Volume 62, Issue 4 Page 1119

Stanford Law Review

THROUGH A SCANNER DARKLY: FUNCTIONAL NEUROIMAGING AS EVIDENCE OF A CRIMINAL DEFENDANT'S PAST MENTAL STATES

Teneille Brown & Emily Murphy

© 2010 by the Board of Trustees of the Leland Stanford Junior University, from the *Stanford Law Review* at 62 STAN. L. REV. 1119 (2010). For more information visit http://www.stanfordlawreview.org.

THROUGH A SCANNER DARKLY: FUNCTIONAL NEUROIMAGING AS EVIDENCE OF A CRIMINAL DEFENDANT'S PAST MENTAL STATES†

Teneille Brown* & Emily Murphy**

As with phrenology and the polygraph, society is again confronted with a device that the media claims is capable of reading our minds. Functional magnetic resonance imaging ("fMRI"), along with other types of functional brain imaging technologies, is currently being introduced at various stages of a criminal trial as evidence of a defendant's past mental state. This Article demonstrates that functional brain images should not currently be admitted as evidence into courts for this purpose. Using the analytical framework provided by Federal Rule of Evidence 403 as a threshold to a Daubert/Frye *analysis*, *we demonstrate that, when fMRI methodology is properly understood, brain images are only minimally probative of a defendant's past mental states and are almost certainly more unfairly prejudicial than probative on balance. Careful and detailed explanation of the underlying science separates this Article from others, which have tended to paint fMRI with a gloss of credibility and certainty for all courtroom-relevant applications. Instead, we argue that this technology may present a particularly strong form of unfair prejudice in addition to its potential to mislead jurors and waste the court's resources. Finally, since fMRI methodology may one day improve such that its probative value is no longer eclipsed by its extreme potential for unfair prejudice, we offer a nonexhaustive checklist that judges and counsel can use to authenticate functional brain images and assess the weight these images are to be accorded by fact finders.*

---

† This Article is dedicated to our incredible friend and mentor Hank Greely.

* Teneille R. Brown, J.D., is an Associate Professor of Law at the University of Utah, S.J. Quinney College of Law. While writing this Article she was a fellow with the Stanford Center for Biomedical Ethics, a fellow at the Stanford Law School Center for Law and the Biosciences, and a research fellow with the MacArthur Foundation Law and Neuroscience Project.

** Emily R. Murphy, Ph.D., is a J.D. candidate (2012) at Stanford Law School. While writing this Article she was a fellow in the Stanford Law School Center for Law and Biosciences and a research fellow on the MacArthur Foundation Law and Neuroscience Project. Thanks to the MacArthur-funded Law and Neuroscience Project, Jeff Cooper, David Faigman, Jaime King, Michael Saks, Hank Greely, Stephen Morse, Kathryn Abrams, Walter Sinnott-Armstrong, Ryan Calo, R. Duncan Luce, William Uttal, and Sean Mackey for their thoughtful comments on earlier drafts of this Article. Special thanks to Nita Farahany and Adina Roskies for their in-depth feedback and guidance.

INTRODUCTION

> *Intent is basically a subjective element, that is, the operation of a person's mind. However,* since we cannot x-ray a person's mind *to determine what he is thinking, you may infer a person's intent by his acts or words or both.*[1]

The model jury instruction above tantalizes those who hope to use technology to improve upon current methods of determining what is in someone's mind. We now know that x-raying a person's brain does not provide direct access to her mind, but that was not always the case.[2] As with phrenology[3] and the polygraph[4] before, courts are again confronted with a technology that the media claims is capable of mind reading: functional neuroimaging. A recent criminal case in California provides an example of the attempted use of functional neuroimaging to address what was—or was not—going on in a defendant's mind at the time of the crime.

In Monterey County, defendant Francisco Saviñon was charged with attempted murder.[5] After the dissolution of his long-term relationship and the loss of his job, Saviñon became increasingly despondent and decided to commit suicide. He testified that while sitting in his car inhaling carbon monoxide fumes from the exhaust pipe, he "remembered" that a "recurring theme" in conversations with his ex-girlfriend (Jane Doe) was that "if they could be alone, free of kids and former spouses and the daily stresses of life—they

1. CHARGES TO THE JURY AND REQUESTS TO CHARGE IN A CRIMINAL CASE IN NEW YORK § 4:18 (2009) (emphasis added).

2. Attempts to use photographs to draw inferences about someone's mental state have been made for hundreds of years. In the nineteenth century, parties relied upon pictures of the testator to indicate that he either was or was not "of sound mind" at the time of execution of his will. *See, e.g.*, Wilcox v. Forbes, 53 N.E. 146, 146 (Mass. 1899) (upholding a judge's refusal to consider photographs of the deceased as evidence of dementia); Varner v. Varner, 9 Ohio Cir. Dec. 273, 273 (Cir. Ct. 1898) (reversing and remanding where a photograph had been introduced to establish a testator's competence, because even "[t]he most devout believer in the efficacy of the X rays has never urged them as a means of discovering the mind . . . .").

3. "Although the claims of the phrenologists were never taken seriously by the mainstream scientific community, they did capture the popular imagination of the time. In fact, a textbook on phrenology published in 1827 sold over 100,000 copies." MARK F. BEAR ET AL., NEUROSCIENCE: EXPLORING THE BRAIN 9 (1996).

4. *See* James R. Wygant, *Uses, Techniques, and Reliability of Polygraph Testing*, 42 AM. JURISPRUDENCE, TRIALS 313, 330 (2008) ("The importance we place upon the ability to discern the truth is reflected in how we characterize other people. If we say that a person is 'naive' or 'gullible,' we usually mean that he or she can be fooled easily. To some extent we relate 'wisdom' to insight . . . . The desire to be wise, to know the truth, has led naturally to the development of aids to observation, including the polygraph.").

5. Press Release, Monterey County Dist. Attorney, Firefighter Sentenced to Thirty Years, Eight Months for Attempted Murder (Feb. 27, 2009) (on file with authors) [hereinafter Saviñon Press Release].

would be happy."[6] Saviñon claimed that in his altered state of consciousness, due to alcohol intoxication and carbon monoxide exposure, he decided "this [recurring theme] meant she wanted to die with him and they would both be happy if they went to heaven together."[7]

Saviñon got into a second car, drove to Jane Doe's condo, and waited for her to emerge. When Jane Doe left her building, Saviñon attacked her and took her at knifepoint in her own car to his apartment. When she gained awareness of her surroundings and struggled against him, he stabbed her at least twice before forcing her into his car. He duct-taped her to the passenger seat and attempted to poison them both with the exhaust fumes. Jane Doe managed to convince him not to kill them both. Saviñon released her on the condition that she promised not to tell anyone who attacked her. Saviñon then bandaged her wounds and took her back to her apartment, where he brought her a cordless phone so she could call 911.[8]

Saviñon entered pleas of not guilty and not guilty by reason of insanity, based on a claim that due to his psychiatric diagnosis of major depressive disorder and his acute carbon monoxide exposure, he was cognitively impaired and thus could not have formed the required mens rea for attempted murder.[9] The defendant's claims were to be supported by family testimony, psychiatric clinical testimony, and by images from a functional neuroimaging device known as single proton emission computerized tomography (SPECT) machine.[10] Dr. William Klindt of San Jose Brain Clinic obtained the SPECT images about five months after the incident.[11]

Functional neuroimaging devices like SPECT are not generally accepted in

6. Defendant's Offer of Proof and Points and Authorities Re SPECT Evidence at 2, People v. Saviñon, No. SS070622A (Cal. Super. Ct. Jan. 12, 2009) (on file with authors) [hereinafter Saviñon Defendant's Offer of Proof].

7. *Id.* This admission in and of itself may acknowledge that the defendant intended to kill his ex-girlfriend, but defense experts planned to testify that the defendant's beliefs were the product of disordered reasoning, and "that defendant was legally insane at the time of the offense." *Id.* at 3-4. This would tend to negate an inference that the defendant had the requisite mens rea to kill.

8. *Id.* at 3.

9. *Id.*

10. *Id.* SPECT scanning uses a radionuclide tracer to image blood flow and brain metabolism, and is similar to fMRI in the reliance on blood flow assessment to make inferences about neural function. SPECT scanning, like fMRI, is not presently useful for making psychiatric diagnoses. We will discuss each device in greater detail *infra* in Part II.

11. *Id.* at 4. The defense counsel in *Saviñon* acknowledged that the brain images could not prove whether the defendant was legally insane or possessed the specific intent to kill. The defense argued instead that Dr. Klindt would testify that "the functioning of the brain, as demonstrated by the SPECT imaging, is consistent with the diagnoses [of] . . . 'major depression' . . . [and] consistent with toxic exposure . . . ." *Id.* at 6-7.

the medical or scientific community for the purpose of validating depression.[12] Even so, Dr. Klindt was prepared to assert that SPECT's appropriate neurological uses transferred validity to other as-yet unproven forensic psychiatric uses.[13] It is improper for advocates to blur together the multiple uses of a technology such as functional neuroimaging.[14]

In addition to highlighting the way functional neuroimaging may be co-opted by parties to a case, the *Saviñon* case also illustrates the important and practical impact of resources in the adversarial system. Despite the prosecution's having a written affidavit by a respected SPECT expert[15] refuting the validity of SPECT for forensic psychiatric purposes, the district attorney's office did not have the funds to pay for this expert to travel to the Bay Area for both the pretrial evidentiary hearing and, if admitted, the actual trial itself. Understandably, the prosecutor facing the pretrial hearing was worried that the paper testimony from her expert would not be as persuasive as the defense expert's oral testimony and accompanying colorful brain images.[16]

At the time of this writing, data are being gathered to assess the frequency

12. Declaration of Alan D. Waxman, M.D. at 2 (Dec. 4, 2008) (unsubmitted prosecution motion in *People v. Saviñon*, No. SS070622A (Cal. Super. Ct. Jan. 12, 2009)) (on file with authors).

13. In a pre-trial defense motion, the defense argued that because SPECT is used clinically to track Alzheimer's disease, strokes, and trauma, it should also be accepted for the very different purpose of diagnosing psychiatric conditions such as depression. Saviñon Defendant's Offer of Proof, *supra* note 6, at 7-8 (relying on a secondary source to describe medical diagnostic use of SPECT imaging).

14. In California, expert scientific evidence is reviewed under the *Kelly* test, which states that for a novel technology to be admitted, it "must be *sufficiently established to have gained general acceptance in the particular field to which it belongs*." People v. Kelly, 549 P.2d 1240, 1244 (Cal. 1976) (quoting Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923)). To see how the *Kelly* test ought to be applied with regard to neuroimaging offered to prove past mental states, see People v. Ford, No. B171801, 2005 WL 236487, at *7 (Cal. Ct. App. Jan. 25, 2005) ("Dr. Woods intended to testify that a SPECT scan performed on appellant's brain confirmed or was consistent with his clinical findings of cognitive impairments and of impairment to the frontal and temporal lobes of appellant's brain. The trial court excluded any reference to the SPECT scan results, finding that the technology was not generally accepted in the medical community for diagnostic purposes in a forensic setting and therefore did not meet the *Kelly* standard for admissibility."). See *infra* Part III for a discussion of the various tests that govern the admissibility of scientific evidence, and how those tests should apply to neuroimages offered to prove mental state.

15. Declaration of Waxman, *supra* note 12.

16. We provided the prosecutor with the arguments laid out in the rest of this paper, but before the admissibility of the SPECT scan and Dr. Klindt's accompanying testimony could be determined in a pretrial evidentiary hearing, the defendant pled to thirty years in prison. *See* Saviñon Press Release, *supra* note 5. The resource issue may cut both ways: indigent defendants reliant on public resources may have less access to brain imaging, while private defense firms may spare no expense and steamroll the prosecution. Conversely, while at present the defense proffers most examples of brain imaging as evidence, so too could the state inappropriately rely on neuroimaging to bolster arguments of dangerousness and civil commitment. Prosecutory use in individual cases is, of course, limited by the compliance of the defendant to submit to a brain scan.

of neuroimaging-based evidence offered or accepted in court.[17] Anecdotal reports such as the *Saviñon* case lead us to hypothesize that functional brain images are showing up with increasing frequency in state trial courts with advocates attempting to rely on brain images to bolster otherwise weak mens-rea-type claims.

In this Article, we argue that functional brain images should not currently be admitted into evidence to prove or rebut criminal mens rea charges.[18] We do this within a framework built on Federal Rule of Evidence 403: assessing probative and prejudicial value of functional brain imaging for mens rea claims. We conclude that functional brain images are minimally probative when introduced to prove a defendant's mens rea, and almost certainly more unfairly prejudicial than probative on balance. Furthermore, we argue that Rule 403, rather than the *Daubert*[19] (or similar) rules governing scientific evidence, provides both (a) the necessary individualized assessment of claims and (b) room to allow the technology, as well as the public understanding of the technology, to improve and adapt, rather than being defined by *Daubert* (or similar) rulings that may be categorically and too broadly applied.[20]

Part I of this Article establishes the background motivation for this

17. The MacArthur Law and Neuroscience Project is currently gathering data on how often neuroimaging evidence is being introduced in California trial courts, by whom, and for what purpose. This project is led by Hank Greely of Stanford Law School and involves contacting district attorneys and public defenders in each California county to assess the relative frequency of reliance upon neuroimaging evidence in lower courts. For details on this project, see The Law & Neuroscience Project, http://www.lawandneuroscienceproject.org (last visited Nov. 17, 2009). The most systematic way to research case frequency is by going to the reported appellate cases, but these cases will not include those where the defendant used neuroimaging successfully and was acquitted. Capital cases where the fMRI successfully resulted in a life sentence rather than the death penalty are unlikely to be appealed, and the evidentiary issues are often not dealt with in great detail. Often it is not even possible to determine whether the discussed evidence was admitted or who the expert witness was. As such, the appellate record only tells a sliver of the story of admissibility, and we need to scour trial court records for a more representative view. An excellent review of the use of neuroscience in juvenile cases from Terry Maroney demonstrates the legal and scientific limitations of such attempts. Terry A. Maroney, *The False Promise of Adolescent Brain Science in Juvenile Justice*, 85 NOTRE DAME L. REV. 89 (2010).

18. While we focus here on criminal mental states such as specific intent or purpose to kill, the bulk of our arguments could also apply to the use of functional brain images to prove mental states such as pain, bias, or experiential knowledge. It is important to reiterate that in narrowing our focus to *functional* brain images addressed to past mental states, we are not evaluating *structural* brain images such as those that result from X-ray, CT, or regular MRI scans. Structural scans do not render data or images of brain activity; rather, they render images of brain structures such as tumors or tissue damage.

19. Daubert v. Merrell Dow Pharm., 509 U.S. 579 (1993).

20. For example, once a *Daubert* hearing is conducted on a category of scientific evidence (think Bendectin and birth defects), it may be thought of as a legal rule rather than a factual assessment that depends greatly on the specifics of the case. This does not have to be the case, but one of our chief concerns is that judges and lawyers may blur together many different uses of a technology such as functional neuroimaging.

argument and lays out the terrain of why and how functional neuroimaging is being used in courts to negate mens rea. Part II explains each step of the imaging methodology—specifically, functional magnetic resonance imaging (fMRI)—in plain terms, critically examining the complex and non-obvious steps required to produce the colorful images. These steps have direct relevance to admissibility and probative value determinations.[21] While previous articles have addressed the admissibility of neuroimages generally,[22] our purposes are (a) to provide sufficient detail that would be useful to a judge or attorney who is actually confronted with an evidentiary hearing and (b) to establish the scientific baseline to which admissibility decisions in the future may be analogized.

Part III applies the rules of evidence to the scientific details from Part II. Part III argues that the unfairly prejudicial effects will likely substantially outweigh any probative value of functional brain imaging for mens rea claims, even if other admissibility hurdles are cleared. In the Appendix we offer guidelines for authenticating images as well as properly assessing the weight to accord to such evidence in the event that functional brain imaging technology advances such that it is truly probative for certain purposes and its prejudicial value can be sufficiently mitigated.

I. FUNCTIONAL NEUROIMAGING FOR MENS REA CLAIMS

The arguments advanced in this Article are narrowly aimed at functional brain imaging evidence, both images and expert testimony, addressed to mens rea claims. Some arguments may generalize to other courtroom uses of

---

21. As the reader will discover, one of the chief problems with functional neuroimaging is the "epistemic mismatch," that is, the difference between the steps a viewer thinks are required to produce an image, and the actual number of steps required. *See* Adina L. Roskies, *Neuroimaging and Inferential Distance*, 1 NEUROETHICS 19, 20-21 (2008). This problem can be addressed by proper authentication of the evidence, something that has not been addressed by any other article.

22. For a thoughtful review of how courts have analyzed the admissibility of PET and SPECT methodology, see Neal Feigenson, *Brain Imaging and Courtroom Evidence: On the Admissibility and Persuasiveness of fMRI*, 2 INT'L J.L. CONTEXT 233 (2006). Feigenson's article sets forth guidelines for how courts may interpret fMRI in a very accessible way, but as it does not apply the rules of evidence to a particular use, it ends up being too charitable in its abstract assessment of probative value. For another interesting article on the admissibility of neuroimaging, see Jane Moriarty, *Flickering Admissibility: Neuroimaging Evidence in the U.S. Courts*, 26 BEHAV. SCI. & L. 29 (2008), which covers many types of brain imaging evidence for many uses. Because there are important differences between structural and functional imaging devices, and critical differences between use of a scan to show brain damage versus use for inferences of mental states such as deception, these two important articles are much broader in their scope than our Article. Furthermore, Mark Pettit addressed the admissibility of fMRI in a 2007 article. However, his piece can be viewed as more of an early issue-spotting article as it does not interpret the rules of evidence by looking directly at the current state of the science. *See* Mark Pettit, Jr., *FMRI and BF Meet FRE: Brain Imaging and the Federal Rules of Evidence*, 33 AM. J.L. & MED. 319, 319 (2007).

functional brain imaging; certainly, we expect the scientific background to be useful in many cases that use functional neuroimaging to make claims about complex mental processes such as lying and recalling memories.[23]

The motivating factors behind this Article are twofold. First, we anticipate an expansion of the use of functional brain imaging as forensic evidence, based on existing case precedent, advancing technology, and increasing availability.[24] For those reasons, this Article attempts to provide a substantial practical grounding for judges and legal practitioners to assess such evidence's utility and potential for unfair prejudice. Second, the emerging field of neuroscience and law has seen some speculation about the scope of the impact of neuroscience on the law, particularly at the level of legal philosophy and justification for punishment. If such impacts are anticipated to take root in the courtroom,[25] a solid grounding in evidentiary analysis is a primary concern.

A. *What Is Functional Neuroimaging?*

Functional neuroimaging (or functional brain imaging) refers to a class of nonsurgical devices and methodologies that allow measurement of living brain activity. This category is distinct from structural imaging, such as a CT scan or MRI. Structural imaging provides images of gross anatomical features, but not of underlying neuronal or metabolic activity.

There are a few different types of functional neuroimaging devices. fMRI is the most popular functional brain imaging device in cognitive neuroscience research, as discussed in Part II. Tremendous excitement abounds regarding the research and clinical applications of fMRI. It is used preoperatively by

---

23. By "mens rea" claims we refer to the legal term of art in criminal law referring to a determination of culpability that attaches to a specific act, such as purpose, knowledge, or recklessness—or the lack thereof—at the time of the crime. We distinguish this use from other "mental state" uses, such as after-the-fact tests attempting to establish truth-telling or presence or absence of crime-related memories. Advocates have offered, and continue to offer, neuroimaging evidence for these other "mental state" uses. The technical and legal conclusions of this Article will largely apply to both types of uses.

24. *See generally* Purvak Patel et al., *The Role of Imaging in United States Courtrooms*, 17 NEUROIMAGING CLINICS N. AM. 557 (2007) (highlighting recent courtroom uses of imaging technology). We might also see courtroom uses swell as more commercial providers of fMRI services enter the market. According to No Lie MRI's website, "[s]ince the discrediting of the polygraph a large market for accurate truth verification and lie detection has been largely untapped. The current estimation of this market, conservatively based on the peak market demand for polygraph testing in the mid-1980s, is $3.6 billion. However it is likely the technology used by No Lie MRI, Inc. will be able to expand the market well beyond this level." No Lie MRI, Investors Overview, http://www.noliemri.com/investors/Overview.htm (last visited Nov. 17, 2009).

25. An evidentiary analysis seems to be a natural starting point for the introduction of neuroscience to the law. However, recent scholarship has demonstrated that the effect of neuroscience may be minimal, and perhaps efforts to effect change in the legal system should be attempted through legislative and policy channels rather than in individualized cases. *See* Maroney, *supra* note 17, at 169.

neurosurgeons to attempt to localize areas of critical perceptive, motor, and cognitive functions so that these abilities are less likely to be destroyed during tumor or tissue resection. It is also used to assess stroke damage and to follow the progression of Alzheimer's and epilepsy. Notwithstanding these important medical uses, much imaging with fMRI is done exclusively in the research context. Even within the research context, however, "fundamental questions concerning the interpretation of fMRI data abound, as the conclusions drawn often ignore the actual limitations of the methodology."[26] These fundamental questions are critical to an assessment of fMRI's evidentiary value. The conclusions drawn in the research context may often not be as far afield as those being asserted in the courtroom.

B. *Mens Rea Claims*

Criminal acts typically must be accompanied by a particular mental state to be punishable.[27] The separate requirements for the voluntary act (actus reus) and the guilty mind (mens rea) endorse a separation between mental state and bodily act that remains the dominant view in American criminal law.[28] The mens rea requirement stems from the common law notion of reserving punishment for those behaving wickedly.[29] However, the doctrine has evolved to have less to do today with the character of the individual and more to do with the coupling of a particular state of mind (or level of deliberation) with the criminal act.[30]

---

26. Nikos K. Logothetis, *What We Can Do and What We Cannot Do with fMRI*, 453 NATURE 869, 869 (2008).

27. Strict liability crimes such as speeding, drunk driving, and statutory rape are the exceptions to the rule, but the large majority of criminal acts require some showing of intent or knowledge. *See* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 10.01, at 117 (5th ed. 2009).

28. As Joshua Greene and Jonathan Cohen have pointed out, the idea that the brain and body are somehow different from the mind (i.e., dualism) runs deep not just in the law, but also in our folk intuitions of justice. Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y LONDON 1775, 1781 (2004) ("As long as the mind remains a black box, there will always be a donkey on which to pin dualist and libertarian intuitions. . . . Arguments are nice, but physical demonstrations are far more compelling. What neuroscience does, and will continue to do at an accelerated pace, is elucidate the 'when', 'where' and 'how' of the mechanical processes that cause behaviour.").

29. Nita A. Farahany & James E. Coleman, Jr., *Genetics, Neuroscience, and Criminal Responsibility*, *in* THE IMPACT OF BEHAVIORAL SCIENCES ON CRIMINAL LAW 218, 219 (Nita A. Farahany ed., 2009).

30. DRESSLER, *supra* note 27, § 10.02[B]-[C], at 118-19. Often intent is the state of mind that is contested, though under the Model Penal Code, it is purposeful action or knowledge. States that have adopted the Model Penal Code define specific intent crimes such as murder as requiring the mental state of "knowingly" or "purposely." MODEL PENAL CODE § 2.02(2)(c)-(d) (2008). The mental state required for criminal manslaughter is recklessness (involving actual awareness of the danger of death), and that for criminally

This Article specifically addresses the use of functional neuroimaging to support what we are calling "mens rea claims," referring simultaneously to defenses that attempt to inject doubt into the prosecution's case of proving a requisite mental state, and to affirmative defenses such as insanity. As illustrated by the defense arguments in *Saviñon*, the doctrinal curtain that exists between negating mens rea and waging an insanity defense is really more like a veil.[31] The principal distinction between the two is quite straightforward, but in practice the boundaries are blurred. Recall that the defense in *Saviñon* used an argument about Saviñon's lack of mens rea to bolster a legal insanity defense. How one argues an insanity defense depends upon the jurisdiction. In most but not all cases, the states retaining an insanity defense require that the defendant bear the initial burden of raising the defense. In many jurisdictions, once the defendant has done this, the burden of persuasion shifts to the prosecution to prove sanity beyond a reasonable doubt.

As "mens rea" translates literally to "guilty mind," some group the two concepts together and think of mens rea as both the absence of an affirmative insanity defense *and* the presence of the mental state that must be coupled with the criminal act.[32] In this way, "mens rea defenses" could include negating mens rea as well as affirmative claims of insanity and diminished capacity. Technically, our thesis could apply to all such mens rea defenses. While the two are distinct (a defendant could in theory be legally insane and still have intended to kill someone, or someone could be legally sane and not have so intended), in practice evidence of a mental disease or defect will often be used to establish either.[33]

The distinction between criminal mind and criminal act informs theories of punishment, but it also performs another function: it establishes a legal

---

negligent homicide is gross negligence. *Id.* §§ 210.3-.4.

31. *See* Christopher Slobogin, *The Supreme Court's Recent Criminal Mental Health Cases*, CRIM. JUST., Fall 2007, at 8, 12 ("[I]nsanity and mens rea are distinct concepts, even though they may overlap to some extent.").

32. Stephen J. Morse & Morris B. Hoffman, *The Uneasy Entente Between Legal Insanity and Mens Rea: Beyond* Clark v. Arizona, 97 J. CRIM. L. & CRIMINOLOGY 1071, 1075 (2007).

33. Mens rea is a matter of what actually occurred: *did* the defendant intend to do the crime at the time of the act? Legal insanity is a question of capacity or control: was the defendant *capable* of understanding the nature of what he was doing and knowing that what he was doing was wrong, or could he not have done otherwise? Because of this distinction, which is really more of a layering of factors, neuroscience evidence may perhaps be more relevant for claims of legal insanity than it is for claims negating mental state to the extent that the former can rely on present-time assessments, whereas the latter must look back to the time of the act itself. In most jurisdictions, legal insanity is a question of capacity and not actual control. In these jurisdictions, therefore, the prosecution need not show that the defendant *actually* appreciated the nature and wrongfulness of his actions, only that he was capable of doing so. For a useful review of the history of the insanity defense and the distinction between the "control" or "capacity" tests of legal insanity, see Richard E. Redding, *The Brain-Disordered Defendant: Neuroscience and Legal Insanity in the Twenty-First Century*, 56 AM. U. L. REV. 51, 80-85 (2006).

requirement that cannot be concretely measured. The task of identifying the subjective criminal mind has always and continues to elude us.[34] The promotion of technological flops such as phrenology and the polygraph may be understood, in part, as attempts to impose greater scientific certainty on the assessment of mens rea and criminal responsibility.[35] Because we cannot presently read someone's mind to determine her mens rea at the time of the crime, the jury is often told it can rely on the objective circumstances surrounding the criminal's conduct to draw inferences about her state of mind.[36] For example, the purpose to kill may be inferred from evidence showing that a defendant used a deadly weapon on a vital part of the victim's body.[37] Intent can also be supported by evidence of flight, motive, or other behaviors consistent with intentional behavior.[38] But all of these measures are indirect, and they rely on observable behaviors to analogize from one's own experienced contingencies between mental states and behavior to inferences of purpose. As Christopher Slobogin and other scholars have pointed out, isolating the criminal state of mind from the act typically relies on a reconstructed narrative that fuses together our beliefs, intentions, and actions.[39]

---

34. While the precise genesis of the mens rea requirement is disputed, the tying of a criminal act with a criminal state of mind can be traced back to "Jewish law, Christian theology, Codex Justinianus, and even tribal English law." Keren Shapira-Ettinger, *The Conundrum of Mental States: Substantive Rules and Evidence Combined*, 28 CARDOZO L. REV. 2577, 2579 (2007); *see also* United States v. Cordoba-Hincapie, 825 F. Supp. 485, 489-92 (E.D.N.Y. 1993) (providing an incredibly thorough history of the mens rea requirement through ancient, Medieval, and modern law).

35. When the polygraph was first introduced, it was touted for its objectivity over medieval attempts to establish the truth. "The interrogation of criminal suspects may not be easier today than formerly, but it is at least on a more objective basis." Paul V. Trovillo, *A History of Lie Detection*, 29 J. CRIM. L. & CRIMINOLOGY 848, 848 (1939). In the same article, Trovillo invokes an Ayur-Veda text dated at 900 BC to frame his argument for greater objectivity: "A person who gives poison may be recognized. He does not answer questions . . . he speaks nonsense, rubs the great toe along the ground, and shivers; his face is discolored . . . ." *Id.* at 849.

36. Bruce Ledewitz, *Mr. Carroll's Mental State or What Is Meant by Intent*, 38 AM. CRIM. L. REV. 71, 72 (2001); *see also* Franklin v. Anderson, 267 F. Supp. 2d 768, 791 (S.D. Ohio 2003) ("[T]he purpose with which a person does an act or brings about a result is determined from the manner in which it was done, the means or weapon used and all of the other facts and circumstances in evidence.").

37. *See, e.g.*, Curry v. State, 657 S.E.2d 218, 220 (Ga. 2008) ("[Defendant] contends that the trial court erred in charging the jury that it could infer the intent to kill from the use of a deadly weapon. . . . [G]iven the overwhelming evidence that [defendant] repeatedly shot the victim without provocation from the victim, the charge constitutes harmless error." (footnotes omitted)); Commonwealth v. Jones, 912 A.2d 268, 279 (Pa. 2006).

38. *See, e.g.*, State v. Aviles, 944 A.2d 994, 1000 (Conn. App. Ct. 2008) (holding that there was sufficient evidence to support a finding that the defendant had intent to kill his girlfriend based on his behavior, which included arguing with his girlfriend, retrieving a gun, and shooting her after she slammed a door); *see also* People v. MacCullough, 274 N.W. 693, 698 (Mich. 1937) ("Evidence of flight is not substantive evidence of guilt, though it may bear upon the purpose and intent of the party." (citation omitted)).

39. *See* CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE: THE ROLE OF LAW,

Determining the mens rea of an accused person is to attempt mind reading (and, indeed, time travel). As noted above, this task is presently accomplished by inferences drawn from observable behavior. Functional brain imaging seems to offer a more direct method of "mind-reading" by offering quantification of some processes going on inside our skulls. It is not our purpose here to delve deeply into the philosophical problems of "knowing" other minds. It is worth noting, however, that it is precisely the question of inferring a mental state from a quantified brain state that makes functional brain imaging as evidence of mental states different in kind from other forms of scientific evidence to which it is sometimes analogized. The obvious comparison is to forensic genetics. While DNA testing can reliably identify suspects and place people at crime scenes, the results of a genome-wide association study or a point mutation analysis cannot yet be used to make inferences about complex mental states. Even so, genetics has successfully overhauled the criminal justice system due to the fixed nature of our genes and the validity of our genetic tests. But unlike our genes, our brains and mental states change.

Any ability to "read our minds" depends on several working assumptions, including: the strength of the contingency between brain states and mental states, the completeness of our knowledge of cognitive neuroscience, and the accuracy of behavioral assessments and self-reporting that comprise the library of existing knowledge about both normal and pathological mental states.[40] These assumptions are not addressed de facto by improving technology or analytical methods.[41] Rather, the specificity and relevance of the behavioral

---

SCIENCE, AND SPECULATION IN ADJUDICATING CULPABILITY AND DANGEROUSNESS 44 (2007) ("[A]lthough ascertaining objective truth might be possible with respect to acts, *narrative thinking* dominates attempts to reconstruct mental state. . . . Science cannot tell us the truth about past mental states because science is meant to identify objective reality, not interpretations of reality. . . . Even if, in theory, particular past mental states can be said to 'exist' in some objective sense, as a practical matter science will not be up to the task of measuring them."); *see also* Deborah W. Denno, *Criminal Law in a Post-Freudian World*, 2005 U. ILL. L. REV. 601, 605 ("[W]hat people intend, think, and believe are paramount to assessing guilt; in some cases, they can mean the difference between life and death. How odd for a legal system to base so much on something about which it seems to know so little.").

40. *See* Martha Farah, *Neuroethics and the Problem of Other Minds: Implications of Neuroscience for the Moral Status of Brain-Damaged Patients and Nonhuman Animals*, 1 NEUROETHICS 9, 16-17 (2008).

41. Edward Cheng claims that "[i]t's not clear whether or not a somewhat reliable but foolproof fMRI machine is any worse than having a jury look at a witness. . . . If you want the status quo, fine, but in this case, the status quo might not be all that good." Alexis Madrigal, *MRI Lie Detection to Get First Day in Court*, WIRED SCI., Mar. 16, 2009, http://blog.wired.com/wiredscience/2009/03/noliemri.html. In fact, it *is* clear. fMRI has not been demonstrated to be better than the status quo. While we acknowledge that the current method of assessing a defendant's mental state is not ideal, functional brain imaging does not save us from the problem of human discretion, malingering, and distortion. What it does is present findings in a way that *appear* incredibly objective, infallible, and robust, without actually being any of these things. For our discussion of the unfairly prejudicial effects of neuroimages, see *infra* Part III.G.

tasks used to generate fMRI data must be improved upon before they should be considered for forensic use—a challenge that is acutely psychological and, perhaps when dealing with specific issues of malicious intent, at least somewhat normative in nature.[42]

Furthermore, the appropriate translation of functional neuroimaging to a forensic context demands that proponents answer the question of what additional explanatory power such data lends to a particular argument about mens rea. Answering this question is particularly crucial when fMRI is derived from a behavioral task and, at present, the mere absence of activation cannot be taken as proof of any deficit, inability, or incapacity. These are all important points that limit the probative value of functional brain imaging for mens rea purposes, and each will be addressed in greater detail below.

C. *Present and Anticipated Future Use of Functional Brain Imaging in Courts*

Brain imaging has been offered as courtroom evidence for a variety of reasons, albeit not always successfully. These purposes include everything from competence to waive Miranda rights,[43] subjective experience of pain in tort cases,[44] custody determinations,[45] mens rea defenses for fraud,[46] kidnapping, burglary,[47] and even murder.[48] A recent article by Terry Maroney catalogs the attempts to use brain imaging in juvenile cases.[49] Functional neuroimaging has already been admitted and relied upon as evidence of an individual's past mental state.[50]

---

42. Emily Bell and Eric Racine provide a succinct reminder of the challenges created by the task dependency of fMRI: "Careful consideration of fMRI task dependency casts doubts on the ease of translation of fMRI to yield robust and substantial real-world uses." Emily Bell & Eric Racine, *Enthusiasm for Functional Magnetic Resonance Imaging (fMRI) Often Overlooks Its Dependence on Task Selection and Performance*, 9 AM. J. BIOETHICS–NEUROSCIENCE 23, 23-24 (2009).

43. *See In re* Roberto H., No. B192678, 2007 WL 4533141, at *2 (Cal. Ct. App. Dec. 27, 2007).

44. *See* Response to Motion of Defendant to Preclude Testimony of Dr. Joy Hirsch, Ph.D. or in the Alternative for a *Frye* Hearing at 3, Koch v. W. Emulsions, Inc., No. C2006-1227 (Super. Ct. Ariz. Sept. 5, 2008).

45. *See In re* Jasmine M., No. B192729, 2007 WL 1139980, at *2 (Cal. Ct. App. Apr. 18, 2007).

46. *See* United States v. Mezvinsky, 206 F. Supp. 2d 661, 663 (E.D. Pa. 2002).

47. *See* People v. Herrera, No. B163516, 2003 WL 22962809, at *1 (Cal. Ct. App. Dec. 17, 2003).

48. *See* Zink v. State, 278 S.W.3d 170, 177-82 (Mo. 2009).

49. *See* Maroney, *supra* note 17.

50. fMRI was recently admitted at the sentencing stage of a the criminal trial of Brian Dugan of Illinois. *See* Greg Miller, *fMRI Evidence Used in Murder Sentencing*, SCI. INSIDER, Nov. 23, 2009, http://blogs.sciencemag.org/scienceinsider/2009/11/fmri-evidence-u.html; *see also* People v. Ward, No. B193719, 2008 WL 3906423, at *3 (Cal. Ct. App. Aug. 26, 2008) ("Ward's primary defense was that a significant preexisting brain injury, combined with alcohol, negated the mental state required to convict him of second degree murder.

One expert, Daniel Martell, has received so many requests from counsel that he began a business called "Forensic Neuroscience Consultants."[51] According to an interview with Martell reported in *The New York Times Magazine*, he has been hired by defense teams and prosecutors alike and has testified in several hundred cases over the last fifteen years.[52] While his experience may not generalize, he believes that death penalty litigation will be the site of neuroscience evidence's most dramatic impact, as "[s]ome sort of organic brain defense has become de rigueur in any sort of capital defense."[53] If Martell and other experts such as neurologist Helen Mayberg are correct, the tide is rising on this type of argument in criminal cases, particularly capital cases.[54]

While prosecutors may one day introduce fMRI as evidence of future dangerousness, presently defense teams appear to be the dominant users of neuroimaging in the courtroom. One practical reason for this is that it would be physically difficult for the state to compel a brain scan of an unwilling person. Neuroimaging has seen the courtroom in the sentencing phase of capital cases,

---

Ward . . . presented the expert testimony of Dr. Joseph Chong-Sang Wu, . . . [who] testified a positron emission tomography (PET) scan of Ward's brain function revealed 'a profound abnormality in the frontal lobe,' which is the part of the brain involved with judgment, awareness of the consequences of one's actions and the ability to regulate improper impulses."); People v. Ford, No. B171801, 2005 WL 236487, at *3 (Cal. Ct. App. Jan. 25, 2005); Commonwealth v. Yancy, 797 N.E.2d 371, 374-75 (Mass. 2003). For examples of EEG-based neuroimaging evidence that was introduced but not admitted, see *Commonwealth v. Henry,* 569 A.2d 929, 936 (Pa. 1990); *State v. Idellfonso-Diaz,* No. M2006-00203-CCA-R9-CD, 2006 WL 3093207 (Tenn. Crim. App. Aug. 8, 2006). For examples in which the defense attempted to obtain neuroimaging evidence but funds were denied, see *People v. Guerra,* 129 P.3d 321, 339 (Cal. 2006); *Mangum v. State*, 765 So. 2d 192, 193-94 (Fla. Dist. Ct. App. 2000).

51. The Forensic Neuroscience Consultants website lists "Neuropsychological Issues in Diminished Capacities/Diminished Actualities and related *Mens Rea* defenses" as an area of expertise. Forensic Neuroscience Consultants, Inc., http://forensicneuroscience.com (last visited Nov. 29, 2009).

52. Jeffrey Rosen, *The Brain on the Stand,* N.Y. TIMES, Mar. 11, 2007, § 6 (Magazine), at 48, 50.

53. *Id.* at 50. Similarly, Art Barnum & Ted Gregory, *Dugan's Brain the Subject of Sentencing Hearing*, CHI. TRIB., Nov. 6, 2009, at 14, describe the testimony of neuroscientist Kent Kiehl in the sentencing phase of convicted child rapist and murderer Brian Dugan: Kiehl testified for four hours that Dugan's fMRI brain scans help show that Dugan is a psychopath with an abnormally functioning brain. Dugan's lawyers asked jurors to consider this abnormality as evidence of Dugan's incapacity to make moral decisions, and thus not put him to death. *Id.* Nevertheless, Dugan was sentenced to death. *Id.*

54. Helen S. Mayberg is a neurologist at Emory University and has testified in many trials. She spoke about the burgeoning courtroom use of imaging studies at a presentation to MacArthur Foundation on Law and Neuroscience members at its annual meeting in Santa Barbara, California on May 29, 2008. *See also* Patel, *supra* note 24. As mentioned *supra* note 17, the MacArthur-funded Law and Neuroscience Project is currently building a database of lower court cases in which neuroimaging was introduced or attempted to be introduced.

where evidentiary hurdles are much lower and a capital defendant may have a constitutional entitlement to a psychiatric evaluation.[55] One might further argue that criminal defendants have a right under the Sixth Amendment to present exculpatory evidence such as fMRI images, though recent Supreme Court precedent places practical limitations on this right.[56]

D. *The Impact of Neuroscience on the Law: Grounding in Evidence*

The final motivation for this Article is to be useful to the nascent and expanding field of neuroscience and law. This field has received vigorous attention due to speculation or hope that neuroscience will be better able than psychology or philosophy to shed light on the mechanisms of the criminal mind and justifications for criminal responsibility.[57] In turn, some scholars argue that

---

55. *See* Ake v. Oklahoma, 470 U.S. 68, 83 (1985) ("[W]hen a defendant demonstrates to the trial judge that his sanity at the time of the offense is to be a significant factor at trial, the State must, at a minimum, assure the defendant access to a competent psychiatrist who will conduct an appropriate examination and assist in evaluation, preparation, and presentation of the defense. This is not to say, of course, that the indigent defendant has a constitutional right to choose a psychiatrist of his personal liking or to receive funds to hire his own. . . . [W]e leave to the States the decision on how to implement this right.").

56. As the defendant's right to compulsory process "would be defeated if judgments were to be founded on a partial or speculative presentation of the facts," the Sixth Amendment does not provide a right to present unreliable, confusing, or cumulative evidence. Donald Dripps, *Relevant but Prejudicial Exculpatory Evidence: Rationality Versus Jury Trial and the Right to Put on a Defense*, 69 S. CAL. L. REV. 1389, 1405 (1996) (quoting Taylor v. Illinois, 484 U.S. 400, 410-12 (1988)); *see also* EDWARD IMWINKELRIED & NORMAN GARLAND, EXCULPATORY EVIDENCE 19-35 (3d ed. 2004); Peter Westen, *The Compulsory Process Clause*, 73 MICH. L. REV. 71, 96-100 (1974) (chronicling the history of the inclusion of the Compulsory Process Clause in the Sixth Amendment).

In *United States v. Scheffer*, the Court held that a per se rule against admission of polygraph evidence in court martial proceedings did not violate the Sixth Amendment rights of the accused to present a defense. 523 U.S. 303, 309 (1998). The state advanced legitimate interests that included "ensuring that only reliable evidence is introduced at trial, preserving the court members' role in determining credibility, and avoiding litigation that is collateral to the primary purpose of the trial." *Id.* For a survey of various cases that further define the right to present exculpatory evidence, see *Holmes v. South Carolina*, 547 U.S. 319, 331 (2006); *Crane v. Kentucky*, 476 U.S. 683, 691 (1986); *California v. Trombetta*, 467 U.S. 479, 488-89 (1984); *United States v. Sparkman*, 500 F.3d 678, 682 (8th Cir. 2007); *Harris v. United States*, 834 A.2d 106, 124 (D.C. 2003); *Commonwealth v. Durning*, 548 N.E.2d 1242, 1248 (Mass. 1990); *King v. State*, 962 So. 2d. 124, 127 (Miss. Ct. App. 2007).

57. Joshua Greene and Robert Sapolsky, among others, make the case that neuroscience will lead us to see humans in more deterministic ways, where defendants with defective brain functioning should be thought of as broken machines, not moral monsters. *See* Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, *in* LAW AND THE BRAIN 207, 207 (Semir Zeki & Oliver Goodenough eds., 2006); Robert Sapolsky, *The Frontal Cortex and the Criminal Justice System*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y LONDON 1787 (2004); *see also* Rosen, *supra* note 52, at 52 ("'You can have a horrendously damaged brain where someone knows the difference between right and wrong but nonetheless can't control their behavior,' says Robert Sapolsky, a neurobiologist at Stanford. 'At that point, you're dealing with a broken machine, and

a better understanding of the mechanisms underlying brain dysfunction ought to lead us to treat defendants for their illness rather than punish them for moral wrongs. While in theory this type of individualized justice may be something to aspire to, its implementation is complicated in a system that has largely abandoned rehabilitation as a theory of punishment. Further, as Carter Snead and Nita Farahany point out, the same evidence offered by defendants as mitigating factors could also be used by the prosecution to demonstrate future dangerousness—an aggravating factor that can increase punishment following the punishment theory of incapacitation rather than retribution or rehabilitation.[58]

Will neuroscience have a direct impact on legal and societal conceptions of responsibility? The first step toward answering this question is gaining a complete understanding of what neuroscience can and cannot tell us when used as evidence in a courtroom. This question is best explored through the laws of evidence.

Regardless of what one thinks about the ultimate utility of neuroscience as courtroom evidence, neuroscience research *has* resulted in a better understanding of the neural basis of psychiatric disorders, addiction, and cognitive and emotional processing across individuals. This improved understanding will likely inform the law first through the development of programs and policy rather than through the adjudication of specific cases relying on individualized facts.[59] For example, one practical application may be to craft better treatment options for use by drug courts and parole boards. Depending on its particular use, neuroscience will have varying degrees of utility for legal scholars and practitioners.

---

concepts like punishment and evil and sin become utterly irrelevant. Does that mean the person should be dumped back on the street? Absolutely not. You have a car with the brakes not working, and it shouldn't be allowed to be near anyone it can hurt.'").

58. O. Carter Snead accurately points out that neuroscience evidence could be used for aggravating as well as mitigating factors. He incorrectly assumes, however, that mitigation is on the agenda for most of those engaged in law and neuroscience research:

> In the short term, these scientists seek to play a role in the process of capital sentencing by serving as mitigation experts for defendants, invoking neuroimaging research on the roots of criminal violence to support their arguments. Over the long term, these same experts (and their like-minded colleagues) hope to appeal to the recent findings of their discipline to embarrass, discredit, and ultimately overthrow retributive justice as a principle of punishment. Taken as a whole, these short- and long-term efforts are ultimately meant to usher in a more compassionate and humane regime for capital defendants.

In fact, most academic research in neuroscience and law is conducted without specific litigation agendas. O. Carter Snead, *Neuroimaging and the "Complexity" of Capital Punishment*, 82 N.Y.U. L. REV. 1265, 1265 (2007); *see also* Nita A. Farahany & James E. Coleman, Jr., *Genetics and Responsibility: To Know the Criminal from the Crime*, 69 LAW & CONTEMP. PROBS. 115, 130 (2006).

59. See Maroney, *supra* note 17, at 117, for a clear explanation of the level of the legal system at which neuroscience is most likely to have a lasting impact: the legislature.

II. SCIENTIFIC BACKGROUND: IMAGING BRAIN ACTIVITY AND MENTAL STATES

A. *The Science of Functional Neuroimaging*

1. *Overview of older methods*

Several functional neuroimaging techniques predate fMRI in development and in the courtroom, and we briefly review them here.[60] Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) rely on the injection of a radioactive tracer into the subject's bloodstream. The tracer emits pairs of gamma rays, which are detected and interpreted by a computer, and eventually result in a 3-D image of the brain.[61] PET's temporal resolution is on the order of seconds or minutes.[62] PET and SPECT have been quite useful clinically to diagnose many types of cancers, heart disease, and brain abnormalities. PET and SPECT share some commonalities with fMRI, including some of the basic methodologies for constructing the image from the data.[63]

Electroencephalography (EEG) measures electrical activity produced by the brain as recorded from electrodes placed on the scalp. Relative to PET or

---

60. For a great overview of the characteristics of and trade-offs among major functional neuroimaging technologies, see Judy Illes and Eric Racine, *Imaging or Imagining? A Neuroethics Challenge Informed by Genetics*, 5 AM. J. BIOETHICS 5, 8 (2005).

61. PET is based on the "fact that changes in the cellular activity of the brain of normal, awake humans . . . are invariably accompanied by changes in local blood flow. This robust, empirical relationship has fascinated scientists for well over a hundred years, but its cellular basis remains largely unexplained . . . ." Marcus E. Raichle, *Behind the Scenes of Functional Brain Imaging: A Historical and Physiological Perspective*, 95 PROC. NAT'L ACAD. SCI. 765, 765 (1998). SPECT works on similar principles, providing an image of blood flow in the brain thought to be closely coupled with neural metabolism. However, two recent studies that directly measured blood flow, blood oxygenation, and neural activity in animals suggest that the coupling of energy metabolism, blood flow and neuronal activity needs to be reexamined. *See* Anna Devor et al., *Stimulus-induced Changes in Blood Flow and 2-Deoxyglucose Uptake Dissociate in Ipsilateral Somatosensory Cortex*, 28 J. NEUROSCIENCE 14347 (2008); Yevgeniy B. Sirotin & Aniruddha Das, *Anticipatory Haemodynamic Signals in Sensory Cortex Not Predicted by Local Neuronal Activity*, 457 NATURE 475 (2009).

62. "Spatial" and "temporal" are two terms used frequently in functional neuroimaging. "Spatial" refers to the measurement of some data point in space, while "temporal" refers to the measurement of some data point in time.

63. fMRI has quite a few things in common with PET. Specifically, both require a careful experimental design that defines the criteria for subject inclusion and behavioral measure. They also obtain time-sequenced data sets that must be algorithmically reconstructed into a 3-D map of activity based on assumptions about the brain's physiology. The 3-D brain data sets must then be warped onto a brain structure, and they are subsequently averaged and subtracted against a control state. Just as with fMRI, "colors are used to substitute for the numbers in the [data set]," and the visible image is born. JOSEPH DUMIT, PICTURING PERSONHOOD 59 (2004).

SPECT, EEG has poor spatial resolution and is limited to assessing neural activity close to the scalp, but its temporal resolution is much better—on the order of milliseconds. Beyond its clinical uses, various forms of EEG-based investigation or interrogation techniques have received considerable media attention. One methodology was recently relied upon by an Indian court to convict a woman based on her "experiential knowledge" of the murder.[64] Another method dubbed "brain fingerprinting" is hailed by its developers as the next generation of biologically-based deception detection,[65] despite strong academic criticism and official rejection.[66] The lack of peer-reviewed data on the various methodologies of both forensic EEG-based technologies makes us similarly skeptical of the scientific validity of their use. Many of our critiques about functional neuroimaging as applied to mens rea claims apply also to functional neuroimaging attempts in the lie detection context, though we do not focus here specifically on that intended use.[67]

Most of the reported court cases that cite to neuroimaging refer to PET or SPECT. However, these methodologies are largely being replaced by fMRI in research and in practice. Unlike PET or SPECT, fMRI does not require the

64. *See* Anand Giridharadas, *India's Novel Use of Brain Scans in Courts Is Debated*, N.Y. TIMES, Sept. 15, 2008, at A10. Court opinion on file with the authors.

65. *See* Brain Fingerprinting, http://www.brainwavescience.com (last visited Dec. 10, 2009) (the website of the technology's inventor). Brain fingerprinting relies on the P300 wave—a positive event-related potential occurring 300 milliseconds after the particular stimulus that evoked it. "Brain Fingerprinting testing can prove that the suspect's brain does not have the salient details of the crime stored in it, that is, when the suspect does not remember or recognize the salient details of the crime." Brain Fingerprinting, Memory Issues, http://www.brainwavescience.com/MemoryIssues.php (last visited Feb. 5, 2010). For the only peer-reviewed paper on this method, see Lawrence A. Farwell & Emanuel Donchin, *The Truth Will Out: Interrogative Polygraphy ("Lie Detection") with Event-Related Brain Potentials*, 28 PSCYHOPHYSIOLOGY 531 (1991). The P300/EEG evidence was admitted in the Iowa case of *Harrington v. State*, though it was not heard by a jury and was not the basis for the postconviction relief by the judge. Harrington v. State, 659 N.W.2d 509, 516 n.6 (Iowa 2003) ("According to Dr. [Lawrence] Farwell, his testing of Harrington established that Harrington's brain did not contain information about Schweer's murder. On the other hand, Dr. Farwell testified, testing did confirm that Harrington's brain contained information consistent with his alibi."); *see also* John G. New, *If You Could Read My Mind: Implications of Neurological Evidence for Twenty-First Century Criminal Jurisprudence*, 29 J. LEGAL MED. 179, 185-87 (2008).

66. A 2001 General Accounting Office report concluded that "[o]fficials representing CIA, DOD, Secret Service, and FBI do not foresee using the Brain Fingerprinting technique for their operations because of its limited application." U.S. GEN. ACCOUNTING OFFICE, INVESTIGATIVE TECHNIQUES: FEDERAL AGENCY VIEWS ON THE POTENTIAL APPLICATION OF "BRAIN FINGERPRINTING," GAO-02-22 (2001), *available at* http://www.fas.org/sgp/othergov/polygraph/brainfinger.pdf. Specifically, "[t]he FBI conducted a thorough review of Dr. Farwell's research and on two occasions provided him an opportunity to present his findings. In this instance, Dr. Farwell did not conduct research that met the FBI's standards for research, nor did his research demonstrate the usefulness of this technique." *Id.* at 23.

67. For a thorough and recently published assessment of the state-of-the-art in neuroimaging for lie detection, see Hank T. Greely & Judy Illes, *Neuroscience-Based Lie Detection: The Urgent Need for Regulation*, 33 AM. J. L. & MED. 377 (2007).

injection of a radioactive tracer. fMRI's temporal and spatial resolution are also superior to PET's. The temporal resolution and signal-to-noise ratio of fMRI is not as good as that of EEG. Though it is too early in the technology's history for many appellate opinions to have discussed fMRI, experts agree that it will dominate older methods as courtroom evidence. This is due in part to the increased availability of fMRI devices and the reduction in their cost.[68] Much of our analysis in this Article applies to functional neuroimaging generally. However, in anticipation of an expanding use of fMRI, our inquiry focuses on the assessment of mental states using fMRI.

2. *Principles of fMRI*

fMRI is a relatively safe and noninvasive technique that *indirectly* measures the brain's activity.[69] The fact that fMRI is an indirect measurement cannot be stressed enough. fMRI does not directly measure neuronal activity or firing.[70] What follows is the rationale for nonetheless using fMRI to measure brain activity in studies of cognition and behavior.

Perceiving, thinking, acting, feeling, and even resting have associated neuronal firing. A growing body of evidence suggests that mental states—such as thoughts and emotions—are represented by patterns of neuronal activation in specific regions or networks of the brain. For many such cognitive or emotional tasks, an increase in neural firing in a particular region or network is interpreted as the brain doing "more" of that particular cognitive or emotional task.

Because neurons do not have internal reserves of energy, when they fire in response to some activity, oxygen-carrying blood must be transported to the neurons. This is called the "hemodynamic response."

Blood that is carrying oxygen behaves differently in magnetic fields than deoxygenated blood does. The difference in the magnetic properties of oxygenated blood allows fMRI to detect changes in blood flow related to activity. This is called the Blood Oxygen Level Dependent (BOLD) response.[71]

---

68. Eric Racine, Ofek Bar-Ilan, & Judy Illes, *fMRI in the Public Eye*, 6 NATURE REVS. NEUROSCIENCE 159, 159 (2005).

69. Joseph Mandeville & Bruce Rosen, *Functional MRI*, *in* BRAIN MAPPING 315, 315 (Arthur Toga & John Mazziotta eds., 2d ed. 2002). While fMRI is typically quite safe, it is not without risks. Contraindications include any metal in the body, pacemakers, and other devices. Even nonobvious metal, such as tiny fragments embedded in the eyes of metalworkers that otherwise may go undetected, could lead to serious burning. Also, the donut-shaped core of the machine is quite a small space. Individuals who are claustrophobic or who have a particularly large body frame may not be able to participate in fMRI research.

70. *See* Raichle, *supra* note 61, at 767-68.

71. Seiji Ogawa and colleagues labeled this enhancement the "BOLD contrast." *See* Seiji Ogawa et al., *Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation*, 87 PROC. NAT'L ACAD. SCI. USA 9868 (1990); *see also* Marcus E. Raichle & Mark A. Mintun, *Brain Work and Brain Imaging*, 29 ANN. REV. NEUROSCIENCE 449, 455 (2006).

In simple terms, when a region of the brain is "activated" in response to a perception or to enable a behavior, that region receives more oxygenated blood. Because oxygenated blood behaves differently in a magnetic field, the large magnet in the fMRI device can measure this influx. If the local oxygen use is more than adequately supplied by the influx of blood, then a positive BOLD response will result. If the local demand for oxygen exceeds that provided by the regional blood flow, then a negative BOLD response will result. Because the change in the blood oxygenation level in a spatial volume (called a voxel, like a three-dimensional pixel) does not directly capture the activity of neurons, fMRI does not yet provide detailed physiological information about the *neural* mechanisms underlying the mental state.

The actual experience of undergoing an fMRI is similar to that of undergoing a scan for medical purposes, which some readers may be familiar with. To begin, individuals are asked to lie on their backs on a thin bed, which slides into the center of a donut-shaped magnet core. While in this core, the individual is told to lie as still as possible; this state is sometimes facilitated by the doctor's fixing the individual's head in a frame. Depending on the research question, the subject is given some task to perform, which can include listening to stimuli or viewing visual stimuli projected into the scanner via a mirror. In tasks requiring a response, the subject may have their hands on a special controller to input their response without moving their head. During this time, the magnetic coils in the scanner receive different amounts of electric current. The exchange of current in the scanner produces a loud knocking sound. Subjects wear headphones to help drown out this noise and to receive any audio instructions or stimuli. While the subject is performing the task, software loaded on the device collects and stores data about the individual's oxygenated blood flow. The data must be heavily processed, aligned, smoothed, and filtered before it can be mapped onto a template of a human brain.

Employing a very strong magnet and the BOLD response, fMRI can aid in determining which regions of the brain are recruited in particular cognitive or perceptual tasks.[72]

3. *Knowns and unknowns about the BOLD response*

Perceiving, thinking, feeling, and acting correlate with changes in oxygen consumption and regional blood flow in the brain.[73] By comparing the BOLD response signal in the experimental or activation state with the control state,

---

72. Blood that is oxygenated is not attracted to a magnetic field, while blood that is deoxygenated is paramagnetic (attracted to a magnetic field). The presence of paramagnetic deoxygenated blood causes the signal-producing proton spins to dephase more rapidly. An increase in blood flow to active brain regions results in higher concentration of oxygenated blood and therefore results in an increased fMRI signal in that active region. *See* GORDON SARTY, COMPUTING BRAIN ACTIVITY MAPS FROM fMRI TIME-SERIES IMAGES 3 (2007).

73. *See* Ogawa, *supra* note 71.

small changes in signal intensity (on the order of four percent) are detectable.[74] This slight change in BOLD signal is what makes fMRI capable of facilitating inferences about brain activity. There are other less common methods for producing an fMRI image, but the BOLD response is by far the most widespread method.[75]

While the BOLD response is correlated with brain activity, researchers are currently hard at work to understand the relationship between the BOLD response and neuronal firing (including phenomena such as the depolarization, spiking of neurons,[76] and neurotransmitter release).[77] Nikos Logothetis suggests that the BOLD response may reflect the neural activity related to the synaptic *input*[78] and processing in a given region rather than the *output* (spiking activity).[79] The mechanism for this three-part process needs to be further confirmed in humans, as most of the experimental work has been done on anesthetized monkeys. Blood flow, oxygen availability, and glucose metabolism are undoubtedly linked, but the precise relationship between these factors is still unknown.[80] Further complicating the picture, a 2008 study found

---

74. *See* R. Todd Constable, *Challenges in fMRI and Its Limitations*, *in* FUNCTIONAL MRI: BASIC PRINCIPLES AND CLINICAL APPLICATIONS 75, 76 (Scott H. Faro & Feroze B. Mohamed eds., 2006).

75. Some question the strength of the BOLD response because the neural signals on which it is based are relative and not individually quantitative. Other methods, such as the Oxygen Extraction Fraction (OEF), have attempted to measure neural activity directly, but the signal-to-noise ratio is extremely low and statistical methods used to extract quantitative data have been largely unsuccessful thus far.

76. "Spiking" is a term in neurophysiology that refers to the generation of action potentials (discrete electrical events that are the basis of neuronal signaling).

77. *See* Constable, *supra* note 74, at 86. Still, there are recent developments that link neurotransmitter release to the BOLD response. *See* Brian Knutson & Sasha E. B. Gibbs, *Linking Nucleus Accumbens Dopamine and Blood Oxygenation*, 191 PSYCHOPHARMACOLOGY 813, 813 (2007) ("Dopamine release in the NAcc appears to increase local BOLD signal via agonism of postsynaptic D1 receptors. Such a physiological mechanism implies that FMRI may be used to track symptoms related to NAcc dopaminergic dysregulation in psychiatric disorders including schizophrenia and attention deficit/hyperactivity disorder.").

78. A synapse is the gap between two neurons. Generally speaking, neurons communicate via the release of chemical neurotransmitters into this gap, where they are detected by receptors and re-interpreted into electrical or intracellular signals.

79. *See* Nikos K. Logothetis et al., *Neurophysiological Investigation of the Basis of the fMRI Signal*, 412 NATURE 150, 154 (2001); Nikos K. Logothetis, *The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal*, 23 J. NEUROSCIENCE 3963, 3969 (2003) ("Simultaneous fMRI and electrophysiological recordings suggest that the BOLD contrast mechanism directly reflects the neural responses elicited by a stimulus. . . . [A]ctivation in an area is often likely to reflect the incoming input and the local processing in a given area rather than the spiking activity. Although it is reasonable to expect that output activity will usually correlate with neurotransmitter release and presynaptic and postsynaptic currents, when input into a particular area plays what is primarily a modulatory role, fMRI experiments may reveal activation in areas in which physiological experiments find no single-unit activity.").

80. *See* Raichle & Mintun, *supra* note 71, at 452 ("Despite the centrality of blood-flow

that astrocytes, a non-neuronal cell type in the brain that provides physical scaffolding and support functions for neurons, may also be involved in the BOLD response.[81] Even more recently, another study reported increases in local blood flow in anticipation of, but without actual, neuronal firing, throwing into question how tight the relationship really is.[82] Each new finding about the BOLD response suggests that our understanding of its neurological basis and correlation to brain activity is just scratching the surface. Understanding this phenomenon at the level of the neuron is critical to understanding if it is capturing little, some, or most of the brain's actual neuronal activity in response to an event. In turn, this knowledge is necessary to bridge the gap between a particular cognition or behavior and the neural mechanism underlying it.

4. *The semantics of "activation"*

Most researchers, including the authors, use the word "activation" to refer to a change in the BOLD signal. This shorthand, however, may be misinterpreted. A change in the BOLD signal might be correlated to greater or lesser neuronal firing, and "activation" when referring to the BOLD response does not necessarily mean "on" or "excitation" in the neuronal sense. Indeed, the brain is always "on" and the neuronal networks therein are constantly in some state of activity, unless they are dead due to trauma. Our brains use a significant amount of energy even when we are in a resting state with our eyes closed.[83] "Activation," taken out of context, may inaccurately imply that unless

---

changes to the imaging signals we observe with PET and fMRI, the complexity of the relationship of blood flow and metabolism to the underlying cellular events has only recently become more fully appreciated.").

81. *See* James Schummers, Hongbo Yu & Mriganka Sur, *Tuned Responses of Astrocytes and Their Influence on Hemodynamic Signals in the Visual Cortex*, 320 SCIENCE 1638, 1638 (2008) ("Though astrocytes are the major class of nonneuronal cell in the brain, their role in brain function is unresolved. . . . [A]strocytes do respond to neural activity in vivo, but fundamental questions about the relationship between neuronal networks, astrocytes, and hemodynamic responses remain unsolved.").

82. *See* Sirotin & Das, *supra* note 61. In two awake monkeys, blood flow increased in the visual cortex in anticipation of a visual stimulus that the monkeys had been trained to expect to appear. This increase in blood flow, not coupled with neuronal activity, appears to be mediated by an unknown preparatory mechanism. *Id.* at 475. A review published alongside the study observed why this hitherto unexpected response may be problematic for the interpretation of fMRI studies:

> For one thing, most fMRI experiments involve the periodic presentation of sensory stimuli, and then rely on the temporal structure of the haemodynamic response for deducing local neural activity. The present study clearly demonstrates that some of the assumptions underlying such analysis—namely, that cyclical variations in blood flow reflect local, stimulus-driven events—may sometimes be incorrect.

David A. Leopold, *Pre-Emptive Blood Flow*, 457 NATURE 387, 388 (2009).

83. Because the BOLD response refers to a contrast between two states, when only the default networks are engaged (i.e., no specific task is being completed), there is no contrast against which the BOLD response can be measured. Marcus E. Raichle et al., *A Default Mode of Brain Function*, 98 PROC. NAT'L ACAD. SCI. 676, 682 (2001) ("[W]hen an

a BOLD response is observed, the brain is "off." Another confusion arising from labeling the BOLD response "activation" is that it is not clear whether the BOLD response captures excitatory and inhibitory firings or whether it confuses the two.[84] And although inhibition is still a neuronal "action," the limitations of language imply that activation somehow means a positive, rather than negative, feedback loop.[85]

B. *The "Function" of Functional Imaging: Task-Dependency and Behavior*

In cognitive neuroscience research, the construction of an fMRI image typically begins with a question about cognition, emotion, or perception. An experimental task is designed to attempt isolation of the relevant psychological process that is being assessed.[86] The performance of this task constitutes the

---

individual is awake and alert and yet not actively engaged in an attention-demanding task, a default state of brain activity exists . . . . [A]n appreciation of the important activities that may underlie the baseline state of the human brain will certainly enrich our understanding of its function.").

84. György Buzsáki et al., *Inhibition and Brain Work*, 56 NEURON 771, 772 (2007) ("These data were interpreted to raise doubts about the possibility that GABA-mediated inhibition will be reflected in changes in the fMRI BOLD signal.").

85. *See* Robert F. Ackermann et al., *Increased Glucose Metabolism During Long-Duration Recurrent Inhibition of Hippocampal Pyramidal Cells*, 4 J. NEUROSCIENCE 251, 251, 261 (1984) ("[S]everal fundamental issues remain unclear, among them the role that neuronal inhibition plays in regional energy demand."); *id.* ("[F]uture research utilizing techniques having greater resolution, together with concomitant and independent measures of physiological function, is needed to allow confident attribution of alterations in glucose metabolism to specific synaptic mechanisms." (citation omitted)). One research team has posited that this scientific discrepancy may be explained by inhibition and questions whether, with an observable increase in inhibitory firings, we might observe glucose metabolism, increases in blood, and oxygen consumption moving in parallel. *See* Buzsáki et al., *supra* note 84, at 773 ("To date, the lack of rigorous experiments prevents one from answering this question.").

86. The fMRI device is gathering data on the subject's brain while the subject is performing the task. There are many possibilities for behavioral tasks. Cognitive neuroscience researchers often rely on widely-used cognitive tasks such as the Stroop Task, the Go/No-Go task, and the Stop Task to allow for points of comparison between labs. Using the same battery of behavioral tasks is useful for this comparative purpose. However, legal paradigms may require the development of new tasks that better model the specific behavior of interest. Major contributions to neuropsychological research are often made by those who question assumptions inherent in commonly used tasks.

The widely used Go/No-Go task measures inhibitory control. Subjects have to press either the left or right response button according to the direction of arrows that are presented on a screen in front of them. Infrequently, arrows pointing upward ("no-go" signals) appear. In these trials, subjects have to inhibit their motor response. *See* Brian A. Nosek & Mahzarin R. Banaji, *The Go/No-Go Association Task*, 19 SOC. COGNITION 625, 627 (2001); Katya Rubia et al., *Tryptophan Depletion Reduces Right Inferior Prefrontal Activation During Response Inhibition in Fast, Event-Related fMRI*, 179 PSYCHOPHARMACOLOGY 791, 793-94 (2005). The stop task was designed to assess inhibition, and it is often used in diagnosing clinical disorders such as ADHD and schizophrenia. *See* James D. Carter et al., *Assessing Inhibitory Control: A Revised Approach to the Stop Signal Task*, 6 J. ATTENTION DISORDERS subject's behavior; the resulting captured data and constructed images are interpreted as the neural correlates of that behavior. fMRI experiments conducted with the intention of introducing them as evidence of mental states are likely to be similar to such studies.

As in the research context, the utility of the fMRI data in a forensic context is highly task-dependent. Emily Bell and Eric Racine recently reiterated this dependency: "Since fMRI operates under the premise that measured brain activity is a reflection of areas of the brain engaged in or involved in carrying out a task, conventional fMRI and the resulting brain images are bound to the context of the task."[87]

Task design is influenced both by theories and previous studies as well as by practical constraints: the task has to be something that can be reasonably done within the scanner.[88] Thus, limitations on interpretation and generalization to real-life behavior are at least two-fold: the physical constraints imposed by the scanner environment and the constraints inherent in tasks focused on a discrete behavioral or psychological process. A major critique of using fMRI for forensic purposes is that the behavior being solicited in response to the task is usually so isolated that the results are difficult to generalize to other real-world functions. We will discuss this problem of ecological validity under our legal analysis of probative value below.

A third limitation in interpretation is added if the fMRI data are being offered as evidence of mens rea. Such use would require a close relationship between the chosen behavioral task and the specific legal question of interest. Validity problems arise here in trying to operationalize legal terms of art such as "intent" or "capacity" to be assessed in discrete behavioral tasks, including appropriate control conditions. For example, there could be a healthy debate surrounding the selection of the appropriate behavioral task for assessing *capacity* to form the specific intent to kill. Even well-validated behavioral tasks, such as those used to assess executive function,[89] do not directly map on

---

153, 153 (2003).

87. *See* Bell & Racine, *supra* note 42, at 23.

88. fMRI does not presently work with uncooperative subjects. While not all fMRI tasks require an instrumental behavior (such as the press of a button in response to a stimulus), they all require active participation in the minimal sense of the subject's paying attention and keeping still. The subject could very easily deviate from the control group or make the scan unusable just by keeping his eyes closed, moving slightly, or letting himself fall asleep (which frequently happens to a healthy control during long scanning sessions). *See* John A. Detre, *Clinical Applicability of Functional MRI*, 23 J. MAGNETIC RESONANCE IMAGING 808, 811 (2006).

89. *See* B.J. Casey et al., *A Developmental Functional MRI Study of Prefrontal Activation During Performance of a Go-No-Go Task*, 9 J. COGNITIVE NEUROSCIENCE 835, 836 (1997); Akira Miyake et al., *The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis*, 41 COGNITIVE PSYCHOL. 49, 92 (2000); Jennifer A. Richeson et al., *An fMRI Investigation of the Impact of Interracial Contact on Executive Function*, 6 NATURE NEUROSCIENCE 1323, 1323 (2003).

to Model Penal Code constructs such as purpose, knowledge, recklessness, or negligence. At best, deficiencies in task performance and correlated abnormalities in brain activity are several inferential steps away from a legal lack of intent or lack of capacity. The task-dependency of fMRI may be obscure to the fact finder, especially when confronted with images of a defendant's brain, leading this critical component of data interpretation to be left unexamined. Bell and Racine summarize these concerns as a warning: "The selection of the appropriate task for drawing out relevant information about brain activity in fMRI investigations should not be neglected in haste to explain behaviors and pathological differences by functional neuroimaging data."[90]

C. *Variables in Data Collection, Processing, and Analysis*

Creating a brain activation map from the raw data requires expertise from the fields of magnetic resonance physics, neurophysiology, cognitive neuroscience, and statistics. The complex method of constructing the fMRI image is critical to the evidentiary analysis.

1. *Hardware and software: the scanner*

Magnetic resonance imaging scanners are very large and expensive pieces of equipment. When installed, the massive magnet must be appropriately shielded for the protection of the device as well as for the surrounding electromagnetic environment. Scanners are also rather finicky machines—they must be carefully calibrated by expert technicians on a regular basis. Even so, scanners have their own idiosyncrasies, and no two scanners can be expected to produce exactly the same set of functional imaging data, even under identical parameters.

The quality and resolution of the fMRI images will vary depending on the strength of the magnet.[91] While the subject is performing the behavioral task, the raw data are acquired through a program that runs on the fMRI's computer.[92] This "pulse sequence program" directs the switching of current in the magnet through gradient coils, the transmission of radio frequency, and the ultimate collection of data with microsecond accuracy.[93] The program captures the signal that is generated at points in space and represented as a time-series of

90. *See* Bell & Racine, *supra* note 42, at 24.

91. *See, e.g.*, Press Release, Univ. of Mich., A Powerful Picture: Two New 3 Tesla MRI Scanners at U-M Health System Will Take Metabolic Imaging, Research to the Next Level (Apr. 22, 2004), http://www.med.umich.edu/opm/newspage/2004/mri.htm; News Release, Univ. of Pa., *Penn Researchers to Get 7 Tesla Whole-Body MRI System* (Aug. 28, 2006), http://www.uphs.upenn.edu/news/News_Releases/aug06/7TMRI.htm.

92. *See* SARTY, *supra* note 72, at 2.

93. *See id.*

cubed units called voxels.[94] These voxels hold information across multiple slices of the brain. The volume of each voxel is correlated to the data's precision. The greater the voxel volume, the lower the spatial resolution.[95]

Spatial resolution is inversely proportionate to temporal resolution and sensitivity. Researchers often do not look for effects across the entire brain, but instead select a region of interest ("ROI") very carefully to focus on brain regions where they have some reason, based on psychological constructs, to believe that the area may be implicated in the activity being measured. Computer programs take the spatial data and reconstruct it into an image using multiple regression statistical techniques and mathematical modeling.[96] This process is often invisible even to the researcher, as it is done by software installed on the fMRI machine.

2. *Processing the raw data*

At this point, the time-series of images is ready for "preprocessing" to clean up the images and remove random effects. Because the fMRI machine is so sensitive and the physiological signal so tiny, the raw data include quite a bit of background noise (signals thought to be unrelated to the experimental BOLD response). The ratio of background noise to activation signal is described by the Signal-to-Noise Ratio (SNR).[97] To amplify this ratio as much as possible, some noise is removed through preprocessing.[98]

---

94. A voxel is a three-dimensional unit of space, similar to a pixel in two dimensions. The voxel volume and the statistical threshold at which the analysis is set will alter the resulting image immensely. *See* Jerzy Bodurka et al., *Mapping the MRI Voxel Volume in Which Thermal Noise Matches Physiological Noise—Implications for fMRI*, 34 NEUROIMAGE 542, 542 (2007).

95. *See* F. DuBois Bowman et al., *Statistical Approaches to Functional Neuroimaging Data*, 17 NEUROIMAGING CLINICS N. AM. 441, 442 (2007). The amount of imaging time required increases in inverse proportion to the sixth power of a voxel dimension. For example, in order to reduce the voxel size from 3.5 mm to 1 mm on each side, the acquisition time must be increased by ten minutes for each excitation measurement (controlling for the signal-to-noise ratio and scanner used). Joseph B. Mandeville & Bruce R. Rosen, *Functional MRI*, *in* BRAIN MAPPING, *supra* note 69, at 315, 321.

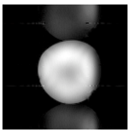
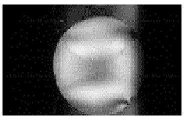
96. *See* Nick F. Ramsey et al., *Functional MRI Experiments: Acquisition, Analysis, and Interpretation of Data*, 12 EUR. NEUROPSYCHOPHARMACOLOGY 517, 521 (2002).

97. *See* Kevin Murphy et al., *How Long to Scan? The Relationship Between fMRI Temporal Signal to Noise Ratio and Necessary Scan Duration*, 34 NEUROIMAGE 565, 565-66 (2007).

98. One example of this is the production of "ghost artifacts," which occur when the image is blurred at the top or bottom. Ghost artifacts such as this one arise when the signal being received from a voxel has an alternating amplitude or phase. Examples of: (a) non-overlapping and (b) overlapping ghost artifacts are below. Images are from Stuart Clare, Functional MRI: Methods and Applications 80 (Oct. 1997) (unpublished Ph.D. dissertation, University of Nottingham), *available at* http://users.fmrib.ox.ac.uk/~stuart/thesis (used with author's permission).

Noise can be created from variations in the scanner's magnetic field and power, in the cognitive strategies of the subject, and in the head motion, swallowing, heartbeat, time in the scanner, and brain architecture of subjects. Studies are designed to increase the SNR and therefore reduce the temporal and spatial noise. One mechanical way to decrease the noise is to use a scanner with a stronger magnet. Another strategy to decrease noise is to average a subject's results over several trials to reduce the effects of random noise, on the theory that a "real" signal will be robust across most or all of the trials and thus survive the averaging. Training the subject to attend to focused stimuli and to remain still may also reduce background noise. However, in order to produce a usable image of the activation and region of interest, the noise often needs to be further smoothed out using complex statistical models.[99]

Once the slices of time-series data sets are transformed into spatial images and the images are correctly oriented in time and space, the individual brain data is then "warped" or normalized to map on to a 3-D template brain structure. A significant point here is that individual brain anatomy varies significantly, particularly in cortical (outermost) areas.[100] It can be very difficult to determine, based on anatomical landmarks, which parts of the brain in one person correspond to the standardized anatomical regions on the "average" structural brain image. With some stretching and realignment, the conformation of structural and functional assessments allows for comparisons between studies and for the pooling of data from different subjects.[101]

(a)  (b) 

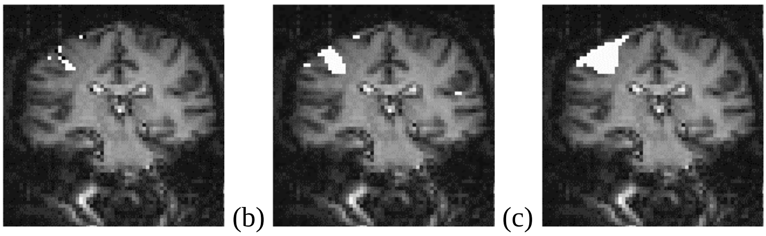
99. *See* Michal Mikl et al., *Effects of Spatial Smoothing on fMRI Group Inferences*, 26 MAGNETIC RESONANCE IMAGING 490, 490 (2008); Patrick L. Purdon & Robert M. Weisskoff, *Effect of Temporal Autocorrelation Due to Physiological Noise and Stimulus Paradigm on Voxel-Level False-Positive Rates in fMRI*, 6 HUMAN BRAIN MAPPING 239, 239 (1998); Amy Scouten et al., *Spatial Resolution, Signal-to-Noise Ratio, and Smoothing in Multi-Subject Functional MRI Studies*, 30 NEUROIMAGE 787, 787 (2006); Christina Triantafyllou et al., *Effect of Spatial Smoothing on Physiological Noise in High-Resolution fMRI*, 32 NEUROIMAGE 551, 551 (2006); *see also* Chloe Hutton et al., *Image Distortion Correction in fMRI: A Quantitative Evaluation*, 16 NEUROIMAGE 217, 217 (2002).

100. *See, e.g.*, Gabriele Lohmann & D. Yves von Cramon, *Automatic Labeling of the Human Cortical Surface Using Sulcal Basins*, 4 MED. IMAGE ANALYSIS 179, 179 (2000) (explaining that one of the most "intriguing problems in [the field of human brain mapping] is the high interpersonal variability of human neuroanatomy which makes studies across many subjects difficult"). In the field of brain mapping, work is being conducted to create useful brain atlases that can be used across individuals with less stretching and distortion than has been the norm. *See, e.g.*, David W. Shattuck et al., *Construction of a 3D Probabilistic Atlas of Human Cortical Structures*, 39 NEUROIMAGE 1064, 1064 (2008) ("[S]ingle subject atlases cannot describe the variability in brain structure that is inherent across the human population. To capture this information, larger numbers of subjects must be examined.").

101. Warping is discussed *infra* Part III.D.2 as a possible problem for the admissibility

Often, spatial smoothing is performed to mitigate the effects of the substantial differences in brain anatomy between subjects. Spatial smoothing refers to the methods employed to blur sharp spikes from the BOLD frequency signals. Some researchers discourage this step because it may inflate the correlations between neighboring voxels and reduce the strength of highly localized effects.[102] Below is an example of the effects of spatial smoothing on a constructed brain image, where the same fMRI data set is subjected to three different spatial smoothing filters. Brain regions that correlate well to the task appear in white, with the same statistical threshold, or p-value, used for all three maps. In the first image on the left, no filtering was applied, in the second, moderate filtering, and, in the third, even more filtering.[103] It is easy to see how the activation appears greater with greater filtering.

Figure 1



(a) (b) (c)

In addition to spatial smoothing, images can be temporally smoothed.[104] This sometimes needs to be done to make sure the functional data accurately tracks the underlying physiological changes over time.[105] Many researchers employ high-pass filtering to remove low frequencies of signal that were identified as likely being the product of random noise or signal drift. Signal drift is often caused by the natural and constant flow of cerebral spinal fluid

of functional neuroimaging data. *See* Bowman et al., *supra* note 95, at 443.

102. *See* RICHARD B. BUXTON, INTRODUCTION TO FUNCTIONAL MAGNETIC RESONANCE IMAGING: PRINCIPLES AND TECHNIQUES 280-82 (2002); Jiongjiong Wang et al., *To Smooth or Not to Smooth? ROC Analysis of Perfusion fMRI Data*, 23 MAGNETIC RESONANCE IMAGING 75, 80 (2005).

103. Image (b) used 7 mm full width at half maximum ("FWHM") filtering, and image (c) employed 15 mm FWHM filtering. In image (a), no filtering was applied. Clare, *supra* note 98, at 115.

104. There is a delay of the peak BOLD response by five to eight seconds due to the gap in time between the neural activity and the brain's blood flow response. There is also a transient increase of blood flow within the first 1000 milliseconds of neuronal activity. *See* Bowman et al., *supra* note 95, at 444.

105. *See* Karl J. Friston, *Statistics I: Experimental Design and Statistical Parametric Mapping*, *in* BRAIN MAPPING, *supra* note 69, at 605, 619-29; Arthur W. Toga & Paul Thompson, *An Introduction to Brain Warping*, *in* BRAIN WARPING 1, 3-5 (Arthur W. Toga ed., 1999); *see also* John Ashburner & Karl J. Friston, *Spatial Normalization*, *in* BRAIN WARPING, *supra*, at 27, 30.

through the hollow cavities deep in the brain.

Despite the well-developed methods for removing such artifacts, there are some means of noise production in the fMRI machine that cannot be completely removed.[106] While the above corrections are typically thought of as *removing* image distortion, the lack of consensus on proper methodology provides an opportunity for manipulation.[107] One expert expressed concern that correction tools can be thought of as "'black boxes' to be trusted implicitly and utilized without attention to their underlying assumptions and inherent limitations."[108] This concern over the lack of standardized methods will be revisited when we analyze the admissibility of fMRI as evidence in Part III.D.5.

Once the raw data are scrubbed up, the activation map must be created.[109] While there is a vigorous ongoing discussion on the topic, currently no single set of best methods for fMRI data analysis exists for a given research question, and therefore no single standard software package to use.[110] Despite the ubiquity of functional neuroimaging, only a small fraction of such research is devoted to standardizing functional neuroimaging methods.[111] Researchers may employ any one of a handful of statistical modeling tools and software packages to create the activation maps, depending in part upon their research question, their familiarity with different options, and what is available to them

106. *See* Chloe Hutton et al., *Effect of Head Motion and Non-Linear Distortions on fMRI Time Series*, 11 NEUROIMAGE S495, S495 (2000) ("This effect can be corrected by characterising the inhomogeneities with a field map and converting this into a map of pixel shifts. . . . Correction of the resulting time-varying distortions may require the generation of a distortion map at each time point, which involves phase unwrapping and image processing techniques that can be laborious and introduce additional noise."); João M. Sanches et al., *Medical Image Noise Reduction Using the Sylvester–Lyapunov Equation*, 17 IEEE TRANSACTIONS ON IMAGE PROCESSING 1522, 1522 (2008) ("Noise reduction in medical images is a difficult task in which linear filtering algorithms usually fail . . . . This paper presents a Bayesian denoising algorithm . . . .").

107. *See* R. Todd Constable, *Challenges in fMRI and Its Limitations*, *in* FUNCTIONAL MRI: BASIC PRINCIPLES AND CLINICAL APPLICATIONS 75, 77 (Scott H. Faro & Feroze B. Mohamed eds., 2006).

108. *See Preface* to BRAIN WARPING, *supra* note 105, at xi.

109. Bowman et al., *supra* note 95, at 443-44.

110. *See* R. Todd Constable et al., *Quantifying and Comparing Region-of-Interest Activation Patterns in Functional Brain MR Imaging: Methodology Considerations*, 16 MAGNETIC RESONANCE IMAGING 289, 289 (1998); Robert W. Cox & James S. Hyde, *Software Tools for Analysis and Visualization of fMRI Data*, 10 NMR IN BIOMEDICINE 171, 175-77 (1997); Wang et al., *supra* note 102, at 75. Through personal communication with Adina Roskies, we have learned that this is something that is being discussed in the fields of neuroscience and bioinformatics.

111. *See* Richard Baumgartner et al., *Are Global Methods Appropriate for fMRI Data Analysis? An In Vivo fMRI Study of the Spatio-temporal Heterogeneity of fMRI Data*, 2002 PROC. OF THE 2002 IEEE CANADIAN CONF. ON ELEC. AND COMPUTER ENG'G 894, 894; Bowman et al., *supra* note 95, at 441.

in their facility.[112]

3. *Individual differences and reliance on the group data*

a. *Something with which to compare: defining "normal"*

While most fMRI research relies on group data sets, the law cares a great deal about the individual and her unique function. In order to make comparative claims about whether an individual has abnormal neurological functioning, a testifying expert must have a basis for comparison—preferably a measurement of what would be considered normal function. To achieve this, data from many subjects and many conditions must be collected. Activation maps from a control group are often added together and then averaged. The normal group map is then compared to the activation map of the individual or group of individuals being tested (which itself is often the result of repeated trials that are averaged).[113] Differences between the individual's averaged data and the group's averaged data is then used to make a claim that the individual is either normal or a certain degree outside of normal with respect to this task.[114]

This discretionary creation of the control map and the subjective definition of normal may cause quite a bit of trouble with respect to the interpretation of the averaged data. Although it may be possible to test *how many* subjects are needed to be included in this control group to present an accurate and reliable brain activation map of the normal population, this sampling frame is not often

---

112. Typically the statistical inferences target the voxel level, or a particular region of interest in the brain. This is based upon some assumptions about where the researcher expects to see the activation. However, nonparametric methods may also be employed, as they do not require knowledge or assumptions about the underlying probability distributions for activation in the brain. For a discussion of statistical methodologies and their benefits and weaknesses, see Christian F. Beckmann et al., *General Multilevel Linear Modeling for Group Analysis in fMRI*, 20 NEUROIMAGE 1052, 1052 (2003); Bowman et al., *supra* note 95, at 454.

113. *See* SARTY, *supra* note 72, at 69.

114. Positron Emission Tomography ("PET"), an older form of functional neuroimaging, uses the subtraction method to compare the baseline state with the experimental state, either across subjects or within a subject. This method is still widely used with fMRI today. The subtraction method measures the BOLD response of an individual in a given state and then subtracts the BOLD response present in a control map. The difference is presented in the ultimate image. This technique may assume that the experimental condition of interest merely adds mental processing (and thus brain processing) that is not present during the control state. It also implies that there are no interactions among various mental components of a task. This assumption may be inaccurate or false, but it does not render the data useless in the research context. *See* Interview, *Marcus E. Raichle*, 8 J. COGNITIVE NEUROSCIENCE 189, 191 (1996). Subtraction as a technique produces information that can be "very useful" for comparisons between discrete tasks and between carefully defined groups. Edson Amaro Jr. & Gareth J. Barker, *Study Design in fMRI: Basic Principles*, 60 BRAIN & COGNITION 220, 223 (2006).

done.[115] The number of subjects in the control group is often selected based on practical concerns such as how much money and scanning time has been allocated to the project, or on experimental bases such as the size and known function of the region of interest. In addition to the problem of control group size, another significant issue involves which, or whose, brains are classified as normal. In theory, someone could be included in the control group of normals because he has never performed poorly on psychiatric evaluations, and yet that person could have abnormal brain activity due to his reliance on different cognitive strategies.[116] Because all of the results are dumped into the mix and averaged, after-the-fact review may not be adequate and may not catch the fact that a member of the control group lies significantly outside the normal mean of the rest of the group and thus may skew the results. We will say more about this problem in Part III.F.1.

b. *Individual differences are important but are often ignored*

Individuals, in addition to having differences in their brain anatomy, often exhibit significant differences in the way they respond to a given stimulus. However, because group data for both the control map and the experimental map typically focus on *shared* neural activations, differences between subjects may be regarded as statistical noise and eliminated.[117] This is deliberate in the research setting, as the mission of cognitive neuroscience is to discover the elements of normal brain function that are common to most people and disrupted in specific patient populations. However, this may be highly problematic in a forensic and individualized legal context.

There are major interpretive problems when individual differences are

115. The number of subjects included in the control group may affect the results. For a particular task, a group of researchers found that about fifteen to twenty subjects were required in order to reliably detect all of the activated regions in the left hemisphere. About thirty to thirty-five subjects were needed to reveal the weaker and more variable activation in right hemisphere regions. *See* Mohamed L. Seghier et al., *Group Analysis and the Subject Factor in Functional Magnetic Resonance Imaging: Analysis of Fifty Right-Handed Healthy Subjects in a Semantic Language Task*, 29 HUMAN BRAIN MAPPING 461, 472 (2008) ("A greater N-sub is required (>30 subjects) when one's objective is also to investigate patients with atypical language representation and where activations are found in areas that are not dominantly involved in healthy controls.").

116. Indeed, a recent report demonstrated that, in the context of gender differences, similar behavioral performance can correspond to significantly different patterns of brain activation; the reverse was also demonstrated, however, as similar patterns of activation could be observed despite variance in behavioral performance. *See* Emily C. Bell et al., *Males and Females Differ in Brain Activation During Cognitive Tasks*, 30 NEUROIMAGE 529, 529 (2006) ("[D]espite these fMRI changes, there were no significant differences between males and females on cognitive performance of the task. . . . In contrast, in the spatial attention task, men performed better . . . than women, but there were no significant functional differences between the two groups.").

117. *See* Seghier et al., *supra* note 115, at 462.

ignored in tasks that involve emotional processing, as variance between individuals in activation patterns may be the rule rather than the exception.[118] Two researchers reviewing the literature on individual differences found that "[r]egional brain activity associated with emotion processing can be influenced by a range of individual differences, including differences in personality, dispositional affect, biological sex, and genotype."[119] Even cutting-edge science still has a crude idea of what "normal" means as captured by group data. So far, researchers have found evidence that state-dependent factors such as sleep deprivation and caffeine intake affect the BOLD response, but the universe of relevant factors is still being populated.[120] Without knowing more about what modulates the BOLD response, comparisons based on averaged group data are not very useful for individual determinations. Michael Miller and his research team have argued that the exclusive reliance on statistical analyses of groups of subjects "may be to the detriment of understanding the true underlying cognitive nature of brain activations."[121] In their study, Miller and co-workers found that information that emerges from the group pattern does not reflect the reliable but non-overlapping patterns from individuals.[122]

The reliance of fMRI methods on averaging activation across individual subjects presents a problem for their utility in criminal law and individualized justice. The figure below makes this point nicely. Each pinhead circle represents the area with the most significant BOLD response for a particular task. The nine individual scans are to the left of the resulting group-averaged map. Interestingly, the group map reveals the most significant activation in a location in the brain where *none* of the individual brains were most active, though a few were close.[123]

118. *See generally* Stephan Hamann & Turhan Canli, *Individual Differences in Emotion Processing*, 14 CURRENT OPINION IN NEUROBIOLOGY 233 (2004).

119. *Id.* at 236.

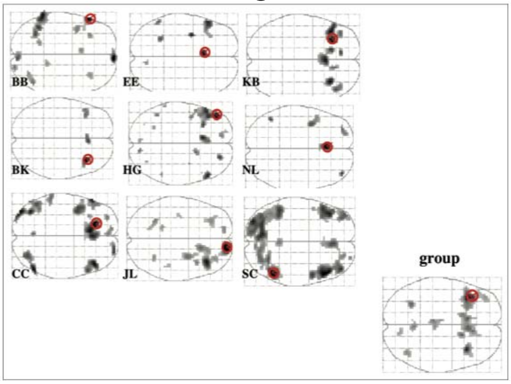
120. *See* Paul J. Laurienti et al., *Dietary Caffeine Consumption Modulates fMRI Measures*, 17 NEUROIMAGE 751, 751 (2002) (showing that BOLD response "in visual cortex was significantly greater" in high caffeine users relative to low caffeine users); Robert Joseph Thomas & Kenneth Kwong, *Modafinil Activates Cortical and Subcortical Sites in the Sleep-Deprived State*, 29 SLEEP 1471, 1475 (2006) (finding that "[s]leep deprivation results in an overall reduction in activation, especially in frontal and prefrontal areas").

121. Michael B. Miller et al., *Extensive Individual Differences in Brain Activations Associated with Episodic Retrieval Are Reliable Over Time*, 14 J. COGNITIVE NEUROSCIENCE 1200, 1200 (2002).

122. *See id.* at 1209.

123. *Id.* at 1203.

Figure 2



To reinforce the problematic nature of averaging, a recent study attempted to pin down exactly how much signal variance was due to individual differences in a semantic language task.[124] To visualize the variability between subjects, the percentage of overlap of each activated voxel was measured across subjects. In order to account for anatomical variation, the method looked to some of the neighboring voxels as well. The researchers found that the activated brain regions varied considerably across subjects in size, localization, and the levels of activation.[125] Of even greater interest, this research team discovered that the results for the same task varied greatly depending on three important and experimenter-controlled variables: (1) the statistical threshold used; (2) the software package employed; and (3) the size of the control population. The potential for manipulating each will be discussed in turn.

4. *Variance: the statistical threshold can be manipulated to affect the results*

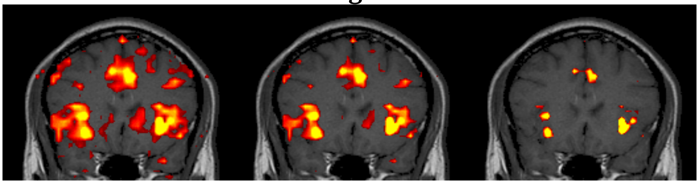
Which of the brain images below shows the greatest level of activation?[126]

124. Subjects were asked to identify whether two nouns belonged in the same semantic family.

125. *See* Seghier et al., *supra* note 115, at 466-67.

126. Images of statistical threshold differences are used here with the permission of Scott Grafton of the University of California, Santa Barbara. From left to right: $P < 0.05$ (1682 voxels), $P < 0.01$ (364 voxels), $P < 0.001$ (32 voxels).

Figure 3



The truth is, these images represent *the same group of brains performing the same task*. Each could be argued to be "statistically significant," in the sense that the likelihood that the result happened due to chance is less than 0.05 in each. The only difference is that the statistical threshold is tighter, and thus the voxel volume is lower, as you move from left to right. Without some background information about the chosen statistical threshold, an individual could infer very different things about how abnormal an activation pattern is.[127] If you were to zoom out enough, only the most significant activation could be seen, while if you zoom in closely, everything may appear to be active. Statistical thresholds can therefore be manipulated to do what neurologist Helen Mayberg has called "dial a defect."[128] If a party does not like the results that are shown at a certain level of zoom, simply altering the statistical precision may provide a more compelling image for one's legal argument.

5. *Variance: the statistical analysis employed can affect the results*

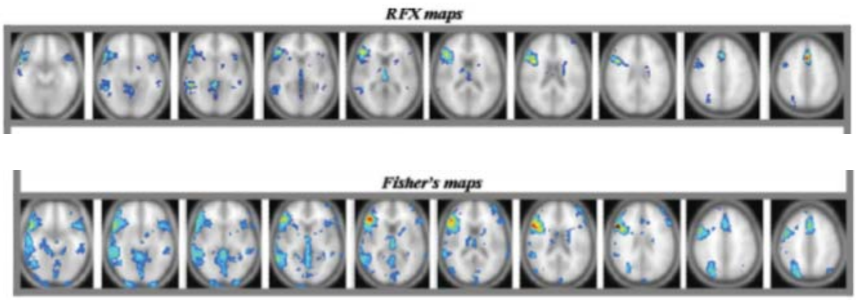
Just as the statistical threshold can affect the resulting activation map, so too can the statistical analysis or test employed. In the figure below, you can see that the map at the top using one method, RFX, is the most restrictive and produces the least activation for this semantic task, while *the same data*

127. While a more lenient statistical threshold may be desirable when conducting more exploratory research, the use of a lenient standard in clinical or forensic diagnosis is much more suspect. *See* Michael B. Miller & John Darrel Van Horn, *Individual Variability in Brain Activations Associated with Episodic Retrieval: A Role for Large-Scale Databases*, 63 INT'L J. PSYCHOPHYSIOLOGY 205, 213 (2007).

128. Helen S. Mayberg spoke about the manipulation of imaging studies at a presentation to the MacArthur Foundation on Law & Neuroscience members at its annual meeting in Santa Barbara, California on May 29, 2008. Of course, the same was said of PET, as Joseph Dumit noted in his book, PICTURING PERSONHOOD, the quantitative data set itself is dynamic and always imperfectly represented visually. The layers of construction making up the image can literally "make it up." DUMIT, *supra* note 63, at 69. David Faigman aptly points out that the problem by which statistical analysis is selected by an expert plagues all applied science: "Applied science is almost invariably probabilistic and so cannot be used adequately without knowledge of probabilities and statistics; [j]udges regularly rely on applied science as an integral part of lawmaking; therefore, it is incumbent on judges to understand probabilities and statistics." David L. Faigman, *Judges as 'Amateur Scientists,'* 86 B.U. L. REV. 1207, 1207 (2006).

analyzed by a different statistical procedure produce a brain that appears more active in the left hemisphere.[129] A good lab will retest its data using different statistical tests, but selecting the analysis provides another opportunity for manipulation.

Figure 4



None of this is to say that skilled researchers deliberately employ such methods to manipulate the interpretation of the data or to dupe peer reviewers. In fact, depending on the purpose of the study, a less restrictive analysis and statistical threshold may be desirable. However, what troubles us is the ability of forensic labs or paid experts to take advantage of these effects to "dial a defect" in the courtroom. Unless standardized and transparent criteria are followed to process the data and construct the activation map, the procedures employed could be subject to distortion. Furthermore, none of these concerns are at all mitigated by the oft-employed claim that data analysis is "computerized"—a virtually meaningless assertion meant to soothe fact finders that such analysis could not possibly be incorrect, as it was not subject to human error. In fact, as we have shown, there are a number of subjective decisions that must be made during image acquisition, preprocessing, and analysis, and the use of a computer does not immunize the process from distortion. Computerized analysis, in this context, should be interpreted by fact finders with skepticism, as it may be a panacea meant to discourage critical questioning about the choices made that underlie the final production of the neuroimages.

Now that we have discussed at length the gathering, processing, and production of group neuroimages, we will walk through the evidentiary steps

129. For our purposes it is not important to understand the specific differences in the methodology. However, it is important to note that researchers are likely not deliberately manipulating the data so much as they are employing various methods for various research purposes. While the RFX analysis may be "suitable for detecting true positive activations with [a] high degree of confidence," the less restrictive analyses may be "particularly useful for appreciating all activated regions of interest during normative database formation." Seghier et al., *supra* note 115, at 472.

required for admissibility, relying on the scientific framework that we have just presented. At various points in the evidentiary analysis it might be useful to refer back to specific methods and concepts. Where possible we will include the section headings to facilitate easy reference.

III. LEGAL ANALYSIS

A. *Admissibility Is Specific to the Evidentiary Purpose*

Admissibility cannot be decided in a vacuum; the first inquiry must be how the party introducing the evidence proposes that it be used. The same evidence may be extremely probative for one purpose and not even relevant for another.[130] As discussed above, this Article focuses on functional neuroimages that are introduced to provide evidence of an individual's mens rea. As such, much of the analysis of the evidentiary requirements could be applied to other applications of fMRI such as in lie detection or assessment of subjective pain. However, this analysis sets aside structural brain images—such as those used to prove harm in a tort action, or those used to validate a relevant medical diagnosis when imaging is a component of the medical standard of care.

In order to be admissible, functional brain images: must be relevant, authenticated, and more probative than unfairly prejudicial; cannot be impermissible hearsay;[131] and, if introduced by an expert, must satisfy the tests for expert scientific testimony in that jurisdiction. In short, there are many hurdles to admissibility. We focus here on the requirements that the evidence be relevant, authenticated, and that the probative value be greater than the potential for unfair prejudice. Even though states have different rules of evidence, each state has some equivalent to Federal Rules of Evidence 401 and 403, which speak, respectively, to relevance and excluding relevant evidence that is unfairly prejudicial.[132] In this Article, we will take advantage of the

130. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 591 (1993) ("The study of the phases of the moon, for example, may provide valid scientific 'knowledge' about whether a certain night was dark, and if darkness is a fact in issue, the knowledge will assist the trier of fact. However (absent creditable grounds supporting such a link), evidence that the moon was full on a certain night will not assist the trier of fact in determining whether an individual was unusually likely to have behaved irrationally on that night.").

131. The hearsay rule applies to "computer-generated evidence which repeats or contains human declarations. Evidence to which this hearsay rule may apply includes accounting records, invoices, summaries or any other types of computer output which reiterate human declarations which have been inputted into the computer." 57 AM. JUR. 3D *Proof of Facts* § 7 (2009).

132. FRE 401 states, "'Relevant evidence' means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." FED. R. EVID. 401. FRE 403 states, "Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless

broad overlap between states on these two rules, and will generally reference the standards for admissibility as codified in the Federal Rules of Evidence (FRE).

B. *Classifying Functional Brain Images*

Evidence can be either direct or circumstantial, and it may come in the form of either demonstrative (real) evidence or witness testimony. Under many evidentiary rules, the classification of evidence does not do much work. But as fMRI challenges the traditional boundaries between categories of evidence, it is worth discussing how the resulting images might be classified.[133] As used in this Article, fMRI images are tangible objects that are often used to illustrate testimony, and are not *as yet* themselves direct proof of the defendant's capacity to form intent or otherwise hold a particular criminal mental state. They are therefore not direct evidence, and for now they are not likely to constitute testimonial evidence.[134] However, if the behavioral task involves the subject testifying to some guilty knowledge or participation in a criminal act, one wonders whether the images might one day be thought of as testimonial. If the methods for producing the images involve the subject declaring or testifying to something as part of the experimental stimuli, the resulting image may very well be categorized as a declaration or statement.[135] If this occurs,

presentation of cumulative evidence." FED. R. EVID. 403.

133. Many thanks to David Faigman for emphasizing the importance of these classifications and the distinctions.

134. Erich Taylor, Note, *A New Wave of Police Interrogation? 'Brain Fingerprinting,' The Constitutional Privilege Against Self-Incrimination, and Hearsay Jurisprudence*, 2006 U. ILL. J.L. TECH. & POL'Y 287, 303 ("The Court has repeatedly considered four factors when determining whether a practice is testimonial in nature: (1) whether the practice 'enlists the mind' of the subject; (2) how the practice 'compares to paradigms of real and testimonial evidence;' (3) how the practice adheres to the values and principles behind the Fifth Amendment; and (4) what implications banning the practice will have on law enforcement practices and procedures."); *see also* Sarah E. Stoller & Paul Root Wolpe, *Emerging Neurotechnologies for Lie Detection and the Fifth Amendment*, 33 AM. J.L. & MED. 359, 366 (2007). Nita Farahany has presented on the potential for brain-derived data to be considered testimonial evidence. *See* Audio recording: Nita Farahany, Incriminating Thoughts, Stanford Law School Junior Scholars in Law and Neuroscience Conference (Apr. 7, 2008), *available at* http://www.law.stanford.edu/display/images/dynamic/events_media/20080405_CLB_FarahanKolber.qtl.

135. Though this technique will undoubtedly be further refined, at present it has been successful in predicting mental states in highly constrained conditions related to visual perception (e.g., asking the subject to recall an image of a hammer or an igloo), and simple intention formation, such as whether the subject decided to add or subtract two numbers. *See* Kendrick N. Kay et al., *Identifying Natural Images from Human Brain Activity*, 452 NATURE 352 (2008); Yoichi Miyawaki et al., *Visual Image Reconstruction from Human Brain Activity Using a Combination of Multiscale Local Image Decoders*, 60 NEURON 915 (2008); Svetlana V. Shinkareva et al., *Using fMRI Brain Activation to Identify Cognitive States Associated with Perception of Tools and Dwellings*, 3 PLoS ONE e1394 (2008); *see also* John-Dylan Haynes et al., *Reading Hidden Intentions in the Human Brain*, 17 CURRENT

the images generated by testimonial statements will also implicate the hearsay doctrine and the Fifth Amendment protection against self-incrimination.[136] For now, however, as the functional images cannot decode thoughts and cannot be used to coerce testimony or reliably reveal someone's "true thoughts," it will be classified for our purpose as circumstantial, demonstrative evidence.[137]

It is analytically important to differentiate between the methodology that the expert relies upon and the interpretation of that methodology. This is true even though the Supreme Court has stated that there must be some nexus between the scientific expert's interpretation and the data and methodology on

---

BIOLOGY 323 (2007); Chun Siong Soon et al., *Unconscious Determinants of Free Decisions in the Human Brain*, 11 NATURE NEUROSCIENCE 543 (2008). The latter study focused on "free" decisions about pressing a right or left button, and found that the outcome of the binary decision was predictable from brain activity well before the decision was reported to have entered conscious awareness. The neuroscientific basis of "free will" is a topic for another manuscript. Furthermore, the construction of such pattern classifiers requires many repeated trials, and the pattern may be idiosyncratic to a particular individual, such that one classifier would not be useful as a general forensic tool unless there is a universal signal in response to the stimuli that is incredibly robust and specific. However, given that visual memory for faces produces a very robust and specific signal in the brain, with sufficient knowledge of the facts of the crime (e.g., exact murder weapon, details about a unique location), it is conceivable that pattern classifiers may be able to predict whether the defendant has ever seen the evidence put before him in the fMRI scanner. *See* John-Dylan Haynes & Geraint Rees, *Decoding Mental States from Brain Activity in Humans*, 7 NATURE REVS. NEUROSCIENCE 523, 528 (2006).

One important caveat is that the pattern classifier at present cannot differentiate between sources of encoding the memory. What this means is that it will not work if the object being tested for recognition triggers a response because it resembles an unrelated object the subject has seen before (such as a similar model of gun or a crime scene that the subject has seen before in a different context). Query whether merely recognizing something indicates anything about whether the defendant committed the crime. To date, pattern classifiers cannot differentiate between knowing something because one has heard it or read it, versus knowing something because one actually experienced it. This severely limits the forensic application of pattern classifiers.

136. If the behavioral measure used to trigger the BOLD response requires making a statement, and that statement is made by someone who cannot testify in court, one wonders whether fMRI might also implicate hearsay rules. *See* Taylor, *supra* note 134, at 306 (questioning whether "hearsay considerations would become an issue with BF [brain-fingerprinting] testing . . . if the test results of one suspect were used in an attempt to prosecute another suspect, possibly a co-conspirator. In such an instance, would the BF test results of the first suspect be admissible as against the second suspect? If BF test results are found to be testimonial statements, they would only be admissible in such an instance if the individual making the statements was 'unavailable' to testify and had previously been subject to cross examination on the test results. Alternatively, the test results could be admitted under any of the recognized exceptions to the general rule of hearsay inadmissibility." (citation omitted)).

137. Demonstrative evidence appeals to the jurors' senses: it is something they can touch, smell, taste, or hear. Circumstantial evidence is evidence that requires the jury to draw an inference, and which on its own cannot be dispositive of the prima facie elements of a case. *See* CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE UNDER THE RULES 50 (2008).

which it rests.[138] In this Article, we are primarily concerned with the presentation of the images themselves, regardless of whether the problems we address would likewise challenge the reliability and analytical validity of the expert's testimony interpreting the images. Practically speaking, functional neuroimages would almost never be introduced without being interpreted by a testifying expert. However, as the images could be thought of as generating independent substance rather than merely illustrating preexisting testimony or data, the presumption that they are demonstrative evidence may not be robust. We will elaborate on this point when discussing the authentication of the images *infra* in Part III.D.

C. *Relevance*

The point of departure for analyzing admissibility of any type of evidence is relevance. Federal Rule of Evidence 402 simply states that only relevant evidence is admissible, and irrelevant evidence is not admissible.[139] Relevance has two parts: materiality and probative value.

The word *materiality* is often used to describe the first prong of relevance, which is the fit between the evidence and the case.[140] The materiality component of relevance reflects the idea that there is no such thing as relevance in the abstract.[141] If brain images were introduced to suggest that the defendant suffered from an irresistible impulse, this would be immaterial in a state such as California that does not recognize the irresistible impulse test for legal insanity. However, so long as the functional brain images are introduced to bolster or challenge some fact that is of consequence under the criminal statute of that state, they will satisfy the materiality requirement.

The second prong of logical relevance is probative value. Quite simply, probative value refers to the tendency of evidence to establish the proposition for which it is offered. Probativeness can also be referred to by another name,

---

138. In *Daubert v. Merrell Dow Pharmaceuticals*, the Supreme Court stated that the "focus, of course, must be solely on principles and methodology, not on the conclusions that they generate." 509 U.S. 579, 595 (1993). *But see* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) ("[C]onclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.").

139. FED. R. EVID. 402.

140. This can be confusing, as unlike some states, the federal rules do not require that evidence be introduced to prove a fact that is in dispute or material to the *outcome*. Materiality is used here to refer to the requirement that there be a nexus between the proffered evidence and the substantive law.

141. *See* GEORGE FISHER, EVIDENCE 18 (2d ed. 2008). Therefore if knowledge is *not* an element of a criminal statute, then a defense based on lack of knowledge (for example, that a previous conviction was punishable by more than one year, and thus the defendant should not have been carrying a firearm) is immaterial.

logical relevance, as it "exists as a relation between an item of evidence and a proposition sought to be proved."[142] In order for a piece of evidence to carry any probative value, there must be a sound reason for relying on it and a probabilistic connection between the evidence and the proposition it seeks to prove. The Federal Rules of Evidence only require that the proffered evidence have *any* tendency to make the existence of any fact of consequence more or less probable.[143] Under the FRE, the evidence could be relevant if it merely helped the jury put some other critical element in context.[144]

Recent legal scholarship has proposed new models for calculating probative value. Some of these call for mathematical representations of the proposition sought to be proved and the probability of proving that proposition.[145] While mathematical models of probative value provide conceptual clarity, they are not very practical:[146] even if judges employ probabilistic models for evaluating evidence, they tend to do so without conscious access to this process.[147]

---

142. George F. James, *Relevancy, Probability, and the Law*, 29 CAL. L. REV. 689, 690 (1941).

143. FED. R. EVID. 401.

144. As we saw in the *Saviñon* case discussed *infra* Part I, functional neuroimages would likely be introduced to bolster other psychological evidence—i.e., psychological testing, behavioral data, and eyewitness testimony. In that particular case, the defense counsel conceded that the image could not prove anything about the criminal's mental state at the time of the crime. Instead, the expert would only use the image as supporting evidence for the psychological theory that the defendant was depressed. *See* Defendant's Offer of Proof and Points and Authorities Re: SPECT Evidence, *supra* note 5, at 5. While evidence may be deemed relevant if it puts something else in context, it still needs to clear other evidentiary hurdles to be admitted.

145. *See* Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 108 (2007).

146. There are a few conceptual models for measuring whether an item has probative value. One model asks whether the probability of the hypothesis (H) is affected by some piece of evidence, (E)—i.e., P(H|E). A second model reverses this inquiry and asks whether the existence of the evidence is affected by an assumed hypothesis P(E|H). To put this into our terms, the first formulation asks whether the probability that an individual was unable to form the requisite intent is more or less likely given the existence of data in the functional brain image. The second formulation asks whether the interpretation of the fMRI data is more or less likely if the individual lacked the ability to inform intent. However, if we already have some reason to believe that the defendant lacked the ability to form intent, then our job in the scanner is a bit easier. The probability that the image reflects some reduced functioning is greater if we think there is a functional impairment a priori, and can build upon existing knowledge of cognitive and behavioral networks to model the experimental task. Because the fMRI image is only as good as the experimental task, having a behavioral task that is tailored to the individual's unique deficit will be easier than throwing darts at the entire frontal lobe and hoping to see a BOLD response that may have nothing to do with the ability to form intent. This is important, as the legal requirement of intent to kill is not a unique phenomenon; there may be many physiological events leading to reduced ability to form a particular type of intent. So the probability of E, given H, is greater than the probability of H, given E. *See* A.W.F. EDWARDS, LIKELIHOOD 44-51 (expanded ed. 1992).

147. A judge typically relies on her "personal experience, general knowledge, and

1. *Logical inference and relationships between brain data and mental states*

The probative value of functional neuroimaging varies depending on the exclusiveness of the relationship between the brain data and the mental state of interest. With respect to complex mental states and cognitive functions, there is virtually no one-to-one mapping of a particular function to a particular brain region. If one sees reduced activity in the frontal lobe (and a corresponding difference in the BOLD response compared to a group of normals) this does not conclusively mean that this individual lacked the ability to form a specific intent. Such a finding certainly *could* implicate mens rea, but the probative value is weak given the many other possible explanations for the result. Without knowing anything about an individual's ability to form the relevant mens rea, or about the precise networks recruited for a particular mental state, a variant BOLD response could mean any number of things, including nothing.[148]

To date, most research investigating the neurological correlates of cognitive processes has relied on "forward inference." With the forward inference approach, researchers ask subjects to perform two experimental tasks

---

understanding of human conduct and motivation." MCCORMICK ON EVIDENCE 736 (Kenneth S. Broun ed., 6th ed. 2006); *see also* D.H. Kaye, Comment, *Quantifying Probative Value*, 66 B.U. L. REV. 761, 766 (1986) ("The search for an interpretation of PV [probative value] should be guided by the uses to which such an expression will be put. Without a quantitative measure of prejudicial effect and the other counterweights of Rule 403, I do not see how a mathematical expression of PV could find direct forensic application. The only purpose that I can see is heuristic. A suitable formulation for PV may clarify our thinking about what it means to say that evidence is very probative, slightly probative, and so on. These rough quantifications are useful in performing the balancing required under Rule 403."). One day, perhaps neuroscience will shed light on this issue as well. *See* Jeffrey M. Beck et al., *Probabilistic Population Codes for Bayesian Decision Making*, 60 NEURON 1142, 1142 (2008) (the paper presents a "neural model of decision making that can perform both evidence accumulation and action selection optimally").

148. As top neuroimaging researcher Steven Petersen explains, "The problem right now with imaging is that doing experiments right is really, really hard, but getting pictures out is really, really easy." Greg Miller, *Growing Pains for fMRI*, 320 SCIENCE 1412, 1412 (2008). Elizabeth Phelps, a cognitive neuroscientist at New York University, adds that although anxiety engages the amygdala, so do intense smells, sexually arousing images, and many other things. *Id.*

This critical point has been elaborated on by Roskies, *supra* note 21, at 24, who states that "[w]hen considering the steps that allow us to draw conclusions about neural activity from imaging data, it becomes evident that a large number of neural states could conceivably give rise to the same signal, and furthermore, we currently lack means of ruling out many of those possibilities as improbable." While this was discussed above, it bears repeating here. Possible alternative explanations include: (1) the behavioral task used in the fMRI is not well correlated with the individual's specific ability to form intent; (2) the differences within the group may be too great, so that the averaged activation pattern does not reflect data that correspond to any one individual; (3) the BOLD response is not capturing reduced functioning, or put differently, the reduced "activation" in the defendant might be the result of efficient processing, increased inhibitory firings, or novel cognitive strategies.

that are thought to differ only in their engagement of a particular mental process. As a crude example, in one condition the subject may be asked to read aloud words aloud that are presented on a screen, and in a second condition he is instructed to silently read the same words on the same screen. If the two tasks differ in their relative BOLD response, it may be inferred that the non-overlapping BOLD activation corresponds with brain regions that are necessary for orally producing, rather than merely visually processing, words. There are problems with the forward inference, as there are many examples of scenarios where regions that are activated during one task are not necessary *and* sufficient for completing that task.

Also, if the task has become highly automatic or if it has been repeated upon many trials, efficient processing may lead to relatively little BOLD activity, even though that region *is* involved in the task.[149] In this case the inference might be that the individual is deficient in her functioning, when in fact she is either on autopilot or has developed expertise. The opposite may also be true, where subjects may recruit additional brain regions if the task is difficult—brain regions that assist in, but that are not normally required for, completion of the task.[150] It is therefore presently impossible to assign anatomical specificity or functional necessity to a region or network in the brain through fMRI alone. In order to make a claim that a brain region is not only specifically recruited for a task, but also functionally required, researchers will need to rely on other techniques that can measure the functional impact of manipulating healthy brain anatomy.[151]

While there are interpretive problems with the forward inference approach, reverse inference reasoning is more troubling. The reverse inference from activation to mental state reflects the logical fallacy of affirming the consequent. In a series of influential articles, cognitive scientist Russell Poldrack lays out this common, but flawed, reasoning.[152] It typically proceeds

149. *See* John Darrell Van Horn & Russell A. Poldrack, *Functional MRI at the Crossroads*, 73 INT'L. J. OF PSYCHOPHYSIOLOGY 3, 4 (2009). "Repetition suppression" is a known phenomenon wherein repeated processing of the same stimulus results in decreasing neural and associated hemodynamic activity. For a review, see R.N.A. Henson & M.D. Rugg, *Neural Response Suppression, Haemodynamic Repetition Effects, and Behavioural Priming*, 41 NEUROPSYCHOLOGIA 263 (2003).

150. *See* Roberto Cabeza et al., *Task-Independent and Task-Specific Age Effects on Brain Activity During Working Memory, Visual Attention and Episodic Retrival*, 14 CEREBRAL CORTEX 364, 373 (2004) ("[T]ask-specific age effects included age-related contralateral recruitments in left PFC during WM and in right PFC during VA. This result suggests that older adults may compensate for deficits in production processes by recruiting monitoring processes, and vice versa.").

151. *See* Van Horn & Poldrack, *supra* note 149, at 4.

152. *See* Russell A. Poldrack, *The Role of fMRI in Cognitive Neuroscience: Where Do We Stand?*, 18 CURRENT OPINION NEUROBIOLOGY 223, 223-24 (2008) [hereinafter Poldrack, *Role*]; *see also* Russell A. Poldrack, *Can Cognitive Processes Be Inferred from Neuroimaging Data?*, 10 TRENDS COGNITIVE SCI. 59, 63 (2006) [hereinafter Poldrack, *Cognitive Processes*].

as follows:

> In the present study, when task comparison A was presented, brain area Z was active. (If A, then Z; or, when chocolate shown, the brain region called the nucleus accumbens is active).
>
> In other studies, when cognitive process X was putatively engaged, then brain area Z was active. (If X, then Z; or, when reward processing is thought to be engaged, the nucleus accumbens is active).
>
> Thus, the activity of area Z in the present study demonstrates the engagement of cognitive process X. (Thus, if Z, when A, then X; or, because the nucleus accumbens is active when chocolate is shown, reward processing must be engaged).

The logical failure is fairly obvious, as the brain region called the nucleus accumbens may be involved in numerous other mental processes that are not specific to signaling reward. Typical reverse inference claims are deductively true only if the *specific* mental process activates the region of interest, but "brain regions observed with fMRI are rarely activated by only one mental process."[153] Reverse inference is relatively common in cognitive neuroscience research articles, where the inference is understood as explaining unexpected findings or developing novel hypotheses to be tested in subsequent experiments.[154] However, reverse inferences may be abused in the courtroom, where the faulty logic is not exposed, and where findings are not being used to bolster future research but rather to make factual or legal claims.[155]

Areas of the brain that are likely to show up in mens-rea-related claims, specifically components of the prefrontal cortex, parietal association cortex, and subcortical components such as the amygdala and hippocampus, are among the best studied yet most complex areas of the human brain. These areas are responsive to a huge variety of stimuli and are engaged in nearly the full spectrum of human behavior. As discussed above, it is well understood with respect to complex perceptual and behavioral functions that the relationships within the human brain between structure and function are not one-to-one. While practicing neuroscientists tend to strongly believe that every mental state must be encoded as a brain state, we are at present a considerable distance

153. Poldrack, *Role*, *supra* note 152, at 223-24 ("It has become increasingly common to use neuroimaging data to infer the presence of specific mental processes . . . . This approach has been particularly common in newer literatures such as neuroeconomics and social cognitive neuroscience, where the fundamental processes underlying task performance are often unknown . . . . However, given that the goal of cognitive neuroscientists is to build explanations rather than deductive laws, reverse inferences may provide some useful and important information even if they are not deductively valid.").

154. Poldrack, *Cognitive Processes*, *supra* note 152, at 63.

155. *Id.* at 60 ("If a [brain] region is activated by a large number of cognitive processes, then activation in that region provides relatively weak evidence of the engagement of the cognitive process; conversely, if the region is activated relatively selectively by the specific process of interest, then one can infer with substantial confidence that the process is engaged given activation in the region.").

away from the precise mapping of complex mental states onto unique brain activation patterns. Knowing that structure and function do not map onto each other in a one-to-one relationship, and given the problems with the forward and reverse inferences in the forensic context, why would anyone ever think that an fMRI directly proves anything about an individual's mental state? At least two alternative avenues exist: First, the images may lend credibility to other psychological theories of the defendant's behavior. Second, the image might be thought of as superior to the status quo of relying on eyewitness or direct testimony.

Given what is known about the unreliability of eyewitness testimony and what may be suspected about the truthfulness of direct testimony, it may be tempting to call upon the more reductionist and seemingly objective brain images to draw inferences about a defendant's mental state.[156] Although brain images may *appear* more scientific and less capable of distortion, as demonstrated in Part II, it is this *appearance*, and not the validity of the science, that parties expect to do the persuading. When used to make inferences of mens rea, fMRI images have even less probative value than the status quo of psychological testing and testimony as to criminal behavior. But given that the construction of the functional brain image requires the researcher to take multiple discretionary steps—steps that are not obvious in the resulting image—the results appear more reliable than they are. This was referred to previously as the "epistemic mismatch."[157] In order to expose the discretionary steps, counsel proffering functional neuroimages should carefully authenticate the images and lay a foundation for their introduction. Over time the authentication process becomes more automated, but particularly with a new technology, the process needs to be meticulously articulated in order make transparent the process behind the compelling images.

---

156. *See* Bethany Shelton, *Turning a Blind Eye to Justice: Kansas Courts Must Integrate Scientific Research Regarding Eyewitness Testimony into the Courtrooms*, 56 U. KAN. L. REV. 949, 954-55 (2008) ("Reid Hastie, a well-known researcher in the 1980s, studied juries and their perspective of eyewitness identifications. Hastie's work indicates that jurors misunderstand many factors that affect eyewitness accuracy. He concluded that juror misconceptions about eyewitness accuracy, combined with potentially inaccurate identifications, pose a serious threat to justice. Today, Dr. Elizabeth Loftus is the most recognized expert on eyewitness testimony. She has participated in many criminal trials where eyewitness identification played a major role in the case against the defendant. Research has led Dr. Loftus to conclude that, depending on the circumstances, eyewitness identification can be highly unreliable and riddled with error. . . . Following in the footsteps of Dr. Loftus and Hastie, researchers continue to discover that juries overestimate the accuracy of eyewitness identification." (citations omitted)); *see also* Cindy O'Hagan, *When Seeing Is Not Believing: The Case for Eyewitness Expert Testimony*, 81 GEO. L.J. 741, 748-49 (1993); Brooke Whisonant Patterson, *The "Tyranny of the Eyewitness,"* 28 LAW & PSYCHOL. REV. 195, 202 (2004).

157. *See supra* note 21.

D. *Authentication*

Before demonstrative evidence can be admitted into the record, the proponent must first offer sufficient proof that it is what she claims it to be. This is done to make sure the evidence has not been adulterated and that it is the product of sound scientific principles.[158] Nothing should be presumed to be authentic, and the proponent bears the burden of establishing authenticity.[159] This standard is embedded in FRE 901(a)[160] and resembles the conditional relevance standard from FRE 104(b).[161] Like conditional relevance, authentication allows for the admission of evidence based upon the fulfillment of some condition of fact (here, authenticity).[162] To satisfy authentication requirements, the proponent must offer enough proof that the jury could reasonably find by a preponderance of the evidence that the exhibit is what the proponent claims it is.[163] In jury trials, the judge will conduct the offer of proof outside of the hearing of the jury and will ask about the underlying methodology for producing the image. Only after an initial offer of proof is satisfied will the judge allow the jury to see the evidence. The jury will then make the ultimate decision as to the item's authenticity.

There is, at present, no standard for authenticating functional brain images. For the moment, analogies to the authentication requirements for more familiar types of evidence may be helpful. As fMRI represents statistical probabilities and requires inferences from a group to an individual, the authentication requirements for social science survey data may be instructive. In this case, the proponent must authenticate the survey data by demonstrating both that the general methodology produces an accurate result and that proper survey methodology was followed in the particular case.[164]

As the ultimate image looks like a still photograph, the authentication requirements for X-rays may be instructive.[165] However, as fMRI relies on

158. *See* EDWARD J. IMWINKELRIED & DANIEL D. BLINKA, CRIMINAL EVIDENTIARY FOUNDATIONS 113-15 (1997).

159. By way of example, before a crime scene photograph may be admitted, the proponent must offer *sufficient proof to enable a jury* to find that the photograph accurately depicts the crime scene in question. CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE UNDER THE RULES 977 (1st ed. 1988).

160. *See* FED. R. EVID. 901(a) ("The requirement of authentication or identification as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims.").

161. *See* FED. R. EVID. 104(b) ("When the relevancy of evidence depends upon the fulfillment of a condition of fact, the court shall admit it upon, or subject to, the introduction of evidence sufficient to support a finding of the fulfillment of the condition.").

162. FISHER, *supra* note 141, at 805-06.

163. *See id.* at 806 (citing Huddleston v. United States, 485 U.S. 681, 688 (1988)).

164. *See* Kevin H. Smith, *External Validity: Representativeness and Projectability in the Probative Value of Sample Surveys*, 39 WAYNE L. REV. 1433, 1445-46 (1993).

165. Christopher J. Buccafusco, *Gaining/Losing Perspective on the Law, or Keeping Visual Evidence in Perspective*, 58 U. MIAMI L. REV. 609, 627 (2004) ("Although the

computer programming and statistical models to produce the brain image, which is similar to the modeling of computer-generated animation videos, comparing the way computer-animated videos have been authenticated may be useful. Some aspects of fMRI will render the social science, X-ray, and computer animation examples inapposite, but this only illustrates the point that authentication requirements are not one-size-fits-all.[166]

1. *The pictorial and silent witness theory of admissibility may accommodate the authentication of fMRI images*

In most situations, a photograph needs to demonstrate "pictorial testimony," meaning that a witness needs to testify according to personal knowledge that the photograph accurately reflects the scene in question. The witness need not have taken the picture, nor does she need to know anything about the conditions at the time the picture was taken. She only needs personal familiarity with how the scene normally looks.

There is an alternative to this approach. The "silent witness theory" was developed to accommodate the situation that arises when no one can personally witness the object being photographed.[167] One such example is X-ray, as no one can testify from personal knowledge as to how bones or tissue normally appear. Instead of speaking to the accuracy of the *content*, the witness need only lay a sufficient foundation that the *process that produced the image is reliable* and capable of capturing whatever it is that the proponent claims it to be.[168] Thus in certain cases we are willing to assume authenticity of content if

---

acceptance of a new form of visual evidence provided a novel question for the jurists of the day, they were quickly able to develop coherent rules for the admission of photographic evidence by analogizing this new medium to more traditional types of evidence that had a long history of admission.").

166. Specifically, the empirical data from computer-generated videos runs the other way: it may be less prejudicial insofar as the images are produced according to an accurate program that does not distort the data—a big assumption. In an empirical study of the impact of computer-generated videos, subjects who viewed testimony with computer animation "recalled information more accurately and in more detail than participants who did not view animation . . . . Providing the animation with verbal narration reduces the processing demands on listeners' short-term memory and maximizes the likelihood of their successful and accurate encoding into long-term memory." Linda C. Morell, *New Technology: Experimental Research on the Influence of Computer-Animated Display on Jurors*, 28 SW. U. L. REV. 411, 414-15 (1999).

167. Not all states subscribe to the "silent witness" theory of authentication, though most do. *See, e.g.,* Dolan v. State, 743 So. 2d 544, 545 (Fla. 1999); State v. Arafat, No. 85847, 2006 WL 871720, at *7 (Ohio Ct. App. 2006); *see also* In the Matter of the Welfare of S.A.M., 570 N.W.2d 162, 165 (Minn. Ct. App. 1997). While many states have adopted this theory, and some have codified it into their rules of evidence, states like Georgia have not. *See* MINN. R. EVID. 901(b)(9); Ross v. State, 585 S.E.2d 666, 670 (Ga. Ct. App. 2003); CHARLES C. SCOTT, PHOTOGRAPHIC EVIDENCE § 800, at 727-29 (1942).

168. Wagner v. State, 707 So. 2d 827, 830 (Fla. Dist. Ct. App. 1998) ("Under the 'silent witness' theory, photographic evidence may be admitted upon proof of the reliability

the process is sound.[169]

For efficiency, we will fold the various components for authenticating evidence under the silent witness theory[170] into the following list, and apply it to fMRI: (1) the image must accurately capture the individual's brain under the same conditions that existed at the time of the crime; (2) the procedure for creating the image should be described in detail to remove any possibility of tampering, error, or distortion (this should include testimony as to the competence of the staff and the reliability of the equipment); and (3) the underlying statistical computer programs must demonstrate reliance on irrefutable scientific principles. Each component will be discussed below with reference to fMRI.

---

of the process which produced the photograph or videotape."); Dep't. of Pub. Safety & Corr. Servs. v. Cole, 672 A.2d 1115, 1119 (Md. 1996); Washington v. State, 943 A.2d 704, 711, 714 (Md. Ct. Spec. App. 2008) (finding that the theory is justified "because [the pictures] accurately represent what they purport to show" and "the photograph speaks with its own probative effect."). Today the silent witness theory for authentication provides another means of introducing visual and photographic evidence in virtually all jurisdictions. MICHAEL H. GRAHAM, FEDERAL RULES OF EVIDENCE IN A NUTSHELL § 401:7 (2010).

169. *See* Straughn v. State, 876 So. 2d 492, 502 (Ala. Crim. App. 2003). Indiana takes the theory a step further than most, and holds that once admitted under the silent witness theory, pictorial evidence may be considered substantive evidence, rather than merely demonstrative. *See* Shepherd v. State, 690 N.E.2d 318, 323 (Ind. Ct. App. 1997). The substantive and demonstrative divide is in many ways a distinction without a difference, however, as there are minimal practical consequences stemming from a designation of substantive as opposed to demonstrative. *See* Robert D. Brain & Daniel J. Broderick, *The Derivative Relevance of Demonstrative Evidence: Charting Its Proper Evidentiary Status*, 25 U.C. DAVIS L. REV. 957, 965-66 (1992) ("Some courts treat demonstrative exhibits exactly like they do substantive exhibits, by formally admitting them into evidence and allowing the jury to view the exhibits during deliberations. Other courts admit demonstrative exhibits into a twilight zone reserved for 'demonstrative purposes only,' apparently indicating that such exhibits can be identified for the record but must be precluded from use by the jury during deliberations.").

Even so, because the photograph itself cannot be cross-examined, most courts hold that there should be a strong showing of authenticity and reliability of the *process*. Specifically in the case of photographs, this may require "evidence as to how and when the camera was loaded, how frequently the camera was activated, when the photographs were taken, and the processing and chain of custody of the film after its removal from the camera." Edwards v. State, 762 N.E.2d 128, 136 (Ind. Ct. App. 2002).

170. This seven-prong test has been dubbed the "*Voudrie* standard" following *Voudrie v. State*, 387 So. 2d 248, 256 (Ala. Crim. App. 1980). It requires: (1) a showing that the device that produced the item being offered as evidence was capable of recording what a witness would have seen or heard; (2) a showing that the operator of the device was competent; (3) establishment of the authenticity and correctness of the resulting recording, photograph, videotape, etc.; (4) a showing that no distortions, additions, or deletions have been made; (5) a showing of the manner in which the recording, photograph, videotape, etc., was preserved, (6) identification of the speakers, or persons pictured; and (7) for criminal cases only, a showing that any statement made in the recording, tape, etc., was voluntarily made without any kind of coercion or improper inducement. *See also* Bryant v. State, 810 So. 2d 532, 536 (Fla. Dist. Ct. App. 2002). For our purposes we have simplified the substance into a three-part test as described in the text.

2. *Images must accurately capture the individual's brain under the same conditions that existed at the time of the crime*

The proponent must show that an X-ray film portrays the part of the body of the person whose condition is at issue in the case.[171] While this threshold element is usually fairly easy to meet, in the case of fMRI, the resulting brain image is *not* just a picture of the individual's brain at a single point in time. The proponent here just needs to reiterate that the image is not the activity of one individual's brain on its own at one time, and in fact it relies on comparison, data manipulation, and warping to templates.

The second common requirement is proof that "the physical condition of the subject at the time of being photographed [is] *the same as at the time in issue in the case*."[172] Here the proponent needs to testify that the subject sustained no injuries after the event in question and before the image is constructed. If a great deal of time has passed, this factor would typically go to weight rather than admissibility.[173] For our purposes, the proponent of fMRI images may satisfy this requirement by stating outright that the fMRI-produced image is not a picture of the brain at the time of the crime, but rather a current assessment of the subject under different circumstances that *may* be correlated with the past mental state. This distinction is critical and ought to dramatically affect the weight that the fact finder gives to the evidence. The jury could not review a picture of the subject's brain at the time of the crime, and there is no way to recreate the emotional reaction and cognitive processes at the time of the crime to assess contemporaneous mens rea.

3. *The procedure for creating the image should be described in detail to remove any possibility of tampering, error, or distortion*

Next, the proponent must show that the apparatus used in making the picture was of a type known to produce dependable results and was in "good working condition."[174] So long as the magnet and the computer equipment were working well and were sufficiently powerful to render accurate data, the proponent could easily clear this requirement. It must also be shown that the person operating the fMRI device was qualified through training and experience to use it,[175] that she followed proper procedures, and that the chain of custody of the resulting image was maintained.

The next authentication prong requires that the manner of creating the

---

171. *See* Call v. City of Burley, 62 P.2d 101, 107 (Idaho 1936).

172. SCOTT, *supra* note 167, § 799, at 724 (emphasis added); *see also* Dennison v. State, 66 So. 2d 552, 554 (Ala. 1953); Wilburn v. U.S. Gypsum Co., 60 P.2d 188 (Cal. Dist. Ct. App. 1936).

173. SCOTT, *supra* note 167, § 799, at 724.

174. *See* Stevens v. Ill. Cent. R.R., 137 N.E. 859, 862 (Ill. 1923).

175. *See id.*

image be described in detail. Specifically, any enhancements or distortions of the image must be disclosed.[176] Here the analogy to X-rays breaks down, as the production of X-ray images involves some, but not much, methodological discretion.[177] Conversely, pursuant to the discussion *infra* Part II, the final product of fMRI is greatly affected by idiosyncratic lab practices.[178] Even though these practices are routinely employed to clean up the data rather than distort it, the degree of methodological freedom invites manipulation.[179] Five of the greatest opportunities for manipulation are: (a) attempts to remove background noise in the data, (b) the selection of the experimental task, (c) the creation of the "control" group and the exclusion of control subjects whose data does not support a certain theory, (d) the chosen threshold of statistical significance, and (e) the statistical software package the lab employs.

The way to resolve the problem of what was earlier dubbed "dialing a defect" is to require a detailed explanation by the proponent of not only what was done to the raw data to produce the image, but also *why* each step was taken. This should be explained in the simplest way possible so that the fact finder can understand. While X-ray images provide helpful guidance as to how fMRI images should be authenticated, the minimal degree of discretion in the production of X-ray images requires reliance on another example. As there is a great deal of literature challenging the authentication of computer-generated videos, this type of evidence may direct the authentication process for fMRI

---

176. In the case of X-rays, courts required information about the distance from the X-ray to the subject, the separation between the film and the subject, the angle from which the X-rays were directed, and the length of exposure. Courts have also required testimony as to whether the film was artificially changed to bring into relief certain features. *See, e.g.*, Bartlesville Zinc Co. v. Fisher, 159 P. 476, 477 (Okla. 1916); *see also* Cooney v. Hughes, 34 N.E.2d 566, 569 (Ill. App. Ct. 1941); West v. Wilson, 4 P.2d 469, 471 (Mont. 1931); State v. Veatch, 740 N.W.2d 817, 826 (Neb. Ct. App. 2007) ("It must also be shown to the satisfaction of the trial court that no substantial change has taken place in the exhibit so as to render it misleading."); SCOTT, *supra* note 167, § 800, at 727.

177. To take an X-ray of the bones, short X-ray pulses are shot through a radiographic film, with the relevant body part in front of the film. The bones then absorb many photons as they are quite electron-dense. The X-rays that do not get absorbed turn the film from white to black, leaving a white silhouette of bone matter. Creating an image of the cardiovascular system requires a few extra steps, as an initial image is taken of the system of interest first. A second image is then taken of the same system after an iodine contrast material has been injected into the blood. The two images are digitally registered rendering one image that contrasts the blood vessels with the background physiology. The registration process does not involve the smoothing or statistical modeling that is required in fMRI.

178. This is true for PET images as well. "If [such] a [large] difference . . . can be obtained on the same subjects with the same injection using different tomographs, then apparent differences in metabolic rates reported by different groups should be interpreted with caution." Cheryl L. Grady, *Quantitative Comparison of Measurements of Cerebral Glucose Metabolic Rate Made with Two Positron Cameras*, 11 J. CEREBRAL BLOOD FLOW & METABOLISM A57, A63 (1991).

179. *See* Jill Witkowski, *Can Juries Really Believe What They See? New Foundational Requirements for the Authentication of Digital Images*, 10 WASH. U. J.L. & POL'Y 267, 269 n.53 (2002).

that also relies on complex statistical software.

4. *Underlying statistical computer programs must demonstrate reliance on irrefutable scientific principles*

The phrase "garbage in, garbage out" describes a key concern when dealing with computer-generated images. Unfortunately, juries may, in their fascination with a computer's ability to produce seemingly objective data, disregard this concern.[180] One legal scholar noted that even with increased visibility of computer technology, "there remains an air of mystery about the computer. Computerization is seen as a magic world, ruled by the wizards who operate the machines."[181] Of course, "computers do not think for themselves; they act on the commands of the programmer."[182] As was illustrated with the various fMRI processing and analytical tests, computers can produce very different functional images, depending on the data put into the computer and the programming commands supplied. As previously mentioned, "computerized analysis" should not be allowed to be a veneer of rigor and reliability, and instead should be questioned as critically as any other piece of evidence.

5. *Authentication should be specific to fMRI and distinct from other image types*

None of the authentication obstacles should be impossible for fMRI to overcome. But just as "an X-ray picture cannot be authenticated in the same manner as an ordinary photograph . . . [as] it requires a higher degree of authentication,"[183] the same is true for fMRI as compared to X-ray. Each new

180. Two examples of being over-awed by computers come from a Pennsylvania case and a recent law review article. *See* Commonwealth v. Serge, 896 A.2d 1170, 1176 n.3 (Pa. 2006) ("Because a [computer generated animation] is a graphic illustration of an expert's reconstruction rather than a simulation based upon scientific principles and computerized calculations, it is not subject to the *Frye* test governing the admissibility of scientific evidence in Pennsylvania."); Leo Kittay, *Admissibility of fMRI Lie Detection: The Cultural Bias Against "Mind Reading" Devices*, 72 BROOK. L. REV. 1351, 1355 (2007) ("[T]he fMRI appears less subject to examiner bias because the exam questions are presented visually on a screen and the analysis is performed using computer software.").

181. Mario Borelli, *The Computer as Advocate: An Approach to Computer-Generated Displays in the Courtroom*, 71 IND. L.J. 439, 439 (1996).

182. Carolyn Smart, *The Computer Must Be Right: Computer Generated Animations, Unfair Prejudice, and* Commonwealth v. Serge, 26 TEMP. J. SCI. TECH. & ENVTL. L. 387, 387 (2007).

183. The departure from the strict foundational requirements for video evidence is a product of "the judicial system's growing familiarity with video evidence, and the widespread social, cultural, and technological acceptance of the medium." JORDAN S. GRUBER, ELECTRONIC EVIDENCE § 8:9, at 386 (1995), *quoted in* Jill Witkowski, Note, *Can Juries Really Believe What They See? New Foundational Requirements for the*

generation of imaging device will require more sophisticated methods for authentication as the steps involved to create the image are obscured by the previous technology. As society becomes savvier as to the production methods for each new type of image or recording, the authentication requirements may become less strict, and judges may start to take judicial notice of the process. Such was the case with video recording and photography, which used to require much more thorough scrutiny for authentication than is typically required today.[184]

Turning from the past to the present, a final analogy comparing fMRI to computer-generated animation may help predict how functional images will be evaluated in the future. By way of background, parties often introduce computer-generated images in car accident cases in order to draw conclusions about how the accident likely occurred.[185] Recognizing the authoritative power of computer-generated images, many jurisdictions require that the underlying software and input data be tested for their ability to authentically model or predict the simulated events.[186] Computer-generated images have been classified as either "simulations" or "animation."[187] Animation evidence merely summarizes the expert's preexisting opinion in graphical form, and the

Authentication of Digital Images, 10 WASH. U. J.L. & POL'Y 267, 280 n.53 (2002).

184. "As time passed, the courts relaxed the requirements for admissibility of video evidence. . . . As long as recordings meet these four requirements, there is generally no need to adhere to the strict seven-part test. . . . In fact, strict foundational requirements for video recordings 'are now almost universally rejected as unnecessary.'" Witkowski, *supra* note 183, at 279, 280 & n.53 (quoting GRUBER, *supra* note 183, § 8:22, at 408).

185. In *Livingston v. Isuzu Motors, Ltd.*, a car accident case, the Montana Federal District Court admitted computer simulation evidence under FRE 702, but only after the plaintiff's expert testified at length to the foundation of the technology. Specifically, he explained the development and use of the simulation program, including how the physics equations were created, how the road conditions were modeled, etc. Livingston v. Isuzu Motors, Ltd., 910 F. Supp. 1473, 1495 (D. Mont. 1995); *see also* 57 AM. JUR. 3D *Proof of Facts* § 3 (2000).

186. *See* State v. Foreman, 954 A.2d 135, 158 n.24 (Conn. 2008); *see also* Schaeffer v. Gen. Motors Corp., 360 N.E.2d 1062, 1067 (Mass. 1977) ("Our concern is not with the precision of electronic calculations, but with the accuracy and completeness of the initial data and equations which are used as ingredients of the computer program."); IMWINKELRIED & BLINKA, *supra* note 158, at 104-06; Fred Galves, *Where the Not-So-Wild Things Are: Computers in the Courtroom, the Federal Rules of Evidence, and the Need for Institutional Reform and More Judicial Acceptance*, 13 HARV. J.L. & TECH. 161, 185 (2000).

187. Simulations require an expert to enter "mathematical formulae or other scientific principles into the computer so that the computer can generate a model—based on the data and scientific assumptions—that the expert will use to form an opinion as to *what must have or could have actually happened*." Galves, *supra* note 186, at 185 (emphasis added); *see also* Robert B. Bennett et al., *Seeing Is Believing; or Is It? An Empirical Study of Computer Simulations as Evidence*, 34 WAKE FOREST L. REV. 257, 260 (1999); Borelli, *supra* note 181, at 450-52.

In contrast, animations "are simply computer-generated drawings assembled frame by frame which, when viewed sequentially, produce the image of motion." Galves, *supra* note 186, at 180 (2000).

data that is put into the animation is based on this subjective opinion. Simulation evidence does more than animate testimony, and in fact provides independent, illustrative content. Because simulation evidence is thought to be more authoritative and objective, it often must meet a higher threshold for authentication.[188]

fMRI may be thought of as a hybrid between animation and simulation evidence, with the scales tipping slightly toward simulation evidence. fMRI provides the basis of the expert's opinion by applying neuropsychological models, laws of physics, and statistical principles in order to draw probabilistic conclusions about an individual's brain activity. In this way it resembles simulation evidence. However, as discussed in detail *infra* Part II, the image can be thought of as animating preexisting psychological theories of behavior and cognition. This is because the interpretation of the BOLD response will vary depending on the assumptions that are built into the computer software package and the theories behind the mental process being studied.[189] This makes fMRI a bit like animation evidence. Even though fMRI may be a hybrid form of evidence, it should require authentication of the scientific principles on which the image construction rests. It should also include an analysis of the reliability of that process under an FRE 702 analysis or its state counterpart. As standards develop in the field of neuroimaging, judges will begin to take judicial notice of the methodology for constructing the brain image, and the authentication process will become much more streamlined.

Before standardization occurs, the authentication of the expert evidence must involve some discussion of the methodology's reliability in order to satisfy the foundational requirement that the evidence is the product of sound scientific principles. While the *reliability* of the process may be unpacked a bit during the authentication phase, the scientific *validity* of the methodology and the interpretations that flow from it should be distinct from authentication. *Daubert* and its progeny have held that this distinction has little meaning, but it still serves a purpose in separating authentication from scientific expert testimony. The film "Willy Wonka and the Chocolate Factory" illustrates this point.[190] Recall the Wonkavision machine that mesmerized Mike Teevee by

188. *See* Fred Galves, *The Admissibility of 3-D Computer Animations Under the Federal Rules of Evidence and the California Evidence Code*, 36 SW. U. L. REV. 723, 726 (2008); Gregory P. Joseph, *A Simplified Approach to Computer-Generated Evidence and Animations*, 43 N.Y.L. SCH. L. REV. 875 (1999); Kurtis A. Kemper, *Admissibility of Computer-Generated Animation*, 111 A.L.R.5th 529, § 3 (2003).

189. *See* Dunkle v. State, 139 P.3d 228, 250-51 (Okla. Crim. App. 2006) ("Although Dalley's use of the computer-generated animations suggested that she was adding a computer-based analysis to the other evidence in the case, a careful review of her testimony reveals that she was simply restating evidence already introduced and re-summarizing areas in which various statements by Dunkle were inconsistent with this evidence.").

190. Wonkavision is a fictional process that is featured in the 1971 film adaptation of Roald Dahl's novel CHARLIE AND THE CHOCOLATE FACTORY. *See* WILLY WONKA & THE CHOCOLATE FACTORY (Warner Bros. 1971).

being able to transport a chocolate bar through a television. Authentication would not ask an expert to opine on whether each piece of chocolate tastes the same, or whether the Wonkavision system made all of the "right" decisions according to the chocolatiers of the world. It just asks the proponent to explain the process used to transmit the candy bars, to make sure that the Wonkavision mechanisms are grounded in sound scientific principles. Whether Wonkavision would stand up to peer review and falsifiability would be taken up later in great detail, when the process undergoes FRE 702 scrutiny. While this example is fanciful, computer simulation evidence is not, and it therefore provides a nice comparison for fMRI, where independent content is likewise generated by algorithms.[191] With computer simulations, jurists have found *Daubert* and authentication criteria to overlap, as the authentication process may need to inquire into the reliability of the methodology of producing the evidence, in order to make sure the programming assumptions are supported by scientific principles.[192]

Even though authentication may cross the threshold of FRE 702, "[t]he authentication standard does not demand the intricate, complex 'pavane' that *Daubert* seemingly entails."[193] With authentication, the judge must only deem the evidence sufficient for a jury to find by a preponderance of the evidence that the neuroimage captures what the proponent claims that it does.[194] Under the expert scientific evidence analysis, the interpretations that flow from the brain images would be heavily scrutinized to make sure that the method is either generally accepted for this purpose or that it is a product of, and subjected to, the scientific method. Authentication of the underlying scientific principles will not answer the question of whether the ultimate interpretation of the evidence is valid. Post-*Daubert* Supreme Court precedent has revealed that the nexus between the methods and conclusions should be reviewed under FRE 702.[195]

---

191. For a discussion of how scientific validity and authentication are related, see Dean A. Morande, *A Class of Their Own: Model Procedural Rules and Evidentiary Evaluation of Computer-Generated "Animations,"* 61 U. MIAMI L. REV. 1069, 1110 (2007) ("Although authentication does not generally involve a *Frye*/*Daubert* analysis, simulations present a unique set of circumstances: providing the jury with scientific, expert-like substantive evidence created by a non-human source.").

192. *See* State v. Kirsch, 820 A.2d 236, 243 (Conn. 2003); Bray v. Bi-State Dev. Corp., 949 S.W.2d 93, 97 (Mo. Ct. App. 1997).

193. Daniel D. Blinka, *Expert Testimony and the Relevancy Rule in the Age of* Daubert, 90 MARQ. L. REV. 173, 195 (2006).

194. "The judge, then, is not expected to mediate scientific disputes on the merits but rather to determine only whether the issues may be fairly disputed within the framework of the adversary trial." *Id.*

195. While *Daubert* drew a sharp line between methodology and conclusions, later Supreme Court precedent recognized that, under a *Daubert* or FRE 702 analysis, the distinction is not clean. *See supra* note 138. A later Supreme Court case followed up on this inability to cleave methods from conclusions. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 153-54 (1999) ("[I]t was the reasonableness of using such an approach, along with [the]

A real-world application of this theory arose in a murder case involving computer-enhanced bitemark photographs. The prosecution used Adobe Photoshop to superimpose the defendant's teeth on top of photographs of the victim's bitten flesh. On appeal, the defense successfully challenged the authentication of the computer-enhanced photographs, as the prosecution's expert had not demonstrated five elements. First, the process for creating the overlays was not demonstrated to have been accepted by the field of odontology. Second, there was insufficient evidence that the proper procedures were followed relating to the processing of data input and output. Third, the Adobe Photoshop software that was used to construct the image was not demonstrated to be reliable for this forensic application. Fourth, the equipment was not demonstrated to have been programmed correctly or administered by a competent technician. Finally, the court found that the Adobe Photoshop software might actually allow for the underlying bitemark photographs to be altered.[196] Note that this analysis is geared more toward reliability than validity, a critical distinction that the *Daubert* Court unfortunately bungled.[197]

For now, the proponent should not encounter much difficulty demonstrating the competence of the fMRI device or the process, so long as she is quite clear at each step of the way that the resulting data is correlative in nature and many idiosyncratic decisions can dramatically affect the constructed image. First, it should be made clear that fMRI does not produce a snapshot of the subject's brain.[198] Second, judges should know that fMRI also typically relies on group data to create a comparison to a "normal" standard. Otherwise the judge may not fully understand how the resulting image relies on data input

particular method of analyzing the data thereby obtained, to draw a conclusion regarding the particular matter to which the expert testimony was directly relevant." (emphasis omitted)). Lastly, a Third Circuit judge stated:

> [T]his distinction has only limited practical import. When a judge disagrees with the conclusions of an expert, it will generally be because he or she thinks that there is a mistake at some step in the investigative or reasoning process of that expert. If the judge thinks that the conclusions of some other expert are correct, it will likely be because the judge thinks that the methodology and reasoning process of the other expert are superior to those of the first expert.

In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 746 (3d Cir. 1994).

196. *See* State v. Swinton, 847 A.2d 921, 951 (Conn. 2004). For a similar case on authenticating computer-generated content, see *Commonwealth v. Serge*, 58 Pa. D. & C.4th 52, 59 (Ct. Com. Pl. 2001).

197. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 590 n.9 (1993) ("We note that scientists typically distinguish between 'validity' . . . and 'reliability' . . . . [O]ur reference here is to evidentiary reliability—that is, trustworthiness. . . . In a case involving scientific evidence, evidentiary reliability will be based upon scientific validity." (emphasis omitted)).

198. *Serge*, 58 Pa. D. & C.4th at 82-83 ("[I]n order to ensure that the jury does not confuse art with reality, an appropriate cautionary or limiting instructions [sic] should be provided . . . . [T]he jury should be reminded that '[a]n animation is only as good as the underlying testimony, physical data, and engineering assumptions that drive its images' and that '[t]he computer maxim "garbage in, garbage out" applies to computer animations.'" (quoting Clark v. Cantrell, 529 S.E.2d 528, 536 (S.C. 2000))).

from other volunteers. Third, in order to remove the possibility of tampering, a witness will need to testify to the procedures employed in analyzing the data and creating the image. In cases where the methodology is unique or particularly opaque, it may be necessary to validate the findings by having an independent lab conduct the same imaging procedure, or at least validate the analysis by subjecting the raw data to the same spatial smoothing and filtering techniques, statistical thresholds, and software package. Finally, once this is done, the image may be considered sufficiently authenticated. While authentication of this complex technology may seem incredibly burdensome to judges with busy schedules and full dockets, we provide a simple checklist in the Appendix as a starting point to authenticate functional neuroimages. While not exhaustive, the checklist may be a tool for unpacking the methodological elements a judge should inquire into before the fMRI data can be considered free from distortion.

E. *Why* Daubert, Frye, *and FRE 702 Should be Secondary Considerations After Rule 403*

For many legal practitioners and scholars, questions surrounding the admissibility of expert scientific evidence begin and end with an analysis of *Daubert* or *Frye*. While it would be impossible to discuss a new type of expert scientific evidence without starting with its relationship to the rules for expert scientific testimony under FRE 702, the inquiry should not end there. In order to push the discussion into a broader arena, we have focused the bulk of our inquiry on whether the images themselves, without any accompanying testimony, may be too unfairly prejudicial to be admitted to speak to mens rea. The first reason for this is that fMRI must be considered with respect to a threshold rule for *all* evidence, not just scientific evidence: FRE 403. Besides being legally necessary to the evidentiary analysis, FRE 403 does a better job of assessing the potential for judicial waste and confusion of the issues. It also responds better to arguments based upon convergent validity.[199] Moreover, FRE 403 provides a flexible standard closely tied to the purpose for which the evidence is being offered, whereas the outcome of a *Daubert*-type hearing

199. Convergent validity refers to the ability of a measurement scale to converge with other measures of the same variable. As such, it can capture the idea that there might be a point after which additional measurements are not cost-justified, and certain types of evidence might be privileged over others. David L. Faigman et al., *How Good Is Good Enough?: Expert Evidence Under* Daubert *and* Kumho, 50 CASE W. RES. L. REV. 645, 654 (2000) ("[W]e do endorse a better evidence principle in our analyses of certain issues in the law of expert testimony. That is to say, we believe that there are circumstances in which a court properly may exclude proffered evidence when other evidence of greater probative value is or should be available." (emphasis omitted)). To the extent that fMRI may one day possess convergent validity with other behavioral measures, or even be simply consistent with (and still reliant upon) a clinical diagnosis, FRE 403 could better deal with the prejudicial impact of wasting the court's time with expensive, convergent data.

could lead to sharper lines being drawn in the minds of courts and counsel that a technology is "admissible" or "inadmissible," even when that technology is still in flux. Even so, there is more than one way to dispose of fMRI addressed to past mental states, and many of the same arguments we will make below in our discussion of probative value could be used to challenge the scientific validity of fMRI addressed to past mental states. For this reason, we provide a brief overview of the evidentiary rules that are specific to expert scientific evidence.

The probative value of functional neuroimages will likely be established through interpretive expert testimony. In this situation, the judge may admit evidence subject to its connection with the testimony, or admit it upon proof of a proper foundation. The underlying expert testimony must meet the requirements of FRE 702 or its state counterpart.[200] FRE 702 requires the judge to determine whether the opinion of the expert is based on sufficient facts or data, whether it is the product of reliable principles and methods, and whether the witness has applied the principles and methods reliably to the facts of the case.[201] This rule is often referred to simply as the *Daubert* standard, following the Supreme Court case that clarified the federal rules.[202] Many states have adopted the essential elements of *Daubert*, either outright or indirectly. Those states that have not adopted *Daubert* typically follow a different test—that of *Frye v. United States*.[203] Under the *Frye* test, "the thing

---

200. *See* Howard L. Nations & Cindy L. Nations, *The Rules of Digital Evidence, in* ANNUAL ADVANCED ALI-ABA COURSE OF STUDY CIVIL PRACTICE AND LITIGATION TECHNIQUES IN FEDERAL AND STATE COURTS 501, 506 (2007).

201. *See* 1 DAVID FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE § 1:4-5, at 6-10 (student ed. 2008).

202. Even though the federal rules are thought to be coterminous with the *Daubert* standard, there are some states (notably Michigan, Arizona, and Missouri) that use language similar to FRE 702 in their evidence codes without subscribing to the specific scientific validity framework espoused by *Daubert*. *See* Alice B. Lustre, *Post-*Daubert *Standards for Admissibility of Scientific and Other Expert Evidence in State Courts***,** 90 A.L.R.5th 453 (2001).

203. In 1923, the Circuit Court of Appeals for the District of Columbia influenced the way other state courts thought about the admissibility of expert scientific testimony for seventy years. In a case dealing with the precursor to the polygraph, the court held that the "test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made." Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). Under the *Frye* test, expert scientific testimony is admissible when the method underlying the data has "*gained general acceptance*" by the relevant community. *Id.* at 1014 (emphasis added). Jurisdictions that have adopted *Daubert* principles include Alaska, Arkansas, Connecticut, Delaware, District of Columbia, Georgia, Kentucky, Louisiana, Massachusetts, Michigan, Mississippi, Montana, Nebraska, New Hampshire, New Mexico, Ohio, Oklahoma, Oregon, Rhode Island, South Dakota, Texas, Vermont, West Virginia, and Wyoming. Victor E. Schwartz & Cary Silverman, *The Draining of* Daubert *and the Recidivism of Junk Science in Federal and State Courts*, 35 HOFSTRA L. REV. 217, 267 n.300 (2006). Conversely, Arizona, California, Florida, Illinois, Kansas, Maryland, Minnesota, New York, North Dakota, Pennsylvania, and Washington

from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."[204] Fourteen states still employ *Frye*, and others employ a similar standard.

The Supreme Court decided *Daubert v. Dow Chemical* partially in response to the perceived weaknesses of *Frye*.[205] In *Daubert*, the Supreme Court held that under the FRE, expert scientific testimony must be scientifically valid in order to be admissible.[206] Specifically, the Court provided a nonexhaustive list of factors that trial courts could use to decide whether evidence is reliable enough to be admitted.[207] This list included asking whether the evidence was falsifiable, refutable, and testable; whether it was subjected to peer review and publication; whether there were known or potential error rates; whether there were standards concerning its operation; and lastly, whether the theory and technique were generally accepted by the relevant scientific community. In deciding *Daubert*, the Court thought it was interpreting the existing FRE, but after *Daubert* was decided, the drafters of the FRE revised the rules to make the endorsement of the *Daubert* standard explicit.[208] About half of the states followed suit and revised their evidentiary rules to follow the new federal standard.

Daubert was a dramatic change from *Frye*. The *Frye* standard did not require judges to know much about a novel scientific method, as they could just rely on its acceptance by an external, relevant community. Supporters of *Frye* have argued that the requirement of general acceptance by the relevant community "assures that those most qualified to assess the general validity of a scientific method will have the determinative voice."[209] Under *Daubert*, however, judges are now required to act as amateur scientists to determine whether the methods are scientifically valid.[210] Depending on its intended use,

---

have rejected *Daubert* and instead continue to follow the *Frye* test of general acceptance. *Id.* at 267, n.301.

204. *Frye,* 293 F. at 1014.

205. The *Frye* test was criticized as both too liberal and too conservative, depending on the type of evidence that the proponent sought to introduce. It seems to privilege well-organized trades or professions and disproportionately excludes reliable evidence that is experimental and not accepted by any particular community. *See* Daubert v. Dow Chem., 509 U.S. 579, 588-89 (1993).

206. 1 FAIGMAN, *supra* note 201, § 1:6-7, at 11-17.

207. *See Daubert*, 509 U.S. at 593-94.

208. "[T]he Rules of Evidence—especially Rule 702—do assign to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand." *Daubert*, 509 U.S. at 597; *see also* FED. R. EVID. 702 advisory committee's note ("Rule 702 has been amended in response to *Daubert v. Merrell Dow Pharmaceuticals, Inc.* . . . In *Daubert* the Court charged trial judges with the responsibility of acting as gatekeepers to exclude unreliable expert testimony . . . .").

209. People v. Kelly, 17 549 P.2d 1240, 1244 (Cal. 1976) (quoting United States v. Addison, 498 F.2d 741, 743-44 (D.C. Cir. 1974)) (emphasis omitted).

210. While Chief Justice Rehnquist derided the new burden the *Daubert* majority imposes upon judges, legal scholars have noted that judges must now be amateur scientists.

expert testimony based on functional neuroimaging may be admissible under *Frye* or *Daubert*. Because fMRI is a valid research tool that has been generally accepted by cognitive scientists and is subject to peer review, and the precise research methods can be replicated and rely on theories that can be tested for error rates, it seems to pass both the relatively permissive *Frye* test and the stricter *Daubert* test when introduced to prove the content of the research itself.[211] The key issue, however, under either a *Frye* or *Daubert* analysis, is the particular purpose for which the brain image is introduced.[212] An expert's testimony is deemed relevant if it fits the facts of a case. Post-*Daubert*, another Supreme Court decision made it clear that the data on which the expert relies must also fit the facts.[213] There must be an appropriate logical nexus between the data and the ultimate conclusions, and the scientific methods and resulting data must be valid *for a particular purpose*.[214]

Proponents of the use of novel scientific evidence in courts often gloss over this significant "fit" requirement. Rather than engage the critical question of whether or not the methodology is generally accepted or scientifically valid for this purpose, they cite the thousands of peer-reviewed articles that rely on fMRI but fail to fully address the purpose for which the findings are relevant.[215] However, the sheer presence of fMRI in peer-reviewed journals is not sufficient to demonstrate its acceptance for a particular purpose and thus permit its admission. The Supreme Court has clarified that the gatekeeping function of

---

See Faigman, *supra* note 128, at 1209 ("In the twenty-first century–and the sooner the better–judges have no choice but to become amateur scientists. The job requires it. This is true well beyond the narrow region of admissibility rules for expert evidence and includes all contexts in which empirical research is relevant to legal decision making.").

211. Under the federal rules, a cognitive neuroscientist could testify as to either memory or deception . . . [W]ith regard to memory, the result of this step of the inquiry would clearly pass the test of admissibility, at least so long as the test was conducted with due regard to the individual variation in brain structure activation. This would require matching neuroimages with stimuli known to be familiar and unfamiliar to that person, before assessing the level of familiarity the subject has with the item whose familiarity level is unknown and relevant. With regard to lie detection . . . the evidence would fail [to be admissible], because there are not currently any reliable "deception signatures" against which the individual's neuroimaging result could be matched . . . .

Charles N.W. Keckler, *Cross-Examining the Brain: A Legal Analysis of Neural Imaging for Credibility Impeachment*, 57 HASTINGS L.J. 509, 541-42 (2006) (footnote and emphasis omitted).

212. Under either test for scientific evidence admissibility, the evidence must still be relevant for something at issue in the case; that is, it must be used to make an argument that has a nexus to the case at hand. "In the area of scientific evidence, the *Daubert* Court explained, relevance foremost is a question of fit. Specifically, whatever the validity of the science, it must pertain to some disputed issue in the case." *See* FAIGMAN, *supra* note 201, § 1:10, at 21.

213. *See* General Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997).

214. *Id.*

215. Cephos Corp., Lie Detection, http://www.cephoscorp.com/lie-detection/index.php#admissibility (last visited Nov. 16, 2009) ("There are over 15,000 fMRI publications so the [Cephos] technique is undoubtedly generally accepted.").

judges extends to conclusions, explaining that "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert."[216] Even the *Frye* test required that technology be generally accepted *for a particular purpose*.[217] Thus, determinations of admissibility of scientific evidence cannot be made in a vacuum and with passing reference to a universe of scholarly publications.

One of the few areas in which fMRI has demonstrated clinical application is in pre-surgical planning. Neurosurgeons may rely on fMRI before surgery to map an individual's critical language areas to assist in decision-making about how the surgery will proceed, so as to minimize functional disruption in a particular patient.[218] fMRI may also be scientifically valid and generally accepted by neuroscientists when introduced to demonstrate brain damage[219] or stroke.[220] When used to determine retrospective and complex mental states in an individual subject, however, fMRI is not sufficiently studied. Nor are there any respected researchers who presently support the use of functional brain images to assess past mental states in a forensic setting.[221] fMRI is

---

216. *Joiner*, 522 U.S. at 146.

217. People v. Ford, No. B171801, 2005 WL 236487, at *7 (Cal. Ct. App. Jan. 25, 2005) ("The trial court excluded any reference to the SPECT scan results, finding that the technology was not generally accepted in the medical community for diagnostic purposes in a forensic setting and therefore did not meet the *Kelly* standard for admissibility." (citing People v. Kelly, 17 Cal.3d 24, 28-30 (1976))).

218. As language centers in the brain are more focused and specific than those for "forming intent," pre-surgical surveying of linguistic function can assist the surgeons in ensuring that they do not obliterate linguistic connections during neurosurgery. *See* John E. Desmond & S.H. Annabel Chen, *Ethical Issues in the Clinical Application of fMRI: Factors Affecting the Validity and Interpretation of Activations*, 50 BRAIN & COGNITION 482, 485-86 (2002); Paul M. Matthews, Garry D. Honey & Edward T. Bullmore, *Applications of fMRI in Translational Medicine and Clinical Practice,* 7 NATURE REVS. NEUROSCIENCE 732, 733-34 (2006).

219. *See* Giovanni B. Frisoni et al., *Mapping Local Hippocampal Changes in Alzheimer's Disease and Normal Ageing with MRI at 3 Tesla*, 131 BRAIN 3266 (2008); C.L. MacDonald et al., *Verbal Memory Deficit Following Traumatic Brain Injury: Assessment Using Advanced MRI Methods,* 71 NEUROLOGY 1199, 1201 (2008).

220. *See* Steven C. Cramer et al., *Use of Functional MRI to Guide Decisions in a Clinical Stroke Trial*, 36 STROKE e50, e51-52 (2005).

221. A research team in Germany has demonstrated some success in predicting the mental state of intention in tightly controlled settings. The research team asked subjects to covertly form an intention to either subtract or add two numbers. After a variable delay the subjects performed their chosen activities, and then indicated which tasks they had prepared. The study demonstrated that during the delay, it was possible to decode from activity in medial and lateral regions of the prefrontal cortex which of two tasks the subjects had covertly intended to perform. It is hard to imagine near-term applications of this finding in a legal setting, and it remains to be replicated in other labs. For further details of the fascinating study, see John-Dylan Haynes et al., *Reading Hidden Intentions in the Human Brain*, 17 CURRENT BIOLOGY 323 (2007). For a discussion of the complexity of disentangling the neural basis of phenomenal consciousness from the neural models, see Ned Block, *Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience,*

therefore neither generally accepted nor scientifically valid for such a purpose as a mental state diagnosis. When properly applied, FRE 702 and the state-specific rules governing expert scientific evidence should lead to the exclusion of functional brain images addressed to past mental states.

But what if the images themselves are not introduced, and the expert merely relies upon them outside of court to form her opinion? Under the Federal Rules of Evidence[222] and thirty state counterparts,[223] the images need not be admissible in order for the expert testimony to be admitted, so long as the images are "of a type reasonably relied upon by experts in the particular field." However, this returns us to something resembling a *Frye* analysis for the factual basis of expert inference, and we have demonstrated that fMRI is not the type of evidence that is relied upon by psychiatrists or neuroscientists to make inferences about past mental states. Moreover, the images should not be disclosed to the jury unless the court determines that their probative value in assisting the jury to evaluate the expert's opinion substantially outweighs their prejudicial effect.[224]

F. *Probative Value*

The limited probative value of fMRI addressed to mens rea forms the crux of our argument. Therefore, much of what we discuss here will apply the findings of Part II to an analysis of probative value and the potential for unfair prejudice. Probative value, in this case, is intimately related to the scientific and methodological limitations of fMRI.

1. *fMRI has limited probative value unless the question of proper base rates is resolved*

As noted above in Part II.C, many imaging methods, and particularly those used in a forensic context, compare an individual subject's brain against the average of a group of controls. The control group is meant to represent the normal population. But "normal" is a statistical creation, and the scientific community does not yet have anything resembling population data on fMRI-assessed brain activity during even the most basic of tasks. There are two

---

30 BEHAV. & BRAIN SCI. 481 (2007). For a discussion of how we might begin to isolate specific neural activity in a sea of constant and complex stimuli, see Hugo J. Spiers & Eleanor A. Maguire, *Decoding Human Brain Activity During Real-World Experiences*, 11 TRENDS COGNITIVE SCI. 356 (2007).

222. FED. R. EVID. 703.

223. Several state courts have held that, under state counterparts to FRE 703, an expert may be allowed to disclose to the trier of fact the facts on which the expert's opinion are based. *See* 89 A.L.R.4th 456, §§ 2, 3 (West 2009).

224. This is the opposite of FRE 403, which will be discussed later, as the presumption under FRE 703 is exclusion, rather than admissibility.

problems with our lack of knowledge about what is normal. First, we do not know the underlying prevalence of the specific functional deficit (i.e., the population base rate), and second, we do not know whom to include in the control group (i.e., the reference class).

Without knowing the prevalence of any functional brain abnormality in the population, we can say very little about the positive predictive value[225] of an fMRI that seeks to establish this abnormality. If the prevalence of an abnormality is low, there will be a great risk of false positives. This problem can be illustrated by considering a population of 10,000 that has a 1% base rate of a particular functional brain abnormality that predisposes one to impulsively commit murder. That means that for every 100 people in the population who have the functional brain abnormality, 9900 do not have it. If an fMRI test for this functional abnormality is 95% specific, meaning that the test will pick up true negatives 95% of the time, then it will falsely tag 5% of the 9900 individuals (or 495) as "positive" even though they lack that abnormality. Even if we assume the fMRI test is 100% sensitive, meaning that it will detect true positives 100% of the time, a positive test result still has the positive predictive value of only 100/595, which is below 17%. This hypothetical example renders 495 false positives, a number that appears too high in the face of potential criminal conviction. Perhaps this number of false positives might not be problematic for civil uses that require proof only by a preponderance of the evidence; this is a normative question, however, that is better explored in other literature.[226]

The lack of empirical data on prior probabilities, or base rates, is troubling for courtroom use of many types of evidence that rely on statistics.[227] This includes the interpretation of survey data, DNA analysis, and individual functional brain images. To help cure the problem, experts relying on brain images and other statistical evidence should speak plainly about the limitations of the process. A judge should question why the specific base rates were used, and perhaps require that experts proffer various statistical probabilities based on unproven but reasonable assumptions.

---

225. Positive predictive value is a statistical term of art. It refers to a test's ability to properly classify people as "positive" who will then develop the disease or dysfunction.

226. We have dealt elsewhere with normative considerations, which in turn can impact upon the predictive value itself:

> This low predictive value of a single positive test for a functional abnormality can be improved by additional tests, but only if the defense allows (or the court requires) those additional tests to be conducted. If the defense gets a positive result that they think supports their claim that the defendant is not fully responsible, then they might be unwilling to subject the defendant to further testing. That will make it difficult in practice to improve upon the low predictive value.

Walter Sinnott-Armstrong et al., *Brain Images as Legal Evidence*, 5 EPISTEME 359, 364 (2008).

227. *See* FAIGMAN, *supra* note 201, § 5:23, at 197; *see also* People v. Collins, 438 P.2d 33, 38 (Cal. 1968).

The next problem is determining the appropriate reference class. Imaging limitations notwithstanding, the selection of the appropriate reference class may be a legal decision, depending on whether the law decides to subjectivize the standard to which a particular defendant is held.[228] Should we measure the defendant's brain activity and compare it against other adults in the community? Other males of the same age and IQ? Other males who take similar drugs?[229] Asking whether we should use a subjective standard for defending criminal conduct based on neural, structural, functional, or behavioral deficits highlights our lack of knowledge regarding the convergent relationship between these features. For any given brain state, there is no predetermined reference class for making references to normal brain functioning.[230] As a result, the probabilistic power of the inference will fluctuate greatly depending on the group of normals to which we compare the subject.[231] Without understanding the significance of an appropriate reference class, juries may give too much weight to the fMRI data and related expert testimony.

As neuroscientists are still identifying which factors influence the BOLD response, group comparisons are nearly worthless for the purpose of making diagnostic predictions at an individual level.[232] While there may be probative

---

228. *See* David M. Paciocco, *Applying the Law of Self-Defence*, 12 CAN. CRIM. L. REV. 25, 44 (2007) ("The practice illustrated by [the] decisions of "subjectivizing" the reasonable person to account for the frailties of the accused has an attractive human dimension as it is sympathetic to the uneven capabilities that individuals possess through no fault of their own. It is, however, problematic as a matter of principle, pragmatics, and legal consistency to evaluate reasonableness using the kinds of variable standards that this line of authority produces.").

229. This issue was raised by Michael Pardo and Ronald Allen when they questioned the utility of probabilistic evidence that does not have a predetermined reference class. *See* Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 113 (2007) ("Instead of being natural facts consigned to predetermined reference classes with labels attached to designate the proper class, the evidence and the events on which it is based are members of an infinite number of reference classes, which lead to various inferences of various strength depending on how the boundary conditions of those classes are specified." (citation omitted)).

230. *See* Edward Cheng, *A Practical Solution to the Reference Class Problem*, 109 COLUM. L. REV. 2081, 2082-83 (2009); Kevin H. Smith, *External Validity: Representativeness and Projectability in the Probative Value of Sample Surveys*, 39 WAYNE L. REV. 1433, 1461 (1993) ("The proper specification of the universe is arguably the most critical step in a survey's design. If the researcher fails to accurately specify the legally relevant universe, the best that will be achieved is a well-designed and well-executed survey that yields good data about the wrong group of people." (citations omitted)).

231. *See* EDWARD J. IMWINKELRIED, THE METHODS OF ATTACKING SCIENTIFIC EVIDENCE § 13-10, at 467-69 (1982).

232. Scientists are still collecting data on what types of environmental and dietary factors can affect glucose metabolism in the blood and therefore the BOLD response. Some data suggest that women who are menstruating have different BOLD responses than those who are not. *See* Jean Claude Dreher et al., *Menstrual Cycle Phase Modulates Reward-Related Neural Function in Women*, 104 PNAS 2465, 2466 (2007). Without a better

value in the weakest of associations, the inference about dysfunction will be meaningless unless more is known about the neurophysiology and variance in activation patterns of normal people who are presented with opportunities to commit crimes. If no rational sampling frame can be composed, the results will be plagued by sampling error, meaning that the "normal" sample will not represent the right population.[233] Because of this deficiency, one researcher stated bluntly that "[a]t the moment, fMRI would be among the most useless things to do" when attempting to draw inferences about an individual's capacity or mental state from group neuroimaging data.[234]

2. *fMRI has limited probative value as it relies on averaged group data and ignores individual differences*

As discussed *supra* in Subpart II.C, the current fMRI methodologies rely heavily on averaged group data to make inferences about whether an individual, by comparison, is normal. These averages often eliminate important cognitive, emotional, and perceptual differences between control subjects, as the data focus on overlapping, or shared, activation.

As noted above, a growing list of external variables can affect the BOLD response. Because of the present inability to account for irrelevant individual differences in the control data, if the defendant is compared to the group map, he may be incorrectly identified as "abnormal" if any shared activation data is subsumed by averaging with a larger set. The opposite is also true, as two individuals may perform very similarly on a behavioral task, and yet have very different brain activity.[235]

Until we can better understand the variables that can lead to individual differences in activation, comparisons between individuals in arbitrarily selected groups will not yield useful legal results. The criminal law cares about individual differences more than most scientific research does. Questions about mens rea do not ask what most *reasonable* people would think at the time of a crime, but rather what the defendant *in fact* thought or was capable of thinking at the time.[236] Therefore, the reliance on averaged group data of what is

---

understanding of external variables that can affect this measure, any group-to-individual inference will have very limited value. *See* Miller, *supra* note 121, at 1201.

233. For a discussion of the problems with surveys that do not capture a random sample of the universe under investigation, see *Amstar Corp. v. Domino's Pizza, Inc.*, 615 F.2d 252, 264 (5th Cir. 1980); *United States ex rel. Free v. Peters*, 806 F. Supp. 705, 715-16 (N.D. Ill. 1992); *see also* FAIGMAN, *supra* note 201, § 4:17, at 148.

234. Richard Robinson, *fMRI Beyond the Clinic: Will It Ever Be Ready for Prime Time?*, 2 PLOS BIOLOGY 0715, 0716 (2004) (quoting John Gabrieli, Associate Professor of Psychology at Stanford University).

235. *See* Bell & Racine, *supra* note 42, at 24.

236. Any reliance on what has been dubbed "social frameworks" theory would be inappropriate here, if only for the narrow reason that a well-researched comparison group of which the defendant is clearly a member does not exist. "Social frameworks" is a term

"normal" is even less relevant for determining a defendant's guilt in a criminal trial.

Inferences are drawn from the control group to the specific individual to help explain a variety of phenomena, such as the reasonableness of a battered woman's self-defense or the reasonableness of a child's delayed response to abuse. As explained above, however, determinations of an accused criminal's mental state cannot be reached based entirely upon what is reasonable; rather, these determinations are factual elements to be proved. For example, murder "may be punished less severely when the killer is merely aware of but ignores a risk of death—'recklessness' —than when the killer desires the victim's death—'purposefulness.' The defendant's *unique thoughts* thus govern the degree of culpability."[237] Averaged data that compares the individual to a group therefore has virtually no probative value for assessing mens rea.

3. *fMRI will have limited probative value for determining mens rea until we know more about the BOLD response*

The limitations on the BOLD response outlined in Subpart II.A underscore the problem of prematurely employing research data for individual diagnostic or courtroom purposes. The BOLD response may provide a helpful, indirect proxy for neural activity in the lab when precision at the individual level is unnecessary. But when the decision is whether to convict or acquit an individual of a serious crime, the lack of information on what the BOLD response actually captures is exceedingly problematic. Does the BOLD response vary based on diet? Does it vary based on hormone levels? Does it capture inhibitory firings?[238] What relationship does it have to underlying synaptic activity? One lab hypothesized that variations in blood circulation may lead to frequent "underestimat[ion]" of activation in fMRI experiments.[239] Much evidence introduced at trial is merely correlative. But given how little is known about the strength of the correlation between the BOLD measure and

---

coined by Laurens Walker and John Monahan to refer to the use of off-the-shelf research results to construct a control group for deciding factual issues crucial to the resolution of a particular case. *See* Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559, 559-60, 570 (1987).

237. Andrew E. Taslitz, *Myself Alone: Individualizing Justice Through Psychological Character Evidence*, 52 MD. L. REV. 1, 3 (1993) (emphasis added) (footnote omitted).

238. *See, e.g.*, Logothetis, *supra* note 26, at 873 ("A frequent explanation of the fMRI data simply assumes an increase in the spiking of many task- or stimulus-specific neurons. This might be correct in some cases, but increases of the BOLD signal may also occur as a result of balanced proportional increases in the excitatory and inhibitory conductances . . . .").

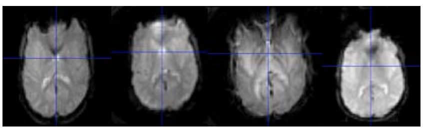
239. Logothetis and his co-workers have found that the BOLD response may reflect the neural activity related to the input and local processing in a given area rather than the output (spiking activity). Nikos K. Logothetis et al., *Neurophysiological Investigation of the Basis of the fMRI Signal*, 412 NATURE 150, 154 (2001).

the brain's physiology, the probative value of fMRI images based on the BOLD response is severely limited.

4. *fMRI will have limited probative value for the purpose of determining mens rea until we have standardized methods for processing the data and creating the activation map*

As we discussed above in Subpart II.C, the techniques of processing data to generate fMRI images, the statistical threshold, and the software package employed can all have a dramatic impact on the resulting activation map.[240] This problem may be mitigated *in part* by vigorous cross-examination that may expose variations in the methods and reduce the jury's confidence in the images. But cross-examination will not do enough, as the problem remains that there are no gold standards for rigorous fMRI analysis in the criminal justice context. Indeed, though it is not our focus here, a *Daubert* challenge might conceivably be raised with respect to the lack of adequate standards for gathering and analyzing the data. In this sense there is currently no objective, measurable truth that each side could work toward in a criminal trial; at present, cross-examination could create a competition over arbitrarily selected software packages. Even if there is no truth at the core of the methodology, the adversarial process of fact-finding would improve if forensic labs were required to standardize their methods. The Functional Biomedical Informatics Research Network ("FBIRN") has developed a suite of tools and methods to calibrate and collect data across diverse environments. FBIRN has also suggested universal methods for storing and managing the data.[241] Without standardized methodology like that proposed by the FBIRN project, there is too great a risk of distortion or manipulation. It remains difficult to know what the fMRI image means and why the particular analysis was employed.[242]

240. The Biomedical Informatics Research Network (BIRN) group demonstrated this by looking at data from the same person, analyzed at four different laboratories with different fMRI machines and use protocols. The image below, available at Function BIRN, http://www.birncommunity.org/current-users/function-birn (last visited Nov. 16, 2009), represents the results.



As Daniel Langleben, Assistant Professor of Psychiatry at the University of Pennsylvania School of Medicine, explained: "[I]f you want to abuse this technique and claim that it works, you can create tests that will produce results—I can see how it could be done. We know enough to rig it." Robinson, *supra* note 234, at 0717.

241. Specifically, FBIRN suggests standardized imaging parameters, quality control measures, and standardized software packages for multi-cite imaging comparisons. All of these tools are available for download from BIRN Home Page, http://www.birncommunity.org.

242. However, even with greater standards for developing fMRI, the BIRN project

5. *fMRI will have limited probative value for the purpose of determining mens rea so long as institutional review boards exist and research ethics are followed*

fMRI is highly constrained by the variety and validity of tasks that can be performed in the scanner.[243] At present, there are no psychological tasks that map directly onto a complex, legally-defined mental state such as intent to harm or kill, or even the capacity to form such intent. This deficiency in the state of knowledge assaults the ecological validity of the fMRI methodology.[244]

As previously mentioned, labs carefully create behavioral tasks to measure specific cognitive phenomena such as executive control or impulsivity. Many of these tasks are quite good at assessing cognitive flexibility,[245] learned and unlearned conditioned fear responses,[246] inhibitory control,[247] and risk

---

cannot get rid of fundamental problems with interpretation at the individual level. *See* Laurence R. Tancredi & Jonathan D. Brodie, *The Brain and Behavior: Limitations in the Legal Use of Functional Magnetic Resonance Imaging*, 33 AM. J.L. & MED. 271, 293-94 (2007) ("Even assuming that the BIRN project succeeds . . . , the difficulties of using fMRI images to draw legal conclusions will continue to remain unresolved. . . . The image doesn't explain by itself psychiatric syndromes, complex behavior, nor even something as discrete as an impulsive reaction. Furthermore, a 99% correlation from a group study between an fMRI image and violent behavior does not tell us why person 'X' committed murder.").

243. *See* Bell & Racine, *supra* note 42, at 23-24.

244. By weak "ecological validity," we mean that the cognitive or behavioral task that the subject is asked to perform will almost never mirror the decision-making process of a real-life criminal setting.

245. The Probabilistic Reversal Learning Task is one measure of cognitive flexibility. On each trial, subjects are presented with two abstract visual patterns. Using trial-and-error feedback, subjects must discover which of the two patterns is correct. Feedback (a green smiley face or red sad face) is presented as soon as the subject has chosen one of the patterns by a left or right button press. The ventrolateral prefrontal cortex is the area of the brain that is often implicated in this task. *See* Roshan Cools et al., *Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging*, 22 J. NEUROSCIENCE 4563, 4563-64 (2002); Chris M. Dodds et al., *Methylphenidate Has Differential Effects on Blood Oxygenation Level-Dependent Signal Related to Cognitive Subprocesses of Reversal Learning*, 28 J. NEUROSCIENCE 5976, 5976-77 (2008).

246. The Pavlovian Fear Extinction task measures the ability to "unlearn" conditioned fear stimuli such as responses to snakes or unfamiliar faces. In this task, a simple discrimination, partial-reinforcement paradigm is used: the conditioned stimuli are colored squares and the unconditioned stimulus is a mild shock to the wrist. *See* Elizabeth A. Phelps et al., *Extinction Learning in Humans: Role of the Amygdala and vmPFC*, 43 NEURON 897, 898, 904 (2004). The ventromedial prefrontal (or medial orbitofrontal) cortex is an area of the brain that is often implicated in this task. *See* Mohammed R. Milad et al., *Thickness of Ventromedial Prefrontal Cortex in Humans Is Correlated with Extinction Memory*, 102 PNAS 10706, 10709-10 (2005); Phelps, *supra*, at 897.

247. The Stop-Signal Task gives a sensitive estimate of inhibitory control—the stop signal reaction time (SSRT)—reflecting the time it takes to internally suppress a response. In each trial, a left- or right-pointing arrow stimulus is displayed on a computer screen. The subject responds by pressing the left or right key as quickly as possible ("go" task) unless

evaluation in decision-making.[248] The important point here is that, even if the behavioral task is reliable for a particular cognitive measure in the lab, it may not be well-suited for diagnostic purposes in the courtroom. Specific intent to kill is not an operationally-defined cognitive state, and in fact it may actually represent multiple cognitive states. While many experimental tasks may be relevant to the ability to form specific intent to kill, the tasks often require extrapolation to a degree that severely limits their relevance in the courtroom. There are several reasons for this conclusion.

First, the emotional stakes in an experiment are comparatively low. No one is actually going to be robbed, stabbed, shot, arrested, or sentenced. The response in a lab will therefore never be as real as the response on the street, and may reflect a different decision-making process entirely (as the fear of sentencing may replace the fear of being caught). Second, in the case of using fMRI to buttress the insanity defense, presenting the defendant with a question about whether or not he could appreciate the wrongfulness of his actions at the time of the crime will be skewed by intervening events (e.g., by his current understanding of the criminal charge, the passage of time, or incentives to malinger). The same incentive to malinger could exist even in a test of executive control that assesses the ability to form intent. Third, and still quite important, is the problem that in some cases, "the very same behavior that might not be deemed criminal in one social context (say, shooting a gun at a target at a shooting range) may be deemed criminal in another (such as shooting a gun in the direction of a crowd of people)."[249] Context obviously matters. The complex situation that gives rise to the criminal behavior will be impossible to replicate, or even approximate, in a lab.

Perfect unison between the experimental task and the criminal setting is impossible to reach. Institutional review boards ("IRBs") oversee research on human subjects to ensure that the experiments are not overly risky to the

---

she hears a beep, in which case she is instructed to withhold a response ("stop" task). The inferior frontal gyrus is an area of the brain that is often implicated in this task. *See* Adam R. Aron et al., *Stop-Signal Inhibition Disrupted by Damage to Right Inferior Frontal Gyrus in Humans*, 6 NATURE NEUROSCIENCE 115, 115-16 (2003).

248. The Cambridge Gamble Task assesses decision-making under risk, with imperfect probability information. The subject is presented with an array of ten red and blue boxes and is instructed that the computer has hidden a token under one of the boxes. After guessing whether the token is hidden in a red or blue box, the subject is asked to wager a proportion of her points on that decision. Potential wagers are offered in an ascending or descending sequence that enables the separation of impulsive response tendencies from genuine risk preference (risk-preferring subjects must wait to place high wagers in the ascend condition). Outcome probabilities of winning and losing are provided explicitly by the ratio of red to blue boxes, and thus the task assesses decision-making under risk rather than ambiguity. The ventromedial prefrontal cortex is an area of the brain that is often implicated in this task. *See* Luke Clark et al., *Differential Effects of Insular and Ventromedial Prefrontal Cortex Lesions on Risky Decision-Making*, 131 BRAIN 1311, 1312-13 (2008).

249. Dean Mobbs et al., *Law, Responsibility, and the Brain*, 5 PLOS BIOLOGY 0693, 0695 (2007).

individual or to society.[250] Because of the regulatory and ethical requirements that the research minimize the risk of harm to the subject, IRBs are incredibly unlikely to approve a set of behavioral tasks that mirror true criminal choices and settings. For this reason, the judge should remind the jury that the behavioral task is at best only loosely related to the mental state at the time of the crime. Again, this does not mean that the data are useless, but rather that their probative value to the legal question of mens rea is severely diminished based on the low ecological validity of the behavioral task.

6. *fMRI has limited probative value for evaluating past mental states as it measures present reactions to present stimuli*

By now, it is clear that using measurements of a behavioral task to attempt to assess a past mental state using fMRI requires significant inferential leaps. The most remarkable leap virtually requires time travel when jurors are asked to view an image from the present to infer capacity or lack of mens rea in the past. Even if the most ecologically valid behavioral task were created, the "perfect" group data were collected, the contrast between the defendant's activation and the group's activation were statistically significant, and regions where circuits involved in intent formation had been localized (these are all huge "ifs"), we still know nothing directly about how the brain was functioning at the time the crime was committed. Each inferential leap raises considerable issues of remoteness.[251] Concededly, many types of evidence require inferences about the present based on the past. But with brain imaging, the time travel aspect deserves special mention because it is not transparent, and the brain states being "captured" are transient rather than fixed.

---

250. In addition to international and federal ethical regulations that prohibit research resulting in significant risk of harm to the human subjects, there may be civil or criminal charges brought for any research that involves the simulation of criminal acts. For a summary of the IRB regulations, see 45 C.F.R. § 46.103(f) (2005). For information on human subject research, see THE NAT'L COMM'N FOR THE PROT. OF HUMAN SUBJECTS OF BIOMEDICAL AND BEHAVIORAL RESEARCH, NAT'L INST. OF HEALTH, THE BELMONT REPORT (1979), *available at* http://ohsr.od.nih.gov/guidelines/belmont.html#goa.

251. Remoteness usually affects the weight of the evidence, rather than its admissibility. *See* 22A C.J.S. *Criminal Law* § 1021 (2009) ("It is held that the probative remoteness of evidence is never alone a reason, or an absolute reason, for its exclusion, but should always be accompanied by some other objection if it is to prevail, such as the confusion which may result or the emotions which the evidence may arouse to disturb impartial decision." (footnotes omitted)); *see also* Matthews v. James, 362 S.E.2d 594, 599-600 (N.C. Ct. App. 1987) ("Although the mental capacity of the decedent to change the beneficiaries of his Blue Bell plans must be determined as of the date of the execution of the forms effecting such change, evidence of the decedent's mental capacity at other times is admissible if it bears on the issue of the decedent's mental capacity at the time he executed the changes."); *In re* Daniels, 313 S.E.2d 269, 271 (N.C. Ct. App. 1984) ("Evidence of mental condition before and after the critical time is admissible, provided it is not too remote to justify an inference that the same condition existed at the latter time." (citing 1 BRANDIS, N.C. EVIDENCE § 127 (1982))).

In the case of an insanity defense, we may need to know whether the defendant was capable of knowing whether his actions were wrong. This still requires time travel, as a defendant may be legally sane at the time of the crime and then rapidly decompensate in prison. Alternatively, a defendant may have been legally insane at the time of the crime (perhaps due to a combination of not taking psychiatric medicine and some triggering event) and legally sane at the time of trial due to adherence to the use of prescribed psychiatric medications and abatement of stress.

7. *Multiple steps in the chain of inference severely limit the probative value of fMRI*

Of course, a lengthy chain of inferences is not itself fatal to relevance.[252] The significant inferential leaps that were described above indicate weaker probative value, but they do not eliminate probative value altogether. Evidence can have enough probative value to be admissible even if the proposition for which it was offered still seems quite implausible once introduced.[253] This is because evidence need only get you to first base, as it were, in order to be relevant; it need not be a home run. The party offering the brain image as evidence merely has to demonstrate that it *tends to increase the likelihood* that the defendant lacked the requisite mens rea, not that it proves the defendant *more likely than not* lacked such mens rea.

For all of the reasons articulated above, we argue that, at present, functional neuroimaging has almost no probative value when introduced to prove an individual's past mental states. But because a brain image may be correlated with a behavioral task that is itself correlated with some mental state, it is at least marginally possible that the image brings the defendant one infinitesimal step closer to making out a mens rea defense based on the lack of intent or insanity. Because the bar is so low, evidence is almost never excluded because it has no probative value.[254] We do not suggest here that fMRI images should be excluded based on their lack of probative value alone, although an argument along these lines could be made. Because the probative value is paltry, FRE 403 and its state counterparts present a stronger threat to admissibility.

G. *Unfairly Prejudicial Effect and the Role of FRE 403*

FRE 403 states that relevant evidence can be excluded if its probative value

---

252. 1 MCCORMICK ON EVIDENCE § 185, at 733 (6th ed. 2006) ("[T]he common objection that the inference for which the fact is offered 'does not necessarily follow' is untenable. It poses a standard of conclusiveness that very few single items of circumstantial evidence ever could meet. A brick is not a wall." (footnote omitted)).

253. *See id.*

254. *See id.*

is substantially outweighed by its unfairly prejudicial effect.[255] Some states do not use the term "substantially outweighed" and require that the evidence be excluded if the potential for unfair prejudice merely outweighs the probative value.[256] Other states require the potential for unfair prejudice to substantially outweigh the probative value, but will only exclude relevant evidence if there is a finding by the judge that it inflames the jurors' emotions or is simply too unfair.[257] We will focus our analysis on FRE 403, as many states have modeled their balancing tests on its language. Also, if evidence is excluded under FRE 403, it will most certainly be excluded under the lower preponderance standard used by some states.

Though all relevant evidence prejudices one party over the other, Rule 403 only challenges evidence that is potentially *unfairly* prejudicial by thwarting the jurors' fact finding or influencing the outcome of the trial by improper means.[258] Scientific evidence's potential for unfair prejudice stems from its objective appearance. The admissibility of many novel scientific tools has been challenged on the grounds that the jury will place too much value in them. Such was the case with the polygraph[259] and mitochondrial DNA.[260]

Unfortunately, neither of these examples is an appropriate comparison for fMRI. fMRI is potentially *more* unfairly prejudicial than the polygraph, as it appears less vulnerable to subjective interpretation.[261] Unlike DNA evidence,

---

255. *See* Maxwell v. Aetna Life Ins. Co., 693 P.2d 348, 356 (Ariz. Ct. App. 1984) ("Only where the prejudice substantially outweighs the probative value will relevant evidence be excluded."); *see also* Farr v. Wright, 833 S.W.2d 597, 602 (Tex. Ct. App. 1992) ("We note that this rule generally comes into play when the probative value of the evidence is low or the problems with the evidence are serious. Only if the probative value of the evidence is *substantially* outweighed by the prejudicial effect is exclusion proper.").

256. *See, e.g.*, ALASKA R. EVID. 403; CONN. CODE EVID. § 4-3; PA. R. EVID. 403.

257. *See* WYO. R. EVID. 403; *In re* MLM, 682 P.2d 982, 986 (Wyo. 1984) ("The showing of prejudice to exclude relevant evidence requires a demonstration that the 'evidence had little or no probative value and that it was extremely inflammatory or introduced for the purpose of inflaming the jury.' Evidence must be deemed unfair to be prejudicial." (citation omitted)).

258. 31A C.J.S. *Evidence* § 292 (2009).

259. *See, e.g.*, United States v. Scheffer, 523 U.S. 303, 314 (1998) (Thomas, J., concurring) (plurality opinion) ("Such jurisdictions may legitimately determine that the aura of infallibility attending polygraph evidence can lead jurors to abandon their duty to assess credibility and guilt.").

260. *See, e.g.*, State v. Pappas, 776 A.2d 1091, 1113 (Conn. 2001) ("The concern is that jurors will overvalue DNA evidence and ignore other types of evidence."); *cf.* Sarah Brashears-Macatee, Note, *A Test Both Lawyers and Scientists Can Live With: The Rigorous Five-Prong Test for the Admission of DNA Profiling Evidence Adopted in* United States v. Matthew Sylvester Two Bulls, 71 NEB. L. REV. 920, 934 (1992) ("Though the principle and techniques behind DNA profiling can be difficult to grasp, examination of an autoradiograph can be illuminating even to lay persons.").

261. The polygraph, like fMRI, relies on an unproven chain of inference from a measured biological function to a mental state. It also, again like fMRI, depends largely on the skill and discretion of the polygrapher. However, the process for producing the polygraph marks is more transparent and does not approach the level of complexity involved

which has been generally adopted as reliable science for identification purposes, fMRI does not yet have any standardized procedures for analysis or replication.[262] Perhaps the best analogy for our analysis is the use of computer-generated images. Hailed as a "silver bullet" in litigation, they seemingly guarantee success by replacing a juror's point of reference for what actually happened.[263] Any evidence that relies on computers is particularly prone to being overvalued, as jurors may view the evidence as "the ultimate arbiter of truth and afford too much weight to [the computer-generated images] while disregarding other crucial pieces of evidence."[264] Similarly, there is great potential for overestimating the objectivity of brain images.

1. *fMRI images may be overvalued due to their glossy portrayal of "hard science"*

Neuroscientist Dean Mobbs has called the allure of fMRI images the "Christmas tree phenomenon";[265] the concern is that jurors may be so impressed by the colorful depiction of the brain that they neglect the expert's interpretation of the image. Likewise, the complexity of the image's meaning may be lost on jurors who "perceptually convert differences of degree (of blood oxygenation levels) into differences in kind: brain activity versus no brain activity."[266] The use of color also has the potential to reinforce the salience and memory of the image (which, if the image is properly described to the jury, can also be a positive effect).[267]

---

in developing a brain image. *But see* Keckler, *supra* note 211, at 514-15 ("What current research offers is what the polygraph fundamentally did not, a way to go beyond the external correlates of deception and into the specific neural processes that underlie the different types of deceptive behavior.").

262. DNA is a much more reliable science, assuming proper specimen procurement and lab practices are followed. It is still imperfect, however, as there is always a risk of contamination. *See, e.g.*, State v. Lyons, 863 P.2d 1303, 1311 (Or. Ct. App. 1993) ("[W]e find nothing about the [method of producing DNA] or the presentation of the results of the . . . method that would undeniably cause jurors to misuse, misinterpret or overvalue the results. . . . [T]he results of the . . . method are not expressed in terms of statistical probabilities capable of creating the aura of absolute identification."). A better example for comparison comes from the use of behavioral genetics data, not to identify suspects, but as character, aggravating, or mitigating evidence. *See* Farahany & Coleman, Jr., *supra* note 58, at 116.

263. Robert B. Bennett, Jr., Jordan H. Leibman & Richard E. Fetter, *Seeing Is Believing; Or Is It? An Empirical Study of Computer Simulations As Evidence*, 34 WAKE FOREST L. REV. 257, 257-58 (1999).

264. Commonwealth v. Serge, 58 Pa. D. & C.4th 52, 79 (Com. Pl. 2001).

265. Dean Mobbs, Hakwan C. Lau, Owen D. Jones & Christopher D. Frith, *Law, Responsibility, and the Brain*, 5 PLOS BIOLOGY 0693, 0699 (2007).

266. Neal Feigenson & Richard K. Sherwin, *Thinking Beyond the Shown: Implicit Inferences in Evidence and Argument*, 6 LAW, PROBABILITY & RISK 295, 299-300 (2007) (emphasis omitted).

267. *See* Cope C. Thomas, *Computer Generated Animation: Identifying New and Subtle Prejudicial Special Effects*, FLA. B.J., Dec. 2000, at 52, 53 ("Scientific experimentation has shown that the consistent use of color is a factor in recognition of an

Scientists and lawyers alike have warned that, due in part to the allure of the illuminated display of the brain, jurors will likewise succumb to the seduction of fMRI. But not all jurors or judges have been swayed by neuroimages. In fact, in some reported cases where images have been admitted to negate mens rea, the jury has still found the defendant guilty of first-degree murder.[268] Judges have also properly rejected images based on the existing evidentiary rules.[269] This demonstrates that fMRI may be more unfairly prejudicial in some cases than in others, depending on the make-up of the jury, the competence of the attorneys to debunk sham claims, and the judge's application of the rules of evidence. Jurors may dismiss fMRI images when the results appear to fly in the face of their instincts regarding the defendant's innocence or guilt, but if the image comports with their theory, they may be more likely to give it exaggerated weight. In this way, fMRI may provide scientific grounding for confirming deep-seated biases.

2. *fMRI gives the unfairly prejudicial illusion that you are directly observing the brain's activity*

Referring to fMRI as a "picture of someone's brain" is as rampant as it is inaccurate.[270] There is a good deal of well-intentioned but misguided

object. One experiment has shown that when test subjects were first shown items in color, and then shown the same items in the same color, they were more likely to recall these items than if they were in black and white."); *see also* Aura Hanna & Roger Remington, *The Representation of Color and Form in Long-Term Memory*, 24 MEMORY & COGNITION 322 (1996). *But see* State v. Syriani, 428 S.E.2d 118, 136 (N.C. 1993), *cert. denied*, 510 U.S. 948 (1993), *reh'g denied*, 510 U.S. 1066 (1993); Sonnier v. State, 913 S.W.2d 511, 518 (Tex. Crim. App. 1995) ("Several factors may be considered in determining whether the danger of unfair prejudice substantially outweighs the probative value of photographs, including . . . 'the number of exhibits offered, their gruesomeness, their detail, their size, [and] whether they are black and white or color . . . .'" (quoting Emery v. State, 881 S.W.2d 702, 710 (Tex. Crim. App. 1994))).

268. In *People v. Musselwhite*, 954 P.2d 475, 478 (Cal. 1998), the prosecution derogatorily referred to neuroimaging analysis as a "marvel machine," and the jury ultimately found the defendant guilty of first-degree murder. Donald Horton, a juror on *State v. Stanko*, 658 S.E.2d 94 (S.C. 2008), *cert. denied*, 129 S. Ct. 182 (2008), stated, "Well, I'll be honest with you when we went in deliberation with that PET scan and all that computerized stuff they did, I said I felt like I'd been dazzled by brilliance and baffled with b.s. That's how I felt. I found the state's witnesses much more credible than the defense." Daniel Schorn, *Murder on His Mind: Can Scientific Images Show What's in the Mind of a Murderer?*, 48 HOURS MYSTERY, Jan. 13, 2007, http://www.cbsnews.com/stories/2007/01/11/48hours/main2352668_page6.shtml. Recently, brain imaging and expert testimony that convicted rapist and murder Brian Dugan was a psychopath did not save him from the death penalty. *See* Barnum & Gregory, *supra* note 53; Art Barnum & Ted Gregory, *Tears of Joy as Brian Dugan Gets Death Penalty*, CHI. TRIB., Nov. 12, 2009, at C1.

269. *See* United States v. Mezvinsky, 206 F. Supp. 2d 661, 677 (E.D. Pa. 2002) ("Upon careful scrutiny, Mezvinsky's proffered mental health defenses are founded upon a miasma of ifs, hypotheses and conjectures that have no relevance to the mental state Mezvinsky disclaims for the twelve years at issue here.").

270. For a few examples of the "fMRI as picture" rhetoric, see *June 2008 Mayo Clinic*

reductionism in the legal literature:

> "New functional magnetic resonance imaging (fMRI) technology can take pictures of a person's brain at the very moment the person is engaged in a task."[271]
>
> "fMRI can be used to take pictures in rapid succession, essentially allowing scientists to create movies of the brain as it performs different tasks."[272]
>
> "You may have seen it on the front cover of Newsweek a year or so ago. It had a picture of the juvenile brain. It's called brain imaging. It's hard science."[273]
>
> "The result is that 'areas [of the brain] with a higher concentration of oxyhemoglobin give a higher signal (a brighter image) than areas with low concentration.' The areas that are more active have more blood and literally 'light-up.'"[274]

Of course, space constraints may not allow a thorough treatment of the nuance underlying fMRI. Journalists and scholars could, however, note that the images are not snapshots, but statistical data mapped onto a 3-D template of a brain (which may not even be the subject's own brain). The overly simplified "picture of the brain" rhetoric reduces fMRI to a misleading sound bite that is

---

Health Letter Highlights New MRI Scanning Technology, Obesity-Related Cancer, and Do-It-Yourself Massage, OBESITY, FITNESS & WELLNESS WK., July 12, 2008, at 3612 ("A new tool, functional magnetic resonance imaging (fMRI), gives a clearer picture of the brain in action and what brain tissue is relevant to accomplishing a given task, such as raising a hand or reading a sentence."); *see also* Vickie Karp, *Third Screen: An Interview with Dr. Oliver Sacks*, HUFFINGTON POST, Sept. 26, 2008, http://www.huffingtonpost.com/vickie-karp/third-screen-an-interview_b_129114.html ("We have pictures from Mars and Jupiter, and we will eventually have a much more precise picture of the brain creating, remembering, and imagining."); Anne McIlroy, *Meditating Through Mental Illness*, GLOBE & MAIL, Aug. 15, 2008, at L1 ("The patients watched and reflected on scenes from sad movies, such as *Terms of Endearment*, while a functional magnetic resonance imager took a picture of their brains."); Milan Stojanovic, *Brain Imaging: A Technological Breakthrough in the Assessment of Pain*, PYSCHCENTRAL, Nov. 4, 2006, http://psychcentral.com/lib/2006/brain-imaging-a-technological-breakthough-in-the-assessment-of-pain/ ("The fMRI resembles the regular magnetic resonance imaging picture of the brain's anatomy, except that areas of the brain that are active during the experience of pain are seen as small red islands on an otherwise gray, detailed picture of the brain."); *New Research Shows How Aging Brain Brings a Healthy Dose of Perspective*, BIOTECH WK., July 2, 2008, at 4376 ("[T]heir brain activity was monitored with a functional magnetic resonance imaging (fMRI) machine, a high-tech device that uses a large magnet to take pictures inside the brain.").

271. Kittay, *supra* note 180, at 1351.

272. William Federspiel, 1984 *Arrives: Thought(Crime), Technology, and the Constitution*, 16 WM. & MARY BILL RTS. J. 865, 867 n.7 (2008).

273. From the edited transcript of the closing argument by defense attorney Craig S. Cooley in the December 2003 penalty phase of the capital murder trial of Lee Boyd Malvo. *See* Robert E. Shepherd, Jr., *Malvo Closing Argument*, CRIM. JUST., Spring 2004, at 73, 74.

274. Matthew Baptiste Holloway, Comment, *One Image, One Thousand Incriminating Words: Images of Brain Activity and the Privilege against Self-Incrimination*, 27 TEMP. J. SCI. TECH. & ENVTL. L. 141, 146 (2008) (citations omitted).

reminiscent of the "Gene for x" debacle that reinforced genetic determinism. This oversimplification will need to be overcome in the courtroom.

3. *fMRI is unfairly prejudicial as the probative content will be presumed based on what the image can prove and what it appears to prove*

For the reasons articulated *supra* in Part III.D.1 regarding the "silent witness theory" of authenticating evidence, a major problem with admitting functional brain images is that the observer may assume that the content of the image speaks with its own probative value. In other words, a testifying expert would only need to put the picture in context, and the content of the image will appear self-evident. Many researchers have cautioned that fMRI "seem[s] to offer direct access to the fact to be proven."[275] Even with proper instructions, judges and juries may subconsciously neglect the complex and idiosyncratic lab methods that are relied upon to create the image, and instead assume that they are "looking at 'someone's brain.'"[276]

The representational strategy of photographs has been referred to as "transparent immediacy," meaning that the image (here, the colorful picture of a statistical brain map) attempts to conceal the process of its construction by making the medium itself invisible.[277] There is a gap between the signifier and the signified—i.e., the *visual representation* of the brain activity is perceived to be the *brain activity itself.*[278] When we see an fMRI image, it is not at all obvious that we are looking at a statistically created map that relies on heavy processing, comparisons to averaged data, and the superimposing of data onto a separately obtained structural brain image. Instead, we think of the brain image as an immediate translation. Even established scientists are not immune; as one researcher explained, "'many people express an interest in using PET . . . because it is obviously so easy! They lack an understanding of what is entailed, I think, because the data comes out as pretty pictures.'"[279] Instead of engaging with the image as a correlative, statistical map, there is "a real danger that pictures of blobs on brains seduce one into thinking that we can now directly observe psychological constructs."[280]

---

275. Feigenson & Sherwin, *supra* note 266, at 299.

276. *Id.*

277. Jay David Bolter, *Remediation and the Desire for Immediacy*, 6 CONVERGENCE 62, 62 (2000); *see also* Robert P. Crease, *Biomedicine in the Age of Imaging*, 261 SCIENCE 554, 561 (1993) ("One obvious solution [to the floods of data] is to rely on images, whose spatial dimensions, shadings, and color codings can easily express large amounts of data.").

278. *See* JAY DAVID BOLTER & RICHARD GRUSIN, REMEDIATION: UNDERSTANDING NEW MEDIA 61 (1999).

279. Dumit, *supra* note 63, at 57.

280. Feigenson & Sherwin, *supra* note 266, at 310 (quoting Richard Henson, *What Can Functional Neuroimaging Tell the Experimental Psychologist?*, Q. J. EXPERIMENTAL PSYCHOL. 58A, 193 (2005)). This unfair prejudice is not limited to formal criminal procedure; in the context of counterterrorism interrogation, too,

fMRI as evidence thus subsumes the underlying data in a sea of colorful images.[281] Many commentators on visual media have opined on its great potential for unfair prejudice:

> [U]nlike words on the page, visual images on the screen are far more likely to directly stimulate heightened emotional responses. . . . Subject to our unthinking gaze, which is mostly how we watch, the screen seems to present a window onto reality. We tend to look *through* the medium rather than *at* it. Moreover, once we *comprehend* what we see, that's usually all we need to *believe* it. In other words, the familiar commonplace that 'seeing is believing' is not just idle folk knowledge . . . ."[282]

In this respect, it is possible that our beliefs about objects follow immediately, perhaps automatically, from our perception provided that we think we can grasp what it is we are viewing.[283] Such a belief would be reinforced by a proponent referring to an fMRI image as a photograph of someone's brain. Put another way, testimony and judicial instructions explaining an fMRI image, may be overshadowed by the apparent epistemic power of the neuroimage to convey its own message.[284]

While it may be harmless for television shows like "CSI" to fail to discriminate between science and science fiction, it is dangerous when government officials and trial counsel do not know the difference.[285] As journalists and television writers continue to perpetuate the misconception that the brain image is a picture of the defendant's "abnormal" brain, this cultural transmission increases the likelihood that a jury might give undue influence to

---

> [t]here is a profound risk that intelligence personnel will be seduced by the glamour of fMRI and its flashy images, and that they will overlook the limitations of the technology (including lack of specificity), the subjectivity of interpretation, and the complexity of brain function outside the realm of playing cards and controlled studies.

Jonathan H. Marks, *Interrogational Neuroimaging in Counterterrorism: A "No-Brainer" or a Human Rights Hazard?*, 33 AM. J.L. & MED. 483, 499 (2007).

281. Dumit, *supra* note 63, at 143 ("[T]he image overtakes the text, overturning the authority of the text.").

282. Richard K. Sherwin, *A Manifesto for Visual Legal Realism*, 40 LOY. L.A. L. REV. 719, 725-26 (2007).

283. Roskies, *supra* note 21, at 22 ("Although seeing a photograph as a representation of a scene involves an inference, the inference is generally automatic, and mediated by implicit background knowledge and judgments of visual similarity.").

284. *See* George J. Annas, Foreword, *Imagining a New Era of Neuroimaging, Neuroethics, and Neurolaw*, 33 AM. J.L. & MED. 163, 168 (2007); *see also* Dumit, *supra* note 63, at 143 ("PET images . . . participate in this reversal of veridictory authority. They do so especially when they leave the close community of researchers who daily deal with their semiotic complexity and are aware of their illustrative rather than veridictory use in scientific presentations. . . . PET images can sometimes become the central argument, with the text as supplement.").

285. *See* N.J. Schweitzer & Michael J. Saks, *The* CSI *Effect: Popular Fiction About Forensic Science Affects Public Expectations About Real Forensic Science*, 47 JURIMETRICS 357, 358 (2007) (explaining how the prosecution can be burdened by shows such as CSI, which create a demand for forensic science and a respect for it that is not supported by its actual validity).

any functional neuroimaging evidence that may be admitted.[286]

This phenomenon of thinking we are seeing a picture of someone's brain derives from what Adina Roskies has called the "inferential distance" of neuroimaging.[287] Roskies uses the term to explain the nonobvious steps that are necessary to go from the raw data to the ultimate conclusions and colorful, glossy pictures.[288] Noting the cultural and epistemic value of pictures, Roskies explains that fMRI images are belief dependent, meaning that the researcher can influence the resulting image by using a certain task, agreeing on a certain sample of controls, lowering the threshold lens through which the images are viewed, or focusing on only a few brain regions.

This dependence on the discretion of the researcher makes functional brain images more like paintings than photographs. Unlike photographs, the visual properties of functional brain images are instantiated by the use of texture, shading, perspective, and color. Thus, the brain image is not an instantiation of the functioning brain, even if it appears so. The inferential distance may not create unfair prejudice that is impossible to correct, but it may be quite difficult to retrain jurors to see a statistical map when looking at a functional image of the brain.

4. *fMRI images may be unfairly prejudicial based on neuro-essentialism*

> The brain scan image—a silhouette of the skull, highlighted with bright splotches of primary color—has also become a staple of popular culture . . . . [T]he black box of the mind has been flung wide open, allowing researchers to search for the cortical source for every flickering thought . . . . The machine, in other words, knows more about you than you do: It's like a high-tech window into the soul.[289]

Our current fascination with the brain may be part of what has been dubbed "neuro-essentialism."[290] This term describes the cultural phenomenon wherein the brain is viewed as capturing the essence of what it means to be human. In

286. *See* David L. Faigman et al., *The Limits of the Polygraph*, 20 ISSUES SCI. & TECH. 40, 45 (2003) (noting, in reference to the polygraph, the potential for prejudice when the media distorts a technology's courtroom value).

287. Roskies, *supra* note 21, at 20-22.

288. Again, the problem of inferential distance was also true with PET. Joel S. Perlmutter & Marcus E. Raichle, *In Vitro or In Vivo Receptor Binding: Where Does the Truth Lie?*, 19 ANNALS NEUROLOGY 384, 384 (1986) ("Inferences drawn from qualitative in vivo measurements . . . must be viewed with extreme caution despite their intuitive visual appeal. Unfortunately, this sort of inference is the rule rather than the exception.").

289. Jonah Lehrer, *Of Course I Love You, and I Have the Brain Scan to Prove It: We're Looking for too Much in Brain Scans*, BOSTON GLOBE, Aug. 17, 2008, at 1K, *available at* http://www.boston.com/bostonglobe/ideas/articles/2008/08/17/picturing_our_thoughts/.

290. "Although studies of the mind and brain are a cornerstone of cognitive neuroscience, neuro-essentialism represents a hasty reduction of identity to the brain." Eric Racine et al., *fMRI in the Public Eye*, 6 NATURE REVS. NEUROSCIENCE 159, 161 (2005).

the last century, advancement of neuroscience has led to the development of brain imaging and has been "said to lift mankind into a state of enlightenment about its own intellectual foundations."[291]

Functional brain images in particular speak to our desire to understand ourselves and to identify the location of the human essence as being inside the body. Therefore, when we reveal images of the brain, viewers commonly presume that these images expose not only an individual's anatomical structure, but also that individual's essence or "self."[292] This causes brain imaging to be an incredibly powerful medium, both culturally and philosophically. Thus, neuro-essentialism, as it applies in the context of functional brain imaging, may underlie a risk of unfair prejudice within the meaning of Federal Rule of Evidence 403.

5. *fMRI has low probative value because alternative, better means of proof exist*

When engaging in a balancing test under FRE 403, the judge may consider the existence of alternative means of proving the same point.[293] Where the proposed evidence risks unfair prejudice to a party, the probative value of that evidence must be discounted by the probative value of alternative means of proof. There are alternatives to fMRI that, while afflicted with serious limitations, may still be preferable to fMRI because those alternatives carry less risk of unfair prejudice.

---

291. Michael Hagner & Cornelius Borck, *Mindful Practices: On the Neurosciences in the Twentieth Century*, 14 SCI. CONTEXT 507, 508 (2001).

Colin Blakemore captured the sentiments of many by saying, "[T]he 'me-ness' of me is undoubtedly situated about two inches behind my eyes, in the very middle of my head." COLIN BLAKEMORE, MECHANICS OF THE MIND 9 (1977); *cf.* MICHAEL R. TRIMBLE, THE SOUL IN THE BRAIN 62 (2007) ("[I]n the eighteenth century, with the rise of the phrenology movement . . . [v]arious parts of the cortex were deemed to be the seats of different functions. . . . Such contentions, and the popularization of the discipline (phrenology), opened the way to considerable misuse of the ideas and to exploitation of the vulnerable at the hands of charlatans. This led to the eventual downfall of the phrenology movement, but the ideas persisted, to permeate and influence thinking in neurology for the next two centuries."). Legal scholar Hank Greely made the same point in a different way:

> If we could successfully transplant my brain into your body, would the resulting person be me with a new body or you with a new brain? I believe almost all of us would say it was me with a new body—that the "essence" of the person is the brain, not the body.

Henry T. Greely, *Neuroethics and ELSI: Similarities and Differences*, 7 MINN. J. L. SCI. & TECH. 599, 624 (2006).

292. *Cf.* Adina Roskies, *Neuroethics for the New Millenium*, 35 NEURON 21, 22 (2002) ("[I]n investigating the brain, we investigate the self.").

293. Old Chief v. United States, 519 U.S. 172, 184 (1997); 29 AM. JUR. 2D *Evidence* § 342 (2009).

a. *The first alternative to fMRI evidence is the defendant's behavior at the time of the crime*

A defendant's mental state may be inferred from the defendant's actions and the circumstances at the time of the crime. For example, intent to kill may be inferred from the manner in which the defendant approached the victim, from the manner of attack, or from the use of a deadly weapon on the victim's person. Inferences about mental state may also be made from evidence of the defendant's flight, her attempts to conceal evidence, or from the defendant and victim's prior relationship. It requires far fewer logical leaps to infer intent from evidence of what actually happened at the time of the crime than from heavily processed and statistical brain-activation maps.[294]

Recall the *Saviñon* case discussed *supra* in the Introduction. In that case, the defendant treated his ex-girlfriend's stab wounds and gave her a phone to call for help. He also made her agree not to tell anyone that he had attacked her. These attempts to conceal evidence seems to indicate the defendant's appreciation of the wrongfulness of his actions, which in turn indicates that he had intended to kill his ex-girlfriend when he attacked her.

fMRI data correlate with behavior. Thus, even if we could model events such as these using fMRI (which we presently cannot), the resulting data may have no greater legally relevant explanatory power than the behavior itself.

b. *Another alternative is behavioral psychology*

At present, fMRI is most often used as a mere legitimizing illustrative accompaniment for social and behavioral psychology data.[295] But in fact, behavioral psychology is much better equipped to introduce the nuance and subjectivity that, while some see them as weaknesses, are currently required for this type of mens rea analysis. The user of fMRI can manipulate psychological constructs and clinical assessment to evaluate an individual's executive functioning, capacity, and inhibition control, throw an incredibly isolated task at a scanner, and then end up with a complementary blob in a brain-scan image that appears to augment her psychological argument.

---

294. *See supra* Part III.F.7.

295. Matthew Hutson, *The 7th Annual Year in Ideas: Neurorealism*, N.Y. TIMES, Dec. 9, 2007, § 6 (Magazine), at 34 ("The way conclusions from cognitive neuroscience studies are reported in the popular press, 'they don't necessarily tell us anything we couldn't have found out without using a brain scanner. . . . It just looks more believable now that we have the pretty pictures.'" (quoting neuroscientist Deena Weisberg)). Some scientists have dubbed this phenomenon "neuro-realism":

> Our concept of 'neuro-realism' describes how coverage of fMRI investigations can make a phenomenon uncritically real, objective or effective in the eyes of the public. This occurs most notably when qualifications about results are not brought to the reader's attention. For example, commenting on an fMRI study of fear, one article states, "Now scientists say the feeling is not only real, but they can show what happens in the brain to cause it."

Racine et al. *supra* note , at 160 (citation omitted).

If the blob only confirms what the psychological assessment gave the user by offering an implicated area of the brain, then why use the blob at all? The reason may be that jurors are suspicious of psychological explanations, and will not accord this evidence much weight.[296] In this sense, brain images may be seen as "saving" psychological explanations in the courtroom. The irony is that, despite the jury's skepticism of nuanced psychological evidence, fMRI data is only as good as the psychological constructs relied upon to develop the experimental task.

The fact that attorneys rely on fMRI, SPECT, EEG and PET, when their experts should know that the functional data cannot make useful individual inferences about past mental states, begs the question of whether the parties are using the images *specifically for* their unfairly prejudicial effect. That is, are attorneys using brain scans because they know they will unduly influence the jury to accept psychological constructs that would otherwise be suspect as "soft" science?

6. *fMRI is unfairly prejudicial as it encourages the fundamental psycholegal error*

"Psycholegal error" is a term coined by Stephen Morse that refers to the tendency to think that an actor is not responsible for behavior that is caused by the actor's brain function or genes.[297] This is improper reasoning because in fact *all* human behavior is caused by the flow of neurotransmitter in the brain and because the law cares about the *effect* of a cause on a legally relevant excusing or mitigating condition rather than the cause itself.

Neuroscientific evidence could lead a factfinder properly to mitigation or excuse of a defendant's behavior, but only if that evidence does more than point to an ordinary biological basis for the defendant's criminal actions. For example, neuroscientific evidence could be used to support a theory of reduced

---

296. *Cf.* Rick Brown, Note, *Limitations on Expert Testimony on the Battered Woman Syndrome in Homicide Cases: The Return of the Ultimate Issue Rule*, 32 ARIZ. L. REV. 665, 673 (1990) ("Jurors are not likely to give unquestioning acceptance to psychological expert testimony."). In addition, psychiatric testimony only needs to "assist the jury" as required by FRE 702. *Id.* at 671. This low standard may be warranted for psychological testimony because there is less risk that the jury will be overawed, as compared with the "hard" sciences. *Id.* at 673.

297. According to Morse, this error "bedevils thinking about the relation between scientific discoveries of causes of behavior and traditional conceptions of responsibility. Discovering a cause for behavior, whether it is biological, psychological or sociological, does not mean that the agent is not responsible for the behavior." Stephen J. Morse, *Criminal Responsibility and the Disappearing Person*, 28 CARDOZO L. REV. 2545, 2569 (2007).

Even if neuroscience develops airtight theories that illustrate biological causes that predispose people to behave as they do, "the science typically supporting the more general and radical claim that conscious will is an illusion—that we do not act and therefore cannot be responsible—is either insufficient empirically or does not have the implications claimed." *Id.* at 2570.

mental capacity, deficient executive function, or lack of inhibitory control that produced a legally relevant excusing condition.[298] Brain images that indicate structural or functional lesions are relevant to such findings. However, fMRI does not by itself explain a behavioral deficit. Rather, it demonstrates which areas of a person's brain are engaged when that person performs particular mental actions. Because it indicates nothing more than a correlation between biological functions and behavior, it provides fodder for factfinders to fall prey to psycholegal error and the mitigation or excuse of criminal behavior. As discussed above, this is an improper basis for juror decision-making, and thus risks unfair prejudice.

7. *fMRI evidence is unfairly prejudicial as it impairs factfinders' ability to assess evidence*

Converging evidence indicates that it may be very difficult for jurors to properly evaluate fMRI-based evidence. For example, recent work demonstrates the public's extreme unfamiliarity with the complexity of producing brain images.[299] Furthermore, researchers have long worried that neuroimaging methodologies may increase the risk for misuse, given that individuals may be lured by their high-tech profile.[300] This problem may be simply one of a technology in transition. Over time, with the cooperation of neuroscientists and journalists, the popular press may better communicate the achievements and limitations of neuroimaging technology, which will hopefully reduce its capacity for unfair prejudice. But we are not there yet, and it is not just the lay public that is confused about the claims that can be made by

---

298. Stephen J. Morse, *Addiction, Genetics, and Criminal Responsibility*, 69 LAW & CONTEMP. PROBS. 165, 174-75 (2006) ("Abnormal physical variables, such as neurotransmitter deficiencies, may cause a genuine excusing condition, such as the lack of rational capacity, but then the lack of rational capacity, not causation, is doing the excusing work.").

299. A Brazilian survey of "neuroscience literacy" showed that the public is quite uninformed about imaging methods. Suzana Herculano-Houzel, *Do You Know Your Brain? A Survey on Public Neuroscience Literacy at the Closing of the Decade of the Brain*, 8 NEUROSCIENTIST 98, 110 (2002).

300. *See, e.g.*, Martha J. Farah & Paul Root Wolpe, *Monitoring and Manipulating Brain Function: New Neuroscience Technologies and Their Ethical Implications*, 34 HASTINGS CENTER REP. 35, 40 (2004) (exploring the potential misuse of neuroimages, which "possess an illusory accuracy and objectivity"); Trisha Gura, *Big Plans for Little Brains*, 435 NATURE 1156, 1158 (2005) ("'You never can tell where research is going to lead . . . [b]ut the danger to everyone . . . is expecting too much, too soon.'" (quoting John Bruer, president of the James S. McDonnell Foundation)); Jennifer Kulynych, *Brain, Mind and Criminal Behavior: Neuroimages as Scientific Evidence*, 36 JURIMETRICS J. 235, 237 (1996) ("[S]ome researchers fear that the premature use of brain images to predict behavior has introduced a new form of 'junk science' in the courtroom."); Robert S. Stufflebeam & William Bechtel, *PET: Exploring the Myth and the Method*, 64 PHIL. SCI. S95, S95-S96 (1997) (praising the "dramatic character of these images," but critically evaluating the limited ability of PET to map onto psychological functions).

fMRI. In fact, as one respected imaging expert pointed out, the "massively multidimensional" aspect of producing neuroimages has led to methodological mistakes by researchers themselves.[301]

Two studies published in 2008 provide empirical support for our argument that it may be extremely difficult for jurors properly to evaluate fMRI evidence.

The first study, from Yale University,[302] began with the hypothesis that subjects would accord too much weight to the explanatory power of irrelevant neuroscience data. The researchers split the eighty-one subjects into three groups: those untrained in psychology and neuroscience, students in a neuroscience course, and neuroscience experts. The subjects were given brief descriptions of psychological phenomena. The researchers then gave the subjects either a "good" explanation or a "bad" explanation (involving circular reasoning) of the psychological phenomenon. The explanations also varied in language. Some used plain English, while others used neuroscience lingo. The varied language was logically irrelevant—it added zero explanatory power.[303]

Subjects of all three groups were capable of judging good explanations more persuasive than bad ones. But bad explanations accompanied by irrelevant neuroscience lingo distorted the understanding of subjects from the two non-expert groups. The researchers found that, for these two groups, the mere presentation of phrases such as "brain scans indicate" turned a bad explanation into a good one, "masking otherwise salient problems in these explanations" and making the explanations appear to be more persuasive.[304] Thus, according to the Yale study, the presentation of logically irrelevant neuroscience data can give people the mistaken feeling that they have just read a good explanation of psychological phenomena.[305] In the courtroom context, this should be characterized as a risk of unfair prejudice. This is demonstrated below in Figure

---

301. Nonindependence bias results when there are correlations or relationships between individual measures in research that render the subsequent outcomes nonrandom. This occurs in fMRI studies when regions of interest are selected for follow-up and their selection affects any comparison to other brain regions for a measure of relative activation. *See* Edward Vul et al., *Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition*, 4 PERSP. ON PSYCHOL. SCI. 274, 279 (2009), *available at* http://www.edvul.com/pdf/VulHarrisWinkielmanPashler-PPS-2009.pdf; *see also* Edward Vul & Nancy Kanwisher, *Begging the Question: The Non-Independence Error in fMRI Data Analysis*, *in* FOUNDATIONAL ISSUES IN HUMAN BRAIN MAPPING (Stephen Jose Hanson & Martin Bunzl eds.) (forthcoming 2010), *available at* http://www.edvul.com/pdf/VulKanwisher-chapter-inpress.pdf ("[The non-independence error] is prevalent in cognitive neuroscience . . . . [O]f the eight papers in a recent special issue of *Neuroimage*, five contained variants of this error.").

302. Deena Skolnick Weisberg et al., *The Seductive Allure of Neuroscience Explanations*, 20 J. COGNITIVE NEUROSCIENCE 470, 470 (2008).

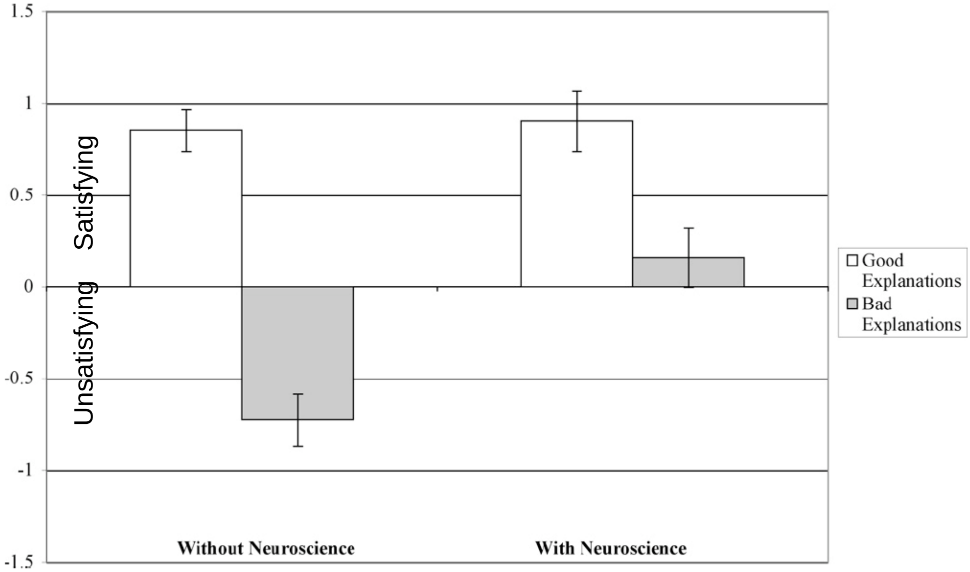
303. *Id.* at 471.

304. *Id.* at 470.

305. *See* J.D. Trout, *Seduction Without Cause: Uncovering Explanatory Neurophilia*, 12 TRENDS COGNITIVE SCI. 281, 281-82 (2008) (discussing the implications of the Weisberg et al. study).

5.

Figure 5



In a separate study conducted at Colorado State University, researchers examined "whether brain images actually do have a particularly powerful persuasive influence on the perceived credibility of cognitive neuroscience data."[306] The researchers presented undergraduates aged eighteen to twenty-five with brief fictional articles that summarized cognitive neuroscience research results. All of the articles contained serious logical errors regarding psychological phenomena. Some of these articles contained neuroscientific images; others did not.[307] After reading the articles, subjects were asked whether "(1) [t]he article was well written, (2) [t]he title was a good description of the results, and (3) [t]he scientific reasoning in the article made sense."[308] The presentation of neuroscientific images did not substantially effect the question about the title, but it substantially impacted the questions of whether the article was well-written and whether the reasoning made sense. As the data below demonstrate, subjects gave higher rankings to text that was accompanied

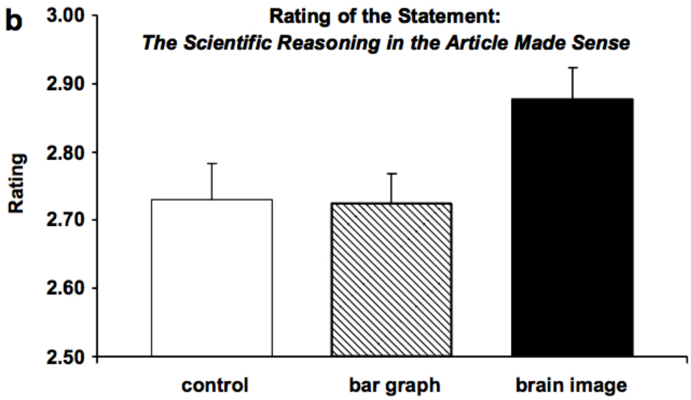
306. David P. McCabe & Alan D. Castel, *Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning*, 107 COGNITION 343, 344 (2008).

307. One example of a stimulus was as follows: Subjects were given an article that reported that watching TV and completing math problems both led to activation in the temporal lobe. The article then concluded that because these two events activated the same lobe of the brain, watching TV improved math skills. The similarity in brain activation was depicted by plain text, bar graph, or brain image. *Id.* at 343.

308. *Id.* at 345 (emphasis omitted).

by a brain image than to text that was accompanied by a bar graph or to text with no image at all.[309]

Figure 6: "Watching TV helps with math ability because both activate the temporal lobe."



These data lend support to the argument that brain images have unique persuasive power, causing viewers to overlook serious logical errors and, therefore, to make improper inferences. Improper bases for decisionmaking such as these are a quintessential form of unfair prejudice, one that is clearly risked by evidentiary use of fMRI.[310]

H. *Cross-Examination Is Not the Cure*

An argument has been made that the answer to the problem of unfair prejudice is to admit more, not fewer, images by allowing rigorous cross-examination and introduction of competing images.[311] Under this theory, more information will lead to jurors assessing brain images more reliably. Ostensibly, competing images would enable jurors to assess the relative

309. *Id.*

310. *Id.* at 343 ("These data lend support to the notion that part of the fascination, and the credibility, of brain imaging research lies in the persuasive power of the actual brain images themselves. We argue that brain images are influential because they provide a physical basis for abstract cognitive processes."). Unfortunately, as much of this type of research is done on psychology undergraduate students, and we are dealing with constructs such as the neurobiology of behavior, the data should be interpreted with a grain of salt. The external validity remains to be tested on a broader pool of the population from which most jurors will actually be selected.

311. *See* Neal Feigenson, *Brain Imaging and Courtroom Evidence: On the Admissibility and Persuasiveness of fMRI*, 2 INT'L J.L. CONTEXT 233, 250 (2006) ("The cure for naïve realism is more, not fewer images.").

strength of the overall story. This argument would be more compelling if we were confident in the robustness of the authentication process. This point aside, cross-examination is a very imperfect process for getting at truth, especially in the case of complex, high-impact technology as fMRI.[312]

Cross-examination might reveal that an expert's statistical thresholds are poor or inconsistent. It might also uncover poor study design or conflicting potential interpretations. But the weaknesses of using fMRI in the courtroom extend far beyond deviation from best research practices. Truth cannot be exposed through the adversarial process when there is no agreed-upon "truth." The current state of the science is up in the air, as the relationship between brain states and mental states is still being worked out. Furthermore, even when the science is done well, there will be serious disagreement about how to interpret the results. Therefore, cross-examination can do little to get us closer to the truth, and it cannot be relied upon to completely purge unfair prejudice.[313]

Unfortunately, neither rehabilitative testimony nor further cross-examination can currently fix this attack. Even after considerable waste of court and party resources, we would be left with jurors' shrugged shoulders and an acknowledgment that the brain images are only a prop developed to bolster the psychological testing or observed behavior. This is precisely what happened in the *Saviñon* case, where the defense counsel ultimately admitted that the brain images were just superfluous, but only after wasting much of the prosecution's and judge's time. Prosecution and public defender offices are already trying to do too much with too little, and ordering expensive fMRI analyses and experts is a waste of precious resources. Because the interpretive standards of producing a functional brain image are subjective, and because the relationship between brain states and mental states is still being worked out, cross-examination is a poor tool for getting at the "true" mental state. Therefore, being presented with more or conflicting images from either side will likely only confuse jurors and waste the court's time. In fact, increasing the jury's cognitive load, or decision-making burden, may lead to less thoughtful decisions.[314] This alone is an acceptable reason to exclude the brain images, as

312. *See* Jeffrey Bellin, *The Significance (if Any) for the Federal Criminal Justice System of Advances in Lie Detector Technology*, 80 TEMP. L. REV. 711, 711 (2007) ("[F]ew would deny that [cross-examination] has achieved . . . exalted status largely by default. Science has simply failed to produce any valid alternative."); Laura Stephens Khoshbin & Shahram Khoshbin, *Imaging the Mind, Minding the Image: An Historical Introduction to Neuroimaging and the Law*, 33 AM. J.L. & MED. 171, 192 (2007) ("[T]he adversarial system is an inadequate forum for determining the evidentiary validity of such evidence.").

313. *See* IMWINKELRIED, *supra* note , § 13-11, at 431.

314. *See, e.g.*, Joshua D. Greene et al., *Cognitive Load Selectively Interferes with Utilitarian Moral Judgment*, 107 COGNITION 1144, 1144 (2008) (finding that "a cognitive load manipulation selectively interferes with utilitarian judgment"); *cf.* Nadine Marcus et al., *Understanding Instructions*, 88 J. EDUC. PSYCHOL. 49, 49 (1996) ("Instructional material may be difficult to understand if it consists of many elements that must be held in working

the judge has "'wide latitude to impose reasonable limits' on cross-examination in order to avoid confusion of the issues or extended discussion of marginally relevant material."[315] If fMRI adds almost no probative value, while introducing much interpretive ambiguity, unfair prejudice, or confusion, the images should be excluded based on FRE 403.

CONCLUSION

A. *FRE 403 Provides Adequate Grounds for Exclusion Based on the Potential for Unfair Prejudice and Nearly Bankrupt Probative Value*

Relevant evidence can be excluded if its probative value is substantially outweighed by the dangers of unfair prejudice, confusion of the issues, and misleading the jury. A great deal of evidence is excluded on these grounds, and trial judges have considerable discretion in making this determination.

We have demonstrated here that the probative value of fMRI addressed to a defendant's mens rea is extraordinarily low, such that a modest amount of unfair prejudice should render these images inadmissible.

Functional neuroimaging promotes unfair prejudice and confusion of the issues because it causes jurors to ground their decisionmaking in emotional responses to images and distracts jurors from logical errors, thus causing them to make decisions founded on improper bases. Additionally, it wastes the court's time and resources.[316] While more research on mock jurors needs to be done to bolster or refute our claims, the preliminary data suggest that fMRI could be excluded based on its potential to bias jurors and make them find otherwise bad arguments more compelling.[317]

B. *Closing Thoughts on the Critical Legal View of the Allure of fMRI*

The familiar story is one of weak circumstantial evidence and impressive scientific findings. The combination of these elements may be a powerful

---

memory simultaneously."); John Sweller, *Cognitive Load Theory, Learning Difficulty, and Instructional Design*, 4 LEARNING & INSTRUCTION 295, 303 (1994).

315. United States v. Gonzalez-Vazquez, 219 F.3d 37, 45 (1st Cir. 2000) (quoting United States v. Twomey, 806 F.2d 1136, 1139 (1st Cir. 1986)); *see also* United States v. Ambers, 85 F.3d 173, 177 (4th Cir. 1996) ("To entitle defense counsel to explore the intricacies of the [evidence] on cross-examination might do much to confuse lay jurors and little to enlighten them.").

316. *See* Michael Avery, *Prejudice vs. Probative Value,* Philadelphia *Style*, 50 ST. LOUIS U. L.J. 1147, 1147 & n.5 (2006) ("[I]f the defense lawyer cannot convince the trial judge to exclude such photographs, the chances of winning the issue on appeal are remote.").

317. *See generally* Edward J. Imwinkelried, *The Meaning of Probative Value and Prejudice in Federal Rule of Evidence 403: Can Rule 403 Be Used to Resurrect the Common Law of Evidence?*, 41 VAND. L. REV. 879, 905 (1988) (discussing when Rule 403 should be used to "shield the jury from evidence").

prescription for injustice: scientific evidence seems so compelling that it could sway even the most skeptical juror and convince him that the elements of a weak case are proved beyond a reasonable doubt.[318] If, on the other hand, the defendant catches the court's sympathies, then the junk science may swing in the opposite direction and make a weak defense appear stronger.[319] This story has played out before with phrenology,[320] the polygraph,[321] and countless other forensic technologies that have since been discredited. Improper reliance on each of these untested and unreliable technologies has led to unjust outcomes.[322]

These older forensic technologies all have the window dressings of science.[323] Each supplies the court with lab coat-wearing experts who will speak to analyses of "matching" criteria with confidence that their methods are

---

318. Perhaps this is what happened in a recent case in Pune, India, where Aditi Sharma was convicted of murdering her ex-fiance, Udit Bharati. The court repeatedly mentioned the fact that the marriage between Aditi and Udit was approved by their parents, and it is difficult not to consider the possibility that Aditi's acts were considered particularly heinous in light of her breaking off her engagement to Udit. The court's opinion contained many logical errors, and the shaky evidence was all drawn in the light most negative to the defendant. With no clear proof of intent or action, an untested brain-based technology was used to prove that Aditi had "guilty knowledge" of the details of the crime. The opinion does not say how the technology is capable of discerning whether Aditi's brain appeared to recognize facts of the case because she experienced them, versus her awareness because these statements were present in her charge sheet. The full opinion is available at http://lawandbiosciences.files.wordpress.com/2008/12/beosruling2.pdf.

319. Dan White, the former San Francisco supervisor who assassinated Harvey Milk and Mayor George Moscone, famously entered a diminished capacity defense based largely on his depression, and minimally upon his diet of junk food. The junk food was introduced as a symptom of his depression rather than the cause, but the press exploited the latter, even thought it was only briefly mentioned by defense attorney Douglas Schmidt in his closing argument. *See* Kelly Snider, *The Infamous Twinkie Defense—Fact or Fiction?*, 9 ANNALS AM. PSYCHOTHERAPY ASS'N. 42, 43 (2006). In any event, the defense worked to reduce White's conviction from first-degree murder to manslaughter. *See* People v. White, 172 Cal. Rptr. 612, 612 (Ct. App. 1981). The people of San Francisco rioted in response, and the diminished capacity defense was later abolished in California. *See In re* Christian S., 872 P.2d 574, 575 (Cal. 1994).

320. *See supra* note 3.

321. *See supra* note 35; *see generally* Wygant, *supra* note 4, at 330-31.

322. Comparative bullet lead analysis (CBLA) is just one example. CBLA has "recently been abandoned after decades of use because an exhaustive analysis by the National Research Council concluded that there was no scientific basis for the claim that it could 'match' crime scene bullets to particular boxes of bullets." Keith A. Findley, *Innocents at Risk: Adversary Imbalance, Forensic Science, and the Search for Truth*, 38 SETON HALL L. REV. 893, 935 (2008); *see also, e.g.*, Bowling v. Commonwealth, No. 2006-SC-000034-MR, 2008 WL 4291670, at *2 (Ky. Sept. 18, 2008) (denying defendant's motion for a new trial, despite the introduction of now-discredited CBLA evidence, on grounds of harmless error).

323. *Cf.* David Faigman, *Anecdotal Forensics, Phrenology, and Other Abject Lessons from the History of Science*, 59 HASTINGS L.J. 979, 979 (2008) (describing the "varying merit" of various "claims to the mantle of science").

sound. But despite popular appeal, phrenology, polygraphy, and fingerprint and handwriting analysis have never had the ringing endorsement of mainstream physical or biological sciences. In addition, empirical studies have confirmed that there is little reliability or validity in many of these methodologies.[324] However, unlike these sensationalized forensic sciences, functional neuroimaging has the imprimatur of the scientific research community. Indeed, it is difficult to open a copy of *Nature* or *Science* without eyeing several colorful functional brain images.

Perhaps, then, the once fledgling field of genetics can provide a more appropriate analogy. So long as genetic samples are not contaminated, the ability to exclude someone from a suspect list based on modern DNA testing is fairly robust. Even so, recall that it took many years for DNA evidence to arrive at the presently-understood state of fallible yet scientifically-valid evidence. However, before the lab standards and analytical models were fully vetted, suspects were unfortunately charged based on DNA samples that were later found to have been carelessly analyzed.[325]

Science can appear to be beyond the reach of human distortion. As a result, the more the scientific evidence relies on complex technologies like computers or imaging devices, the greater the risk that it may be endowed with powers to solve difficult legal questions. Litigants have long used this fact to their advantage, stretching scientific findings in order to retrofit them to legal conclusions. This may be what is happening with fMRI. The device is not yet capable of capturing past mental states, but because the criminal law is sometimes desperate to prove the unprovable, there will almost surely be an increase in proffered evidence and testimony based on this new technology. However, until fMRI is able to reliably capture past mental states, this evidence should not be admissible for such purposes either under FRE 403 or under local standards for admissibility of scientific evidence.

---

324. *See* Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 892 (2005) ("The assumption of discernible uniqueness that resides at the core of these fields is weakened by evidence of errors in proficiency testing and in actual cases."); *see also* Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 TEMP. L. REV. 55, 56 (1998) ("[H]air identification, bitemark analysis, voice spectography, handwriting analysis . . . and . . . fingerprinting[] have crept into court with virtually no demonstration of their scientific bases."); Craig M. Cooley, *Reforming the Forensic Science Community to Avert the Ultimate Injustice*, 15 STAN. L. & POL'Y REV. 381, 381-82 (2004) (providing examples of many forensic science errors that have carried great weight in trials).

325. In one instance of carelessly analyzed DNA, the defense thankfully was able to challenge the genetic testing procedure, and teased out the lack of testing on the criteria used to determine a DNA "match." *See* People v. Castro, 545 N.Y.S.2d 985, 997 (N.Y. Sup. Ct. 1989). For an example of how Australian courts have treated errors in DNA analysis, see R. v. Lucas, (1991) 55 A. Crim. R. 361.

APPENDIX: CHECKLIST FOR JUDGES CONFRONTED WITH FUNCTIONAL NEUROIMAGING EVIDENCE

It is possible that the validity and probative value of fMRI in assessing mental states will improve in the future. It is also possible that the public understanding of the inferential leaps required by functional imaging techniques will progress such that the technology carries less risk of unfair prejudice. In the event that these twin events occur, judges and opposing counsel may appreciate having access to a simple checklist of questions that they can pose when deciding whether to admit fMRI evidence. Although our thesis is addressed to functional brain images used to prove mens rea, this checklist could also be applied to fMRI used as evidence of lie detection and other mental states.

General questions to ask counsel seeking to introduce functional neuroimaging evidence:

(a) *Behavioral task*. What is the particular behavior assessed during the scan? Why was the particular behavioral task chosen? Is it well supported in the psychological literature as best capturing this type of mental state? Did the subject perform the behavioral task adequately? Is the task vulnerable to manipulation, countermeasures, or malingering? Are the subject's behavioral data within or significantly outside the normal distribution of performance on the task?

(b) *Controls*. How were the controls selected to be in the control group? Are they the correct reference class? What sort of testing was done on the controls to make sure that they were in fact, "normal"? Is the sample size large enough to capture normal variance between subjects?

(c) *Variance*. Can you show us the brain scans of the control group, and are there significant differences among the individuals in this group? How much difference between individuals do we see?

(d) *Image construction*. Please walk us through the process for developing the image. How did you go from the raw data in the scanner to the color picture of the brain? Can you provide the raw data and exact methodology to an independent party for verification of the image creation process?

(e) *Alternative explanations*. What are possible alternate explanations for this behavior and corresponding neural activation correlates (i.e., expertise in the task, medication status, drug abuse history, hormonal fluctuations, language or motor limitation, etc.)?

(f) *Purpose of fMRI evidence*. What justifies the introduction of this brain image over evidence of the accused's behavior at the time of the crime?

(g) *Statistical threshold*. What statistical threshold was used to create the image? Why was it used?

(h) *Causal connection*. Is there a known or hypothesized mechanism causally connecting any perceived brain abnormality to a functional deficit? Do we have any data on the incidence of reduced metabolic or hemodynamic activity of this kind resulting in this type of cognitive deficit?