
<th>Education</th>
<th>Online subjects (%)</th>
<th>In-person subjects (%)</th>
<th>U.S. Census (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than HS</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>High school/GED</td>
<td>11</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Some college</td>
<td>29</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Assoc. degree</td>
<td>11</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>36</td>
<td>36</td>
<td>17</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>14</td>
<td>23</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income</th>
<th>Online subjects (%)</th>
<th>In-person subjects (%)</th>
<th>U.S. Census (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $20,000</td>
<td>28</td>
<td>18</td>
<td>$1–$24,999: 22</td>
</tr>
<tr>
<td>$20,000–$40,000</td>
<td>26</td>
<td>16</td>
<td>$25,000–$34,999: 19</td>
</tr>
<tr>
<td>$40,000–$60,000</td>
<td>22</td>
<td>23</td>
<td>$35,000–$49,999: 21</td>
</tr>
<tr>
<td>$60,000–$80,000</td>
<td>12</td>
<td>14</td>
<td>$50,000–$64,999: 14</td>
</tr>
<tr>
<td>$80,000–$100,000</td>
<td>7</td>
<td>11</td>
<td>$65,000–$74,999: 6</td>
</tr>
<tr>
<td>&gt; $100,000</td>
<td>6</td>
<td>18</td>
<td>$75,000–$99,999: 8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Online subjects (%)</th>
<th>In-person subjects (%)</th>
<th>U.S. Census (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>43</td>
<td>36</td>
<td>49</td>
</tr>
<tr>
<td>Female</td>
<td>57</td>
<td>64</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age groups</th>
<th>Online subjects (%)</th>
<th>In-person subjects (%)</th>
<th>U.S. Census (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–24</td>
<td>19</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>25–34</td>
<td>35</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>35–44</td>
<td>20</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>45–59</td>
<td>18</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>60+</td>
<td>7</td>
<td>18</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Online subjects (%)</th>
<th>In-person subjects (%)</th>
<th>U.S. Census (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>79</td>
<td>90</td>
<td>74</td>
</tr>
<tr>
<td>Non-white</td>
<td>21</td>
<td>10</td>
<td>26</td>
</tr>
</tbody>
</table>