Journal of Law and the Biosciences, 208–211
doi:10.1093/jlb/lsy007
Peer Commentary

# The difficulty of discerning the effect of neuroscience: a peer commentary of Shen et al. 2018

John B. Meixner, Jr

School of Law, Northwestern University, Chicago, Illinois, USA
Corresponding author. E-mail: john.meixner@gmail.com



## INTRODUCTION

Neuroscience is often considered to have a certain 'seductive allure'.[1] Its mystique should not be surprising. Seeking to understand the network of nearly 100 billion neurons that make up the human brain, neuroscience examines some of the most difficult questions imaginable. And yet, it is also a deeply personal discipline—questions like, 'How do we create memories?' and 'What causes emotions?' touch on experiences shared by all people.

Does the mystique of neuroscience cause individuals to ascribe undue weight to neuroscientific findings, or assume that neuroimages indicate research quality? Over the past decade, a literature has sprung up seeking to answer questions like these. One important context for that literature is the courtroom. The use of neuroscience in legal contexts has increased in recent years,[2] and scholarly interest for the topic has skyrocketed,[3] leading to the very natural question: will judges and jurors—who are typically not scientists—be able to give neuroscientific evidence presented in court appropriate weight, or will they be overly swayed by alluring images and conclusions? Should the rules of evidence protect factfinders from such undue influence?

In a recent issue of *The Journal of Law and the Biosciences*, Francis Shen and his coauthors offer a new dataset examining this issue. It is among the first, to my knowledge, in the context of neuroscience-based memory recognition detection—a relatively new field in which an examiner attempts to detect whether an individual possesses knowledge that allows an inference as to that person's role in a prior event (such as a crime),

1 Deena Skolnick Weisberg et al., *The Seductive Allure of Neuroscience Explanations*, 20 J. Cogn. Neurosci. 470 (2008).

2 See eg Nita A. Farahany, *Neuroscience and Behavioral Genetics in U.S. Criminal Law: An Empirical Analysis*, 2 J. L. & Biosci. 485 (2016).

3 Owen D. Jones & Francis X. Shen, *Law & Neuroscience in the United States*, *in* International Neurolaw, 349, 351 (2012).

and which has been regularly referenced for its potential application in court. In this brief commentary, I seek to critically examine what we can learn from that dataset, and provide some framework for future research in this area.

## GLEANING THE EFFECT OF NEUROSCIENCE FROM A LIMITED TASK

The study takes a straightforward approach in its two experiments, presenting a brief written narrative to mock jurors placed in one of a number of conditions.[4] In both experiments, the circumstantial evidence against the defendant is varied between strong, medium, or weak. The key manipulation, however, involves neuroscience—in both experiments, two of the conditions involved presenting mock jurors with neuroscience-based memory recognition evidence stating that the defendant either recognizes information relevant to the bad act or does not recognize that information. Both experiments also use control conditions in which no expert evidence is presented, and experiment 1, interestingly, uses two polygraph conditions, in which the participants were presented with expert evidence regarding a *polygraph*-based (not neuroscience-based) memory detection test (indicating either that the individual was lying or was not).

The key result is depicted in figures 1 and 2 of the paper—presenting the subjects with culpability-implicating neuroscience-based recognition evidence increased the mock jurors' likelihood of ascribing culpability to the defendant (a 'yes' or 'no' dichotomous decision) by 0.23 and 0.13 in experiments 1 and 2, respectively—a much smaller effect than changing the strength of the circumstantial evidence from weak/medium to strong, which increased the likelihood of ascribing culpability by 0.58 and 0.32, respectively. Based primarily on this result, the authors properly conclude that, at least in the context of these experiments, 'the effect of neuroscientific evidence is not nearly enough to outweigh the overall strength of the case', and that the findings thus 'suggest that there is not a seductive allure to neuroscientific evidence' and that 'neuroscience is unlikely to be an outcome-determinative silver bullet'.[5]

For the most part, I agree that the results support those conclusions, but I think that there are several good reasons to temper them, in light of the methods of these experiments. First, and most importantly, the study uses a vignette methodology—which has some issues in and of itself[6]—that gives the mock jurors very little understanding what neuroscience-based memory assessment is or how it works. The entire description of the expert evidence is only three sentences.[7] In contrast, in a real trial scenario, a substantial amount of time would be dedicated to an explanation of the test and the extent of its reliability. The authors understand and identify this in the paper, but I think they underestimate the extent to which this may skew their results—based on the current literature, at a minimum, an expert in this area would seek to testify

4 Francis X. Shen et al., *The Limited Effect of Electroencephalography Memory Recognition Evidence on Assessments of Defendant Credibility*, 4 J. L. & Biosci. 330, 348 (2017).

5 *Id.* at 358.

6 See eg Jonathan J. Koehler & John B. Meixner Jr., *Jury Simulation Goals*, *in* The Psychology of Juries: Current Knowledge and a Research Agenda For The Future (2017).

