Cross-Examining the Brain: A Legal Analysis of Neural Imaging for Credibility Impeachment

Charles N. W. Keckler

Follow this and additional works at: https://repository.uchastings.edu/hastings_law_journal

Part of the Law Commons

Recommended Citation
Available at: https://repository.uchastings.edu/hastings_law_journal/vol57/iss3/2

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Law Journal by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.
Cross-Examining the Brain:  
A Legal Analysis of Neural Imaging for Credibility Impeachment

CHARLES N. W. KECKLER*

But where are these practical psychological tests, which will detect specifically the memory-failure and the lie on the witness-stand? ... If there is ever devised a psychological test for the valuation of witnesses, the law will run to meet it ... Whenever the Psychologist is really ready for the Courts, the Courts are ready for him.

INTRODUCTION

The premise of this Article is simple: "the Psychologist" is almost ready for "the Courts" several decades after Wigmore began waiting for him, but the Courts must be made ready, rather than be relied upon to "rush to embrace" any advances in detecting deception. Since technology now allows the observation of the internal neurological processes by which deceptive information is produced, an observer adjudicating credibility need no longer be limited to the psychosomatic responses measured by either demeanor observation or polygraph interpretation, which are poorly correlated in principle, and practice, to deceptive acts. In Part I, this Article describes and evaluates the scientific changes that have created this potential for detecting deception, which is only just beginning to be exploited. A major element has been the expansion of cognitive neuroscience, which is a theoretical orientation toward mental phenomena that first characterizes with precision the information and processing steps needed to accomplish some mental task, and secondly identifies within the brain those structures that actually perform the

* Visiting Assistant Professor, George Mason University School of Law. Ph.D. (all but dissertation, Human Evolutionary Ecology), University of New Mexico; J.D., University of Michigan, 1999; M.A. (Anthropology), University of Michigan, 1992; A.B., Harvard College, 1990. Initial research for this Article occurred during my residence as John M. Olin Fellow at Northwestern University School of Law. Further support during the preparation of the Article was provided by both the Law and Economics Center, and the National Center for Technology and Law, both located at George Mason University School of Law. The author may be reached at ckeckler@gmu.edu.

1. JOHN H. WIGMORE, WIGMORE ON EVIDENCE § 875 (2d ed. 1935).
information processing steps involved. It is the recent technological changes in brain imaging, particularly the visualization of the brain while it is actively working—so called “functional imaging”—that have allowed cognitive neuroscience the potential to identify relatively subtle processes such as deception. Moreover, the ability to examine in real time the response of the subject brain during a question and answer session makes it feasible to use this technique forensically, so long as the pattern of brain activity corresponding to deception is sufficiently well-characterized.

In order to more precisely specify the research and evidentiary potential of this technique, I break down deception into three different types of mental operations, which show distinct patterns and differ in their practical or potential tractability to accurate measurement. Perhaps the most straightforward testing paradigm is feigned ignorance, sometimes called “guilty knowledge” in polygraph literature, or malingering. In this mental operation, it is relevant whether the subject knows X or not; the deceptive subject (falsely) denies knowledge of X, and the detection of deception device (DDD) attempts to accurately distinguish between the presence and absence of X, which, respectively would either bolster or contradict this denial. More problematic but potentially more valuable is the use of a DDD in the classic circumstances of the lie, wherein the subject is asked as to the fact of some matter and instead of responding sincerely with X, responds with Y, a falsehood about that state of affairs to which X also refers. The DDD in this context is required to accurately identify when the subject’s brain is formulating a verbal response in conflict with a (different and presumably sincere) response evoked directly from the subject’s memory. Most generally and most problematically, we would ideally like to detect deception when the subject is neither denying knowledge nor uttering falsehoods, but is simply misleading the questioner because they have a subjective intent to deceive. Note that it is only in this final paradigm that we are actually measuring “deception” per se, rather than mental operations that may distinguish deception from normal communication. Hypothetically, a DDD could be used to identify the presence of this intent in the subject.

2. Gershon Ben-Shakhar et al., Trial by Polygraph: Reconsidering the Use of the Guilty Knowledge Technique in Court, 26 LAW & HUM. BEHAV. 527, 530 (2002).
4. This type of deception, of course, encompasses the preceding two kinds; that is, if there truly were a DDD capable of showing intent to deceive, a subject denying knowledge he possesses, or choosing to utter a lie, would be detected by the intent—more specifically, the hypothetical distinctive cognitive planning necessary to deceive—that necessarily preceded his deceptive behavior. It should be pointed out that none of these definitions precisely track the law of perjury, which would require a false material statement (and this can be a denial of knowledge), along with criminal intent. See
The goal and measure of a DDD will be to demonstrate accuracy in terms of empirical sensitivity (picking up all instances of the deception type) and specificity (not picking up other phenomena not demonstrating deception). As statistical measures, sensitivity corresponds to a low rate of "Type II" errors, sometimes called "false negatives," while specificity reduces "Type I" errors, or "false positives." Moreover, the test will also have to possess sufficient theoretical justification for what it measures in order to satisfy all the scientific criteria for admissibility. Along with discussing neuroimaging, and in particular, functional magnetic resonance imaging (fMRI), as the most promising techniques to lay the groundwork of this research, I also discuss in a more cursory fashion the use of electroencephalograms (EEGs). Instead of showing actual brain structures, these devices record electrical activity throughout the brain by the attachment of external electrodes, and have been promoted under the label "brain fingerprinting" as a DDD for denied knowledge; in addition, I will occasionally for purposes of contrast, discuss the polygraph research as it relates to the denied knowledge, the lie, or the intent to deceive paradigms.

Having reviewed the facts, I present in Part II a method of potentially integrating this research into the legal arena, beginning with an assessment under Federal Rule of Evidence 702, of its capacity to satisfy the scientific criteria of admissibility. In order to create a "virtuous cycle" of increasing accuracy and increasing court use, I propose a model that begins with limited admissibility in those contexts most likely to encourage increased rigor—namely, when the proponent is adverse to the witness tested, a circumstance that implies the use of fMRI initially for impeachment rather than substantive evidence. This way there can be mutual benefit for civil plaintiffs and defendants to "cross-examine" the brain of a witness whose credibility has been put in doubt.

**Historical Context**

Dean Wigmore's confident prediction reflects the dubious quality of legal judgments of credibility, and the unfilled demand for any form of accurate assistance. The exclusion of DDDs from the trial setting is no small part of the history of modern evidence. Until very recently, this interaction involved only a recapitulation of the perennial questions surrounding the polygraph, and repeated rejections modeled on the very Frye decision which had made the art of the lie detector the gold (or perhaps pyrite) standard of an unscientific means to ascertain the truth.\(^5\)

---

\(^5\) Frye v. United States, 293 F.1013 (D.C. Cir. 1923).

\(^6\) DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: SOCIAL AND BEHAVIORAL SCIENCE ISSUES 584--
Hastings Law Journal

Hypothetically, at least in the limited universe of civil cases where other evidence is quite equivocal, even a questionable instrument like the polygraph might be of value, but an absolute bar rather than a conditional one has always been the rule. Or as Justice Thomas put it in conclusive terms: "A fundamental premise of our criminal trial system is that 'the jury is the lie detector.'"

This blanket exclusion of a method, of course, provides little incentive for its scientific improvement. Indeed the polygraph of today differs little in its essential features (and vices) from the earliest models. Blaming Frye for this, or even the professional interest of the bar, would be excessive, as the primary culprit is the polygraph itself, whose theoretical and practical flaws are such that the debate has generally been whether it is "completely useless" or just "usually useless" (which suggests also "occasionally useful"). The experience of Massachusetts, for instance, suggests that even with an incentive to improve, the ability of the polygraph to progress remains weak. Having recognized the negative incentives of a blanket prohibition, Massachusetts made polygraph evidence admissible by discretion, with the explicit hope that this would cause advocates and researchers to improve their techniques now that there would be some kind of marginal benefit in the form of increased admissibility for increased accuracy. After waiting more than a decade, however, Massachusetts reverted to its older rule—noting the conspicuous failure of the polygraph to respond to the opportunity they had presented.

7. Insofar as an instrument gives results that are even slightly above chance—something that even strenuous critics of polygraphy, including myself, do not deny—it will alter the probability of reaching the "correct" result in a case whose probable outcome absent polygraphic evidence is fifty-fifty. Such circumstances blunt one of the major objections to polygraph evidence, because even if fact-finders mistakenly believe the device to be more accurate than it actually is, they will arrive (on average) at the correct legal outcome, their only error of "prejudice" being that they are more confident of the decision than they ought to be.
8. United States v. Sanchez, 118 F.3d 192, 197 (4th Cir. 1997) (holding that the traditional rule that polygraph evidence is never admissible to impeach the credibility of a witness "is binding upon us in this case" (emphasis added)); accord United States v. Scheffer, 523 U.S. 303, 311 (1998); United States v. Prince-Oyibo 320 F.3d 494, 501 (4th Cir. 2003).
9. Scheffer, 523 U.S. at 313. Although this might be so at the present time—it is not "fundamental" in the way that term is usually employed, as a historically invariant or logically necessary feature. See George Fisher, The Jury's Rise as a Lie Detector, 107 Yale L.J. 575, 579 (1997) ("When and why did the system declare that jurors had the wisdom to arbitrate unvarnished credibility conflicts at criminal trials? To the question 'when,' the surprising answer is very recently.").
11. For an updated version of this debate, see David L. Faigman et al., Modern Scientific Evidence: The Law and Science of Expert Testimony ch. 10 (2d ed. 2002).
Given the stagnation in polygraph technology and technique, the more interesting question might be why its admissibility remains such a live and recurrent issue. One explanation might be that some litigants, when the facts and the law seem to be going against them, may resort to a polygraph, thinking that a favorable result might occur and they have in effect nothing to lose (since they would, in the absence of stipulation, be left free to introduce the test or not). *Ceteris paribus*, this would attract those with the weakest cases, who are disproportionately deceptive in their litigation position. Consequently they would be attracted by the very *unreliability* of the polygraph test (and the more unreliably administered the better for them), since a favorable result would necessarily be an inaccurate one. Fifty-fifty odds or worse of such a result (technically, a false negative showing no deception)\(^4\) might be worth the gamble to such a desperate party. For instance, it might generate reasonable doubt for a criminal defendant—a common context for an attempt at polygraph admissibility.\(^5\)

Assuming this is *part* of the answer that explains when admissibility is actually sought, the more fundamental reason why polygraph evidence will not "go away" is that it remains relevant to recurrent issues of credibility in adjudication, and the substitute methods of adjudicating these issues—usually through the naked eye assessment of witness demeanor—are not superior to the polygraph. Jurors assess demeanor evidence when evaluating the credibility of the accompanying statements...

---

\(^4\) For purposes of clarity, what I mean by accuracy is a low rate of Type I error, detecting "something" when there is actually nothing, as well as a low rate of Type II error, detecting "nothing" when what is being sought for is actually present. A Type I error is sometimes called a "false positive," and a Type II error a "false negative." In this Article, the "something" is deception—truth or falsity itself. I generally avoid these labels, as they would lead to difficult locutions, such as an unwarranted positive being a "false falsity." Rather, I will speak of the rates of undetected deceit or "evasions" (the Type II problem) and truth mischaracterized or "artifacts" (the Type I problem).

\(^5\) An uncontrolled "market" for self-administered detection of deception devices has many of the classic hallmarks of a "market for lemons." George Akerlof, *The Market for Lemons: Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488, 488–500 (1973). The polygraph is known to produce errors, at rate \(E\), and different methods and testers have different rates of error, which are not visible to the consumer. The consumer comes in two types, either honest or deceptive, and therefore favors a low \(E\) or high \(E\), respectively. Quite apart from any estimate of the relative frequency of these types who become criminal defendants, the chances are that any other evidence in the case will tend to corroborate honesty and contradict deceit. Therefore, the honest person will find the evidence largely redundant (at rate 1 - \(E\)) and damaging at rate \(E\)—a risk rarely worth taking. The deceitful will more often find themselves in circumstances where their expected rates of loss exceed 1 - \(E\), meaning they attempt a polygraph, because it gives them \(E\) chance of an inaccurate result, reasonable doubt, and acquittal. The rate at which the evidence is admitted is irrelevant to this calculation since it merely reduces the potential benefit of an erroneous test, without its ceasing to be of marginal benefit. To the extent the foregoing is true, the less accurate methods of the polygraph will be chosen by the dominant type of consumer, and the average level of \(E\) will increase (or at the very least not fall), meaning the discipline would become less accurate over time, or least no more accurate. I discuss some ways to get around this problem of adverse selection. Part II.E *infra*. In this case, one might say that the polygraphy industry has taken lemons and learned to make lemonade.
in an artificial context (the courtroom) under questioning by an “expert” (the lawyer). So it is hardly distinguishable in form from polygraph evidence. The fact is that the court uses the human mind as a DDD, providing it with visual and paraverbal data, and this built-in “device” is generally inaccurate. Even individuals in law enforcement generally perform at levels barely above chance, and they presumably have more experience and incentive than juries or even judges in assessing statement credibility. Consequently, the whole process of making witnesses sweat on the stand, however integral it might be to the self-image of the bar, is highly dubious as an aid to the truth. Supposing the polygraph to have improved, but not much, the best that could be said of live examination over the last century is the same, meaning its relative value for adjudication has not really changed, and there is still plenty of demand for greater accuracy in this area. At some level, most courts realize this, or so one would hope, and this continued and unfilled demand explains their willingness to reconsider any apparent system whose improved ability might fill this need.

Moreover, and more importantly, from a biological standpoint, nonverbal demeanor evidence and polygraph measure are essentially metrics of the same phenomenon, a general level of sympathetic nervous system arousal loosely associated with an anticipated risk of detection or the transgression of norms (fear or embarrassment). The likelihood is that changes in blood pressure, sweating, and so forth have been used to assess credibility of accompanying verbal statements since the beginning of spoken language; likewise these markers have always been unreliable. What current research offers is what the polygraph


17. See Holly Orcutt et al., Detecting Deception in Children’s Testimony: Factfinders’ Abilities to Reach the Truth in Open Court and Closed Circuit Trials, 22 Law & Hum. Behav. 339, 365 (2001) (finding jurors incapable of telling when witnesses are dissembling). “I am led by my investigations to serious doubt concerning the ability of a trial jury to perform the central task assigned to them: to assess credibility. And, I must add, insofar as I can determine, the laws of evidence and the contribution of the trial courts in interpreting and applying the laws do little to enhance confidence.” H. Richard Uviller, Credence, Character, and the Rules of Evidence: Seeing Through the Liar’s Tale, 42 Duke L.J. 776, 778 (1993).