7 Shen et al., *supra* note 4, at 352.

(i) That the test has a near-zero false positive rate—that is, when the data indicate that the individual recognizes the item being tested for as meaningful, that conclusion is almost always correct;[8] and
(ii) That the test has a very low (10%–20%) false negative rate—that is, when the data indicate that the individual does not recognize the item being tested for as meaningful, that conclusion is very often correct.[9]

The expert would also, of course, describe the methodology in substantial detail, and provide other background information about P300 (the brainwave associated with memory recognition), neuroscience, and the difference between this sort of testing and lie-detection testing. Without at least some presentation of these details, I think it is difficult to know the true effect that presentation of such a test would have on a jury.

This issue is potentially further exacerbated by the language used in the vignettes describing the outcome of the memory recognition test—in both experiments, rather than simply stating whether the test indicated recognition, the vignettes describe whether the test implied that the defendant was being dishonest. While accurate in its form, that type of description may have caused confusion in the participants as to whether the test described was a lie-detector test—which may come with dubious connotations.

The second issue that I think is worth consideration is the substantial difference in the strength of evidence between the strong, medium, and weak conditions. In the first experiment, for example, the strong condition contains four items of circumstantial evidence that seriously implicate the defendant: his presence at the scene of a theft near the time of the theft, his access to the location of the theft, his past history of theft, and a text message consistent with theft.[10] In the weak condition, similar items are strongly exculpatory, including the defendant's lack of presence at the time of the theft, lack of access to the location of the theft, lack of criminal history, and lack of any text message indicating involvement. Not surprisingly, the effect of the strength-of-evidence variable easily overcomes the effect of the neuroscience evidence, but is that strength-of-evidence variability representative of real trials? Real cases tend to go to trial at least in part because they're close, where the overall evidence does not clearly favor one side. To draw an illustration: If the Shen et al. study is using evidence strength drawn from the tails of the normal distribution, but real trials tend to occur where the evidence is in a band near the center of that distribution, then the effect of neuroscience-based recognition detection evidence could be more pronounced in real trials than as modeled here, because the other evidence in real trials does not overwhelm so easily.

Of course, it could be the case that memory-detection evidence is potentially powerful (perhaps due to its high sensitivity and very high specificity, as described above), but not necessarily because it is *neuroscience-based*. Indeed, the underlying goal of Shen et al. was to inform the ongoing debate as to whether neuroscience evidence has a potentially prejudicial 'seductive allure'. To that point, the first experiment provides an

---

8 See eg John B. Meixner Jr., *Admissibility and Constitutional Issues of the Concealed Information Test in American Courts: An Update, in* DETECTING CONCEALED INFORMATION AND DECEPTION: RECENT UPDATES (forthcoming 2018).

9 See *Id.*

10 Shen et al., *supra* note 4, at 349.

interesting comparison demonstrating no difference between the effect of neuroscience-based memory detection evidence and polygraph-based memory detection evidence. Because the description of the two methods in the vignettes was so limited, and because the term 'polygraph' is so connected with lie detection (rather than memory detection) and potentially connected with preexisting presumptions about its validity, it is difficult to know what to make of this null result. Regardless, however, it provides a sort of control point that is helpful, and may be useful in future work in this arena.

## CONCLUSIONS AND FUTURE RESEARCH

In sum, though I think it is difficult to fully assess the results of this experiment, it serves as a wonderful introduction into how both neuroscience-based memory detection evidence and memory-detection evidence generally would be perceived by juries. I provide a few potential avenues for future research in this area.

One important—and underutilized—resource in the memory-detection arena is Japan. Japanese police have regularly used memory-detection paradigms in criminal investigations for more than half a century, and results from those tests are frequently used in court as well.[11] To my knowledge, those tests are all polygraph-based, rather than neuroscience-based, but Japan still provides a unique source of insight into how memory-detection paradigms work in practice, and how judges and jurors are likely to assess them. Though collaborations across countries and languages are obviously difficult, research on this topic in Japan could be conducted on individuals who have actually observed this type of evidence in court, and data from Japan as to how this type of evidence is actually presented would be invaluable.

Second, and perhaps more realistically (at least in the short term), I think replication and extension of this study are warranted. In my view, the next natural steps are to use more elaborate and realistic vignettes, or even videotaped presentations of evidence, to provide a more realistic look at the weight ascribed to neuroscience-based memory detection evidence when participants are fully aware of what the evidence means. Such future studies would be most useful if they can control for the effect of neuroscience independent of the use of the test itself, as the polygraph-based conditions in experiment 1 do at least in part.

Neuroscience continues to make advances that will make it attractive to use as evidence in court in a variety of ways, from criminal sentencing, to competency determinations, to credibility assessments. As that evidence begins to be offered, litigants will have to present arguments as the proper way to evaluate it. Research like the Shen et al. paper provides a rich source of data for those arguments, and further work in this area should be instrumental in allowing courts to arrive at the right decisions as to whether the evidentiary value gained is worth the potential for misuse.

[11] Akemi Osugi, *Daily Application of the Concealed Information Test: Japan*, *in* MEMORY DETECTION: THEORY AND APPLICATION OF THE CONCEALED INFORMATION TEST 253, 274 (B. Verschuere, G. Ben-Shakhar & E. Meijer eds., 2011).