18. See Christian A. Meissner & Saul M. Kassin, “He’s Guilty!”: Investigator Bias in Judgments of Truth and Deception, 26 Law & Hum. Behav. 469, 472 (2002) (showing in a review of studies no effect of training, except increased likelihood of labeling all individuals as deceitful, yielding more Type II errors, along with increased false confidence in one’s abilities).

19. It is important to distinguish this from the questioning process used to identify inconsistencies or evasions; this is clearly valuable for credibility purposes, since by fleshing out the witness’s account, its actual level of relative plausibility is more easily assessed. What I question is the marginal value of conducting this questioning live, in order to achieve parallel transmission of non-verbal demeanor evidence that will inevitably affect credibility where that is at issue, despite the proven unreliability of such evidence.

20. A larger anthropological question is implicated by this point, regarding why our system of
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.

I. THE SCIENCE AND TECHNOLOGY OF DETECTING DECEPTION

A. THE METHOD OF COGNITIVE NEUROSCIENCE

The cognitive neuroscience way of looking at the mind is fundamentally driven by a desire to know how the brain produces behavior. Computer science has inspired cognitive neuroscience as it has emerged over the last quarter-century to ask "how a machine with the physical properties of the brain can produce specific behaviors when given specific inputs." What cognitive scientists mean by “specific” in this regard is usually defined in terms of an information-processing goal—or for inputs, a type of information—that can be specified computationally as a series of steps. For deception, no one has yet specified precisely what one must do to deceive, because there are many different kinds of deception, with different demands. What can be noted is what is not needed as an informational matter. Specifically, it is not necessary that a person “feel bad” or signal fear or otherwise engage the body in somatic arousal. Indeed this behavior is probably counterproductive for what might be taken as the working definition of the cognitive goal: to induce a belief in the receiver that the sender thinks plausible but false, a process that entails the receiver supposing a signal to be one emitted in order to convey information the sender believes to be true. Consequently, a cognitive neuroscientist would begin the study

verbal communication has apparently stabilized around imperfect deception, given that there are countervailing selective pressures for a signaler to deceive and for a receiver to detect deception. See generally Richard W. Byrne, Tracing the Evolutionary Path of Cognition, in THE SOCIAL BRAIN: EVOLUTION AND PATHOLOGY 43, 43-60 (M. Brune et al. eds., 2003) (inferring that the basic cognitive equipment necessary for tactical deception arose approximately twelve million years ago in primate evolution, but that language and consciousness, if not indeed driven by an “arms race” around deception, have significantly expanded the opportunities for it). Recent research suggests that in systems with repeated interactions, “bad liars” may be favored. See Paul W. Andrews, The Influence of Postreliance Detection on the Deceptive Efficacy of Dishonest Signals of Intent: Understanding Facial Clues to Deceit as the Outcome of Signaling Tradeoffs, 23 EVOLUTION & HUM. BEHAV. 103, 115 (2002). This result, which could maintain a mix of types genetically, is in line with a theoretical model that would assume “tells” (indicators of deception) are not selected against because it will be easier to forgive and trust someone after a revealed deception if their deceptions are accompanied by “tells”; hence such signals are a form of costly insurance advantageous to certain types. By contrast, if a lie is sent without such insurance, the “once-bitten, twice-shy” deceived receiver will not be able to categorize future transmissions and may simply mistrust the sender.


22. There are to be sure inadequacies with this definition—for instance whether merely concealing an expected signal counts, but these are beside the point for the moment. Note, however, that this definition does not imply that the receiver is denied “the truth.” The deceitful sender may also be mistaken about what is actually true, and yet may still act to mislead, under this definition, as
of deception not by looking at bodily arousal as is the case with polygraphy or the visual observations of stress reactions, but by examining "directly the organ that produces lies, the brain . . . identifying specific patterns of neural activation that underlie deception."  

Ideally, problem specification in cognitive neuroscience is to be done quite apart from, and prior to, actually looking at the physical properties and relationships of groups of neurons, which founding cognitive neuroscientists, such as Marr, referred to as the "level of implementation." 24 The implementation level corresponds conceptually to building a working piece of computing technology, and was contrasted by Marr with the computational level of "abstract problem analysis," which consisted of decomposing the problem into its primary parts (for example, the need to combine visual data from two eyes into a single image, and then to store this image). The intermediate level for Marr was that of the algorithm, giving a formal "programming" procedure that would in principle yield the right output for the inputs. 25 Once these levels had been penetrated the researcher would see how one would actually build this machine out of neurons within the brain (or rather, how evolution had potentially done so). In practice, however, these levels are not pursued independently, because the structure and organization of the brain acts not only to constrain theorizing, but to suggest more basically "what problems need to be solved," as well as giving hints about how the brain actually does solve them by processing information.

Therefore, an inquiry into detection of fabricated responses would begin quite differently from the polygraph, if this inquiry were informed by cognitive neuroscience. The first theoretical goal would be to define how to build a machine for lying, 26 not one for lie detection—because the structure of the latter must be wholly dependent on the structure of the

what he (the deceiver) thinks is true. Accord Augustine, De Mendacio, ¶ 3 (lying is the intentional negation of the subjective truth). This is close to the working definition used by biological researchers for tactical deception, which is said to involve a "successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue." Sean A. Spence et al., A Cognitive Neurobiological Account of Deception: Evidence from Functional Neuroimaging, 359 Phil. Transactions Royal Society B: Biological Sciences 1755, 1755 (2004) (quoting with approval the definition employed by A. Vrij & S. Mann, Telling and Detecting Lies in a High-Stake Situation: The Case of a Convicted Murderer, 15 Applied Cognitive Psychol. 187-203 (2001)).

23. G. Ganis et al., Neural Correlates of Different Types of Deception: An fMRI Investigation, 13 Cerebral Cortex 830, 830 (2003).

24. See Frontiers in Cognitive Neuroscience, supra note 21, at xiii.


26. I will allow the reader to insert his own witticisms regarding lawyers, politicians, or expert witnesses.
former. Only by understanding the nature of the device we possess for lying to other human beings could we hope to build another external machine that detects when this internal machine is operating. For precision greater than the detector we already possess, it is necessary to identify those distinct processes involved in implementing "the lying function," an identification that is greatly facilitated by knowing (1) what additional information or operations are needed to create a lie as opposed to uttering the truth, and (2) where in the brain information and operations of this type are performed.

Hence, a function-driven perspective is melded with a traditional approach in neuropsychology, which, starting from a "natural history" of the brain, sought to identify what various anatomical structures did and how they did it. One way to characterize the distinction is that cognitive neuroscience starts with the question of what tasks the brain must accomplish in order to carry out its functions, and then goes looking in the neuroanatomy for the mechanisms that actually accomplish these tasks. By contrast, the more traditional approach, which was often tied to the clinical examination of patients with particular injuries, would begin with the anatomical structure and go on to infer its function from its observed effect on behavior.27 Put more simply, the ultimate goal of cognitive neuroscience was to identify the neurological origin for every behavior that the brain performs, and the ultimate goal of neuropsychology was to find the associated behavioral function for every piece of the brain. These questions ultimately converge.

One of the most important techniques carried over from the naturalistic study of neuroanatomy is the identification of the dissociation and double dissociation of different cognitive capacities, which show the independence of different types of information processing.28 A dissociation between cognitive functions is demonstrated when the first function (for example, short term memory) disappears although a similar function (for example, long-term memory) remains. A double dissociation is more informative, because it more conclusively shows independence of the capacities. This would occur if there were also cases where long-term memory disappeared but the individual could recall material in the short term. For many years, such dissociations—describing the way cognition is "carved at the joints"—were painstakingly identified through the study and comparison of brain-

27. In this context, clinical neuropsychology would attempt to find what one might term a "clinical George Washington" someone "who could not tell a lie," literally, because they lacked the equipment to do so. (This is unlike the punctilious truth-teller, of whom it is more accurate to say that they would not tell a lie.) See generally Charles V. Ford, Lies! Lies! Lies! The Psychology of Deceit (1996) (noting that certain psychological deficits, such as autism, can create an excess of honesty).

damaged patients with peculiar deficits. To the extent function could be tied to structure, the nature of the patient's injury would provide the basis for any inference, even when it was difficult to tell exactly the role the damaged part played in the lost behavior (that is, the most that one could infer was that the anatomical structure was somehow necessary).

Neuroimaging allows one to pose a particular task and observe the parts of the brain that respond to it, and consequently has obvious advantages over the adventitious method formerly used to identify anatomical correlates, making it the most important methodological advance for cognitive neuroscience. Nevertheless, the logic of inquiry as sketched above, although greatly accelerated, remains the same: if a structure (usually a group of interlinked structures) is activated by one task, but not by other similar tasks, the tasks can be considered dissociated, and if the second task activates a distinct pattern or "signature," there is double dissociation and "independence"—although the more complex picture of the brain revealed by neuroimaging usually shows overlaps for those processing steps common to both tasks. As one commentator acknowledges:

With the development of functional imaging techniques capable of monitoring the brain's physiological response to cognitive tasks, researchers are rapidly gaining insight into the neural mechanisms that underlie vision, sensation, hearing, movement, language, learning, memory, and certain sex differences in language processing. Functional neuroimaging allows researchers to confirm long-standing hypotheses—first formulated from neuropsychological testing of brain-damaged patients—about structure-function relationships in the normal brain.

The result of this type of empirical work, together with the task-based orientation of cognitive neuroscience, led it to view the mind as essentially "modular." Without delving too deeply into the various debates over how modular, it is enough to simply note that there are "[f]unctional and/or anatomical components that are relatively specialized to process only certain kinds of information." The modular

29. COGNITIVE NEUROSCIENCE 136 (Michael S. Gazzaniga et al. eds., 2d ed. 2002) ("The most exciting methodological advances for cognitive neuroscience have been provided by new imaging techniques [that] ... enable researchers to identify brain regions that are activated during these tasks, and to test hypotheses about functional anatomy.").


31. See STEPHEN PINKER, THE BLANK SLATE 39-41 & passim (2002) (discussing different views on this point, as well as the broader implications of the modular view for the psychological and social sciences).

32. Ralph Adolphs, Cognitive Neuroscience of Human Social Behavior, 4 NATURE REV. NEUROSCIENCE 165, 166 (2003) (noting "evidence of domain-specific processing that is specialized for specific ecological categories (such as faces and social contract violations)").
view has been particularly conducive to evolutionary theorizing, because the "cognitive task" can be recharacterized as a relatively discrete adaptive problem to be solved by natural selection, the ultimate result of this being a specialized mental organ adapted to fulfill the information processing need of the organism. As applied to the study of deception, this does not necessarily indicate there exists a "module" for deception, since deceit is an ancient biological feature that is part of many different social behaviors and predates the evolution of language, although some would argue language led it to a new efflorescence of complexity. On the other hand, an ambitious detection of deception project would rely on the possibility that there may be features that are distinct to and common among the class of tasks we can identify as deceptive, even if those tasks show significant variation among them.

B. ELECTROENCEPHALOGRAMS AND THE "BRAIN FINGERPRINTING" METHOD

Before examining the application of neuroimaging to deception in more detail, it may be useful to distinguish this technique from an alternative method of monitoring mental activity, already being applied to some extent as a way of measuring deception. An event-related potential (ERP) is recorded by an electrode on the skull of a subject performing a task, and usually presented as a set of waveforms on an electroencephalogram (EEG). Because actual brain activity is electrical, this is a direct measure of the presence of information processing. However, the primary drawback is that the electrical activity impinging on the skull electrode is very difficult to localize to a particular piece of...
neuroanatomy within the "black box" of the brain.\textsuperscript{38}

As in neuroimaging, or polygraphy for that matter, the goal of the EEG deception researcher is to find a distinctive "signature" associated with a deceptive response. An EEG will not reveal precisely why and how the signature is created; in this sense it is much like a polygraph. However, consistent with cognitive neuroscience and unlike polygraphy, the brain-based measure used is more closely tied to the cognitive activity required for the deception to occur, as opposed to more contingent physiological correlates such as sweating or blood pressure. The type of deception that has attracted the most attention in this area is guilty knowledge. As general research into ERP and memory has demonstrated, there exist consistently different reactions of the brain in response to "significant" (i.e., remembered) information as opposed to non-significant information. Consequently, if a person was genuinely unfamiliar with a piece of evidence X (say a damaging memo or a murder weapon), he should treat it as no more "significant" than similar memos or weapons with which they are presented. This marker for the reaction is a particular waveform commonly known as the P300 wave (indicating its position next to the parietal lobe, which is involved in memory and recall).

One clinical application of this technique is the detection of "malingerers," people who pretend for a variety of reasons, including for fraudulent insurance or legal claims, to possess an illness they do not have.\textsuperscript{39} Psychologists attempt to spot those pretending to be brain damaged by giving such patients memory tests. Genuinely disabled people will of course not recognize a stimulus they have forgotten, while fakers will feign ignorance although the P300 wave of their brains will be consistent with the item being significant.\textsuperscript{39} Current approaches allow the classification of approximately eighty-seven percent of the deceptive subjects, with no misclassification of truthful subjects.\textsuperscript{40} A series of laboratory studies in which subjects were given incentives to lie about "guilty knowledge" of a simulated crime showed an ability to correctly classify ninety-six percent of the subjects as honest or deceptive.\textsuperscript{41}

The forensic use of this technique has acquired the sobriquet of "brain fingerprinting,"\textsuperscript{42} although its primary promotor, Dr. Lawrence

\textsuperscript{38} The use of several different electrodes is common, providing some capacity to "triangulate" through differential levels of activity what broad region of the brain is being used.

\textsuperscript{39} See Mossman, supra note 3, at 231 (discussing case law involving possible feigning of cognitive deficits to avoid standing trial and for other reasons).

\textsuperscript{40} See J. Peter Rosenfeld et al., P300 Scalp Amplitude Distribution as an Index of Deception in a Simulated Cognitive Deficit Model, 33 INT’L J. PSYCHOPHYSIOLOGY 3, 17 (1999).

\textsuperscript{41} See id. at 16.


\textsuperscript{43} See 3 DAVID L. FAIGMAN, SCIENCE IN THE LAW § 10.3-3.13[2] (2d ed. 2002); Andre A.
Farwell, refers to it as the Farwell MERMER technique. In any event, Farwell has attempted to promote this as a replacement for the polygraph, but with limited success. Farwell's attempt has received a fair amount of media interest, particularly after an Iowa court admitted the technique as competent evidence in 2001. More recently, Farwell has been less successful in introducing this material in Oklahoma, after his examination of a capital defendant there.

Despite the hype, the amount of peer-reviewed material available to study the efficacy of this method is almost nonexistent; it turns out that much of the basic research was funded and or conducted by the CIA in the late 1980s and early 1990s and therefore is publicly unavailable. The CIA terminated its research after Farwell refused to reveal the algorithm used to analyze the EEG readings, on the basis that he considered it proprietary. He subsequently appears to have received two patents and to have gone into the private business of detecting deception, a process that it seems has hampered the ability of the scientific community to either validate or build upon his research.

Moenssens, Brain Fingerprinting—Can It Be Used to Detect the Innocence of Persons Charged With a Crime?, 70 UMKC L. Rev. 891, 893–95 (2002).

44. Lawrence A. Farwell & Sharon S. Smith, Using Brain MERMER Testing to Detect Knowledge Despite Efforts to Conceal, 46 J. Forensic Sci. 1, 1–9 (2001) (MERMER is Farwell's acronym for the particular waveforms he uses, one of which is derived from P300).


46. See Slaughter v. State, 105 P.3d 832, 834 (Okla. Crim. App. 2005). “Dr. Farwell allegedly asked numerous details concerning ‘salient details of the crime scene that, according to [Petitioner's] attorneys and the records in the case,... the perpetrator experienced in the course of committing the crime for which Mr. Slaughter was convicted.’ According to Dr. Farwell, Petitioner's brain response to that information indicated ‘information absent.’” Id. In part because Farwell failed to provide the Oklahoma court with a promised report on his work, including the salient details tested or peer validation of the technique, the court rejected this material as “new evidence” and concluded that “Brain Fingerprinting, based solely on the MERMER effect, would [not] survive a Daubert analysis.” Id. at 836.

47. United States General Accounting Office, Investigative Techniques: Federal Agency Views on the Potential Application of “Brain Fingerprinting” 9, 14 n.1. (2001) [hereinafter GAO Report]. Apparently Farwell conducted a forty subject study in which one-half (twenty) of the subjects had participated in a simulation, and one-half had not. All subjects were then presented with pictorial stimuli—presumably of the simulation—and their EEGs were compared. Farwell claimed one hundred percent classification had occurred in that study, which used only the P300 measure. See Farwell & Smith, supra note 44, at 137.

48. See GAO Report, supra note 47, at 7 (indicating his acquisition of two patents in 1995).

49. If the history of the polygraph is a guide, commercialization and proprietary control of a scientific technique is unlikely to improve the technique, at least if this occurs before the background science is well understood and accepted (which would then allow researchers to compete on relative accuracy). In an atmosphere of bitter conflict among the early developers of the polygraph, one of
Nevertheless, other ERP researchers are able to replicate what Farwell is able to do (or claims to be able to do) in terms of detecting when subjects deny the significance of significant material. Moreover, there is some theoretical basis for explaining the electrical activity of the P300 wave as what is termed an "orienting response" toward surprising or important information, a response that can be detected by either EEG or by polygraph, or by a combination of the two. As argued recently by proponents of using this method forensically, responses to "guilty knowledge" mediated by an orienting response are not in fact premised on the type of fear arousal usually associated with measurement made by the polygraph, but are wholly cognitive, based on the fact that the organism will focus attention on any item of importance, independent of how they might emotionally react to this item. The main problem with the use of the physiological correlates of this response is that they track the stimulus by periods of up to twenty-five seconds, making them vulnerable to countermeasures. Therefore, the authors fall back ultimately on the contention that the P300 response would be a suitable measure, as it appears and disappears within about two seconds after the stimulus.

Recent reports of experiments conducted in Beijing at the National Laboratory on Machine Perception support the promise of this technique in using event-related potentials (ERP) measurements, and is of particular relevance for forensic work. The experimental subjects (hooked to thirty different electrodes) were presented with nine faces, three of whom were strange, and six of whom were familiar, and were told to deny familiarity with three of the known faces (the "targets"). The subjects were told they would be penalized if the "computer" caught them lying, and regardless of the actual ERP of the subjects, the computer pretended to "catch them" five times out of the thirty times

---

them, Leonard Keeler, patented his own version, and when the patent expired, began a commercial training service for examiners. Although this produced a standardized, mass-produced machine, Keeler's training methods were not oriented to scientific validation but rather to the desires of his customers (employers and police departments) to maximize the chance of subject confession. Ken Alder, A Social History of Untruth: Lie Detection and Trust in Twentieth-Century America, 80 REPRESENTATIONS 1, 22 (2002). Moreover, it is probable that the limited paraprofessional training of polygraphers on Keeler's machine (or its descendants) would have the potential to create a network effect that would make them resistant to hardware changes away from the design on which they were trained.

50. Gershon Ben-Shakhar et al., Trial by Polygraph: Reconsidering the Use of the Guilty Knowledge Technique in Court, 26 LAW & HUM. BEHAV. 527, 530 (2002).

51. Id. at 532. Any somatic polygraph measures would also remain susceptible to the criticism that stray thoughts or other stimuli could induce arousal that would be very difficult to distinguish from that correlated with cognitive activity, whether or not this arousal was intentional.

52. Id.

they responded (a feature apparently used to maintain the subject's fear).

This particular protocol reveals a serious problem with the use of P300 measure of "significance" in deception research. As Dr. Fang astutely discusses, with regard to stimuli like faces, an orienting response might occur to a familiar face because that would be something of interest. However, it might be equally expected to occur when exposed to the unfamiliar face of a stranger, thus making the equation of "significance" with "knowledge" extremely problematic. Indeed the Chinese researchers found that familiar and strange faces were not distinguishable based only on the "orienting response." Fang and his colleagues were easily and reliably able to use statistical techniques to distinguish when a "target" face had been presented and the subjects had denied knowing this person by combining the electrical activity of all thirty electrodes. Other electrical measures showed a distinguishing pattern, and this complex statistical signature—rather than the "orienting response"—served as the functional and neural basis for the deception marker. Because of the experimental design, fear or indeed anticipation of detection might explain much of the electrical signature, along with a greater amount of cognitive activity devoted to deception related efforts such as self-control. Consequently, although the Fang method may be relatively useful in detecting deception as to whether one knows another person, any ability to make inferences between brain and behavior have been sacrificed in order to achieve the clearest marker of deceit.

54. Id. at 249.
55. Id. at 252–53.
56. The subjects were instructed specifically to "keep their minds calm" in order to avoid detection. See id. at 249. Since any cognitive activity directed to this end would increase electrical activity, following this advice would actually contribute to the observed effect. Although this confuses the scientific interpretation of the result, it would be useful in practical terms because it suggests that many polygraph countermeasures that could control items like blood pressure still require cognitive effort and so would be much less effective against ERP detection.
57. The feigned ignorance paradigm is difficult to apply in most criminal investigations or civil trials, particularly given the discovery and disclosure rules governing these proceedings. Nevertheless, its applicability is perhaps wider than realized. In the United States there is a relatively small amount of research into what one might call "forbidden knowledge." For a distillation of commonly held norms, see NICHOLAS RESCHER, FORBIDDEN KNOWLEDGE AND OTHER ESSAYS ON THE PHILOSOPHY OF COGNITION 9 (1987) ("There seems to be no knowledge whose possession is morally inappropriate per se. Here inappropriateness lies only in the mode of acquisition or in the prospect of misuse. With information, possession in and of itself—indeed independently of the matter of its acquisition and utilization—cannot involve moral impropriety.") In other words, it would incriminate a subject if he were shown to be familiar with bomb-making equipment and procedures, given that he was under suspicion of misusing this knowledge, and familiarity with classified or proprietary information is punishable as probative of complicity in its unlawful transfer. This criminalization of knowledge can be much more expansive in an authoritarian political system, making the use of knowledge detection technology potentially pervasive. Especially when combined with a lack of concern about false positives, it would be relatively straightforward to obtain results indicating a subject's mere recognition of forbidden writers and passages of their works, banned religious rituals or symbols,
It is unlikely that such an atheoretical approach to understanding deception will be satisfying scientifically or legally. Despite the claims of those using ERPs to detect "information" stored in the brains of the subjects, in fact only a responsive orientation reaction is usually detected, and this is correlated only in a rather loose way with whether the information has already been encoded. Without observing the actual processes of memory encoding and memory retrieval, as they pertain to a stimulus (i.e., whether the former or the latter occurs), it remains speculative whether or not the stimulus was present in the brain prior to its presentation. More fundamentally, even if in a particular context a global EEG signature can tell true from false responses, without understanding the basis of this difference, it becomes difficult to know to what extent the result can be generalized, or to what extent other behaviors (outside the limited range available in a laboratory experiment) could produce the same result. Logically, the only way to doubly dissociate deception is to observe and understand the anatomical correlates that are required to implement the deceptive act—the extra processing steps necessary to deceive as opposed to giving a similar but honest behavioral response. Then, \textit{a priori}, we could ask if in the particular context in which credibility is assessed, those functions would be called upon by a deceptive response. If so, we could, in effect, watch the lie as it was being constructed.

C. \textbf{FUNCTIONAL NEUROIMAGING, ESPECIALLY USING MAGNETIC RESONANCE}

Having now argued that for purposes of assuring scientific validity, detection of deception will require functional neuroimaging, at least in the medium term, it then remains to describe more fully the limits and capacities of these techniques as they are currently practiced. The main methods are positron-emission tomography (PET) and magnetic resonance imaging (MRI), but I will concentrate on the latter for several outlawed works of art or media broadcasts, or even suppressed or "enemy" languages. Clearly such interrogation about the contents of one's mind would (among other things) violate the freedom of thought protection at the core of the First Amendment. See Vaughn v. Lawrenceburg Power Sys., 269 F.3d 703, 717–18 (6th Cir. 2001). For an initial assessment of magnetic resonance imaging in this light, see Sean Kevin Thompson, Note, \textit{The Legality of the Use of Psychiatric Neuroimaging in Intelligence Interrogation}, 90 CORNELL L. REV. 1601, 1615–18 (2005).

58. One efficient way to solve the specificity problems associated with ERPs would be to simultaneously record an EEG trace while the subject is undergoing a neural imaging scan. If in fact the neuroimaging is able to specify a distinctive activation pattern for deceit, it would then be possible to use it as a standard against which one could measure patterns of electrical activity on the skull. Then if one or more of these patterns possesses a precise correlation with actual brain states during deceptive activity, it would be possible to dispense with the MRI equipment in later studies that were essentially replicating the circumstances where the EEG-MRI relationship had been established. Since none of the preconditions for this shortcut have yet been fulfilled, I consider it useful to concentrate on identifying the brain-behavior relationship using MRI.
reasons: (1) although the use of PET predates the use of MRI, the hardware involved in the latter is far less expensive and now much more common, (2) the temporal and spatial resolution of MRI is superior, allowing easier correlation of short-duration events with specific regions in the brain, and (3) most determinative for legal use, PET involves, among other things, the injection of radioactive tracers into the blood stream, which for clinical use strictly limits the amount of PET scanning one can do on an individual, and probably bars it entirely from use as a quasi-deposition technique that any United States court would countenance.59

MRI scanning, as done in most major hospitals, is non-invasive; it uses a very strong magnet and directs radio waves at the subject's body. For neural imaging by MRI the subject lies on a table, with his head surrounded by a large magnet. The magnet causes some of the protons within the atoms inside the subject's brain to align with the magnetic field. A pulse of radio waves is then directed at the patient's head and some of it is absorbed by the protons, knocking them out of alignment. The protons, however, gradually realign themselves, emitting radio waves as they do. These radio waves are captured by a radio receiver and are sent to a computer, which constructs the brain image. The patient cannot sense either the magnet or the radio waves; in fact, the patient only knows the machine is working because of the noise it makes during scanning. Different parts of the brain respond to the radio waves differently, and emit slightly different radio signals depending, among other things, on the local water and fat content. This provides a picture of brain structure.

In order to use an MRI scanner to view the functioning brain as it responds to a task, the following important but not heroic assumptions need to be noted. First, there is the general cognitive neuroscience premise that particular types of cognitive activity can be reliably associated with specific areas of the brain, and that some of those areas in fact provide the neurological machinery for carrying out the cognitive task. As indicated by the caveat, it is sometimes difficult to distinguish between parts of the brain whose stimulation is correlated with a cognitive event and those that actually process it. For instance, speaking involves, among other things, hearing one's own words, so the auditory portions of the brain are normally active during speech production. These areas are not, however, necessary for speech production, since the hearing impaired (either clinically or experimentally induced) can in fact accomplish this task, albeit with less fluency. Second, it assumes that active portions of the brain during the task can be identified by an increased amount of blood flow to those areas during the task. In order

to operate, brain tissue requires oxygen carried by blood hemoglobin. If an area has increased demands for oxygen, the circulatory system responds by increasing blood flow (and thus hemoglobin and oxygen) to the area. What fMRI actually measures is the ratio of oxygenated hemoglobin to deoxygenated hemoglobin, $O+/O-$, which is possible because the deoxygenated hemoglobin responds more readily to the magnetic field emitted by the machine.

Obviously, there is a small time delay between any event calling upon greater cognitive resources, the response of neurons in some region with increased electrical activity, and then what is actually measured, a circulatory adjustment coupled to the increased cellular demands in that area of the brain. An event can be identified as having occurred within about a two second window. What studies have shown about the circulatory, or hemodynamic, response to neural activity is the following: in the first couple of seconds of processing, the activated brain region uses the oxygen at hand, and the $O+/O-$ ratio falls from the baseline state; after that, blood flow increases and in fact “overcompensates” by providing more fresh oxygenated blood than the tissue can absorb, and hence the $O+/O-$ ratio increases and stays above baseline while the region is actively working. In actual physical terms, the hemoglobin in the area under observation becomes less sensitive to the magnetic field of the fMRI. (This can be confusing because most pictures of an active area represent it as glowing colorfully—while the machine actually is measuring activity by the lowered emission of energy in an active area now flooded with less sensitive molecules). The key measure is usually reported as the blood oxygenation level dependent (BOLD) effect, measured by subtracting the difference between the background activation shown in the area prior to the task from the level of activation during the task. That is, where the researcher has shown a BOLD effect in an area, that area is implicated by the stimulus or task.

D. NEUROIMAGING OF DECEPTION

Because much of the previous discussion has focused on the attempt to discover whether a subject is familiar with a particular piece of information, it is worth noting that preliminary studies directly investigated whether a subject can falsely deny such familiarity. However, these direct investigations, which I will refer to for

60. Luis Hernandez et al., Temporal Sensitivity of Event-Related fMRI, 17 NEUROIMAGE 1018, 1025 (2002). The brain should return to a baseline state within about twenty seconds, which is therefore the spacing required between stimuli whose effect is to be separately assessed (as for instance in a series of deposition questions whose answers are to be monitored).

61. COGNITIVE NEUROSCIENCE, supra note 29, at 94. This is almost certainly because the key demand during brain activity is for the blood to remove the by-products of metabolic activity rather than to supply energy and oxygen.
convenience as the Langleben (2002) Study and the Lee (2002) Study, do not exhaust the relevant neuropsychological research in this area. The encoding of memory, one part of which seeks to distinguish processing of novel and familiar stimuli, has been a central focus of cognitive neuroscience from its start, and has continued during the current era of functional neuroimaging studies. This research involves comparisons of stimuli seen for the first time and stimuli that have become familiar through previous presentation. Novelty-driven activation in various parts of the brain has been repeatedly demonstrated for scenes, words, object-noun pairs, word pairs, and line drawings. Similar results have been obtained for faces: when a novel face is encountered, the left prefrontal cortex is activated along with the hippocampus (which is involved in many memory storage operations); whereas when the face is recognized later, completely distinct regions of activation are seen in the prefrontal cortex but not the hippocampus. In fact, recent fMRI studies have been able to distinguish among new items, familiar items from a memorized list, and words the subject falsely believes to have been on the list.

Consequently, it seems highly likely that any form of malingering with regard to specific information should be detectable by observation of brain responses, because even good-faith mistaken responses in this area are detectable. While memory impairment is perhaps too simple to debunk, malingering more generally is an active and increasingly successful area of research, particularly with regard to testing claims of pain or paralysis. In subjects actually incapable of motion, areas of the brain responsible for initiation of motor activity are silent, while fakers show activation in the prefrontal cortex, apparently where they issue commands to inhibit the motion by conscious effort. Quite apart from any use this application might ultimately have in assessing credibility by

67. *See id*. at 340–41. The false items are “implanted” by having the subject memorize a series of words such as “thread, pin, eye, sewing, sharp, point, haystack, pain, injection.” Most subjects will later be convinced that they saw “needle” on this list, but an fMRI can show a different pattern despite this subjective belief. *Id*.
68. David Oakley et al., *Differential Brain Activations for Malingered and Subjectively “Real” Paralysis*, in *Malingering and Illness Deception* 280 (Peter Halligan et al. eds., 2003).
measurement of familiarity with relevant details, neurological examination by fMRI is of obvious forensic value, most clearly in cases such as insurance fraud.

1. The Langleben (2002) Study

In the Langleben (2002) Study, participants were given an envelope containing twenty dollars and the five of clubs, and were told they would be allowed to keep the money if they could keep the identity of the card a secret from “the computer.”69 The computer displayed pictures of different playing cards on a screen inside the MRI machine, and asked if the subjects had the card, to which the subjects responded by pressing a “yes” or “no” button in their hand. (In fact, they always responded “no”). Significant differences were identified in a variety of areas between the pattern of brain activity when they truthfully denied having other cards and when they falsely denied having the five of clubs.70 However, no attempt was made to examine the subjects on an individual basis to classify them as truthful or deceptive. It was suggested that areas showing greater activation indicate that lying required greater activity in areas responsible for “executive” functions or decision-making, as well as in premotor areas required to suppress or inhibit forelimb movement (presumably to press “yes”).

Phan and his colleagues recently replicated a number of the findings of the Langleben (2002) Study, using a purposefully similar structure of having people deny possession of playing cards (although they did not reward participants for successful deception).71 Several areas of the brain were more active when lying—none were more active from the baseline condition when telling the truth.72 The Phan replication, which used fourteen subjects, did find one area of the prefrontal cortex activated by

69. As the study notes, one difference between this experiment and actual deception circumstances is the lie is in some sense permitted. See Langleben et al., supra note 62, at 728. However, this could also be true in certain conditions where the individual believed he was acting morally for ideological reasons, was self-deceived about the truth status of the statement, or was essentially amoral. In an individual who believed he was violating a social norm, however, any activated regions for deception would be accompanied by those cognitive areas responsible for what is behaviorally termed a “conscience” or “morals.” Although it would be beyond the scope of this Article, the identification of this neural area is of legal interest, since under certain theories of responsibility the activation of this region is considered to enhance “responsibility” leading to an increase in punishment. Somewhat paradoxically, the failure to activate the region during proscribed conduct is often considered, with justification, to bespeak caution regarding the future propensity of the subject to commit the conduct. Rather than extend this into the suggestion of how a full neurological specification of conduct may clarify philosophical concepts, and thereby affect concepts of criminal and civil liability, I note only that it is likely to do so.

70. See Langleben et al., supra note 62, at 730 (the areas showing greater activation were Brodmann areas 1-4, 6, 8, 24, 32, 40).


72. See id. (the areas showing greater activation were Brodmann areas 8, 9, 21, 22, 37, 40, 45, and 47).
all individuals when denying possession of the card, which they suggested therefore might serve as a “neural signature for the generation of lies.” Aside from the inconsistent activation of this region in other studies, a confound of this study was the attempt of the researchers to simulate “conditions of the polygraph” in order to make subjects feel “anxious”; thereby the fMRI results are tied to emotional responses (which as discussed above, are inherently imprecise), rather than to the cognitive challenges specific to deception.73

2. The Lee (2002) Study

The Lee (2002) Study was a feigned memory impairment test, in which subjects were told to intentionally do poorly on a test of recall. They responded with a yes/no button. Subjects demonstrated activation of areas responsible for memory recall, as well as for planning and calculation (because there were only two responses, the test paradigm forced the subject to keep track of how many wrong answers he gave in order to approximate chance). They also showed an apparent inhibition of the relevant motor area associated with responding through use of their forelimb.74 As in the Langleben (2002) Study, it appeared there was an instinctive movement or “pre-movement” toward the true answer, which was suppressed. No individual variation or classification was done.

As with earlier technology exploring “guilty” knowledge, there is the inherent problem of a subject possessing familiarity in an “innocent” fashion. For instance, in the paradigmatic case with a presentation of several firearms, one of which is a murder weapon, the subject may turn out to show greater recognition of the weapon of interest, simply because it is the most common of the set presented, or because the subject himself possesses a weapon of that generic type.75 This complicates the interpretation of such evidence, and weakens the strength of the conclusion that can be drawn from a “test-positive” result, but any error would be an inferential one, or arise through lack of proper rebuttal, and would not be intrinsic to the method. Moreover, fMRI memory research may be able to partially vitiate this problem to the extent it can detect the context in which a subject became familiar with an item, in addition to its mere familiarity. Thus, it is theoretically possible to distinguish between someone who has become familiar with a person or item by being shown it or told about it, and one who saw the person or item at a

73. See id. Although presumably certain emotional responses (and their reflections in neural activity) do correlate more strongly with lying than veracity, the individual variation in this, as reflected in the history of the polygraph, could easily create similar difficulties of interpretation and implementation if transferred to a neuroimaging context. Therefore purposely introducing anxiety and looking for it seems a somewhat questionable strategy for improving accuracy in deception detection.

74. See Lee, supra note 63, at 162 (the areas showing greater activation were Brodmann areas 6, 9, 10, 21, 23, 40, 46).

75. For more on this problem, see Ben-Shakhar et al., supra note 50.
particular place and time.\textsuperscript{76} Were DDD evidence to begin to be employed to validate memory claims, and counter-strategies of "inoculation" by familiarization developed, context-retrieval would presumably become of increasing importance, although it does not appear to have yet been adapted for deception research.

The use of fMRI in the area of memory would have relatively little impact on the legal process, at least in comparison with the use of neuroimaging that would genuinely function as a "lie detector." For this to occur, the technology would need to go beyond determining the presence or absence of bits of information and actually distinguish confabulated statements from "true" statements. In this context, "truth" means more specifically a verbal expression of what the speaker actually recalls from memory when prompted by a stimulus question, and "a lie" means an alternative statement that the speaker has creatively generated to take the place of the true expression facilitated spontaneously within the brain by the stimulus. This area, the key one for assuring the veracity of testimony, has now been the subject of four studies, all of which have become available only within the last few years, and all of which purport to lay the groundwork for the use of neuroimaging as a lie detector. I will refer to these for convenience as the Kozel (2004) Study,\textsuperscript{77} the Ganis (2003) Study,\textsuperscript{78} the Spence (2001) Study,\textsuperscript{79} and most recently, the Nunez (2005) Study.\textsuperscript{80} These might be termed the first-generation of investigations. At this writing, there are three additional research reports, all currently in press, tying deception to brain patterns through the use of fMRI. These investigations build on the initial studies, and bring neuroimaging significantly closer to practical use as a DDD.\textsuperscript{81}

\textsuperscript{76} See Ken Paller et al., \textit{Neural Correlates of Person Recognition}, \textit{10 Learning \\& Memory} 256 (2003) (showing specific additional networks activated when individuals retrieve "the spatiotemporal context of the initial episode of viewing the face"); Scott M. Hayes et al., \textit{An fMRI Study of Episodic Memory: Retrieval of Object, Spatial, and Temporal Information}, \textit{118 Behavioral Neuroscience} 885, 885-96 (2004) (describing detectable differences when subjects, after viewing a videotaped tour of four houses filled with memorized items, were presented with objects in their original setting, versus isolated or alternative settings).

\textsuperscript{77} F. Andrew Kozel et al., \textit{A Pilot Study of Functional Magnetic Resonance Imaging Brain Correlates of Deception in Healthy Young Men}, \textit{16 Neuropsychiatry \\& Clinical Neurosciences} 295, 295-305 (2004).

\textsuperscript{78} Ganis et al., supra note 23, at 830.

\textsuperscript{79} Sean A. Spence et al., \textit{Behavioural and Functional Anatomical Correlates of Deception in Humans}, \textit{12 NeuroReport} 2849, 2850 (2001).


\textsuperscript{81} Two of the studies arise from the University of Pennsylvania group headed by Langleben, and continue in the "feigned ignorance" paradigm employed by this group. See Ben-Shakhar et al., supra note 50, at 532; C. Davatzikos et al., \textit{Classifying Spatial Patterns of Brain Activity with Machine Learning Methods: Application to Lie Detection}, \textit{28 NeuroImage} 693 (2005) (using computer analysis on the results of the Langelben group's scans); Daniel D. Langelben et al., \textit{Telling Truth From Lie in Individual Subjects with Fast Event-Related fMRI}, \textit{26 Human Brain Mapping} 262 (2005); supra text
CROSS-EXAMINING THE BRAIN


The Kozel research program has so far been the one most closely tied to a simulated possible crime. In the primary study, the subjects were taken to two rooms, Room 1 and Room 2; in each of the rooms were five objects; beneath one of these five objects was a fifty dollar bill. The subjects “discovered” the bills, but did not disturb them. In the MRI, they were then shown pictures of the objects and asked if the bill was hidden under them. This differs from the familiarity-based studies reported above, because the subjects were “familiar” with all the objects displayed. They responded “yes” by holding up one finger, “no” by holding up two. As a control, they were told to respond truthfully as to the objects in Room 1 and they would get the fifty dollars there. As to Room 2, however, they were instructed to lie twice, denying the correct hiding place, and then picking one of the four object locations and affirmatively claiming that was the hiding place. If they successfully deceived, they were told they would get the other fifty dollars.

As a group, the subjects showed greater activation in the “executive” function area of the anterior cingulate cortex, as well as other areas, when they were responding with a lie as opposed to when they were giving truthful responses. Interestingly, however, in this study, each of the eight subjects was individually assessed. One subject showed no greater activation with regard to lying and truth-telling, and although the other seven did, their activations were in quite heterogeneous parts of the brain. Moreover, there was no area of the brain that was differentially activated in all seven subjects so as to provide a “marker” for deception.

In a subsequent replication of this research, using ten subjects, aggregating the subjects once again allowed a statistical discrimination of the lying brain, and showed—as a group phenomenon—similar activation in areas such as the anterior cingulate cortex and prefrontal areas. In this replication, the researchers also managed to get somewhat more consistent individual results, in that seven of the ten subjects showed greater activation when lying in the same general area of the brain. What remained to be shown was that each of these individuals, if retested, would show the same pattern, thereby allowing their responses to be matched against a standard calibrated to that particular individual.

accompanying notes 69–70. In addition there has been a further research development by the Medical University of South Carolina group headed by Kozel. See F. Andrew Kozel et al., Detecting Deception Using Functional Magnetic Resonance Imaging, 58 BIOLOGICAL PSYCHIATRY 605 (2005). See also supra text accompanying notes 79–81. This program has been more focused on clinically replicating the circumstances of a crime accompanied by a false statement of fact.

82. See Kozel et al., supra note 77.

83. Frank Andrew Kozel et al., A Replication Study of the Neural Correlates of Deception, 118 BEHAVIORAL NEUROSCIENCE 852, 855 (2004).
As the researchers acknowledge, but have not yet addressed, interscan variability even with the same subjects can be high, and this would be extremely problematic if there is already sufficient inter-individual variability such that reliable research requires an individual profile of deceptive and truth brain states.

One of the ten subjects, for example, showed clear differences but seemed to activate a different area of the brain, and as in the earlier study, some subjects (here two out of ten) failed to show greater activation when lying than when responding truthfully. Although this suggests that some individuals could be assessed for deceptive responses, it also carries the corollary that the process of deception could be sufficiently variable among subjects that a certain percentage of the population could essentially be immune from this type of examination. From the legal perspective, this would not pose a significant problem so long as this group could be identified and if it were difficult or impossible for a person to make themselves immune through countermeasures.


In the Ganis (2003) Study, the subjects were interviewed about their work and vacation experiences. They were then asked to respond with short verbal answers about where these took place. Sometimes the subjects were asked to respond truthfully, sometimes they were asked to respond with a prepared falsehood, and for some questions they were asked to make up a spontaneous lie (although the lie was supposed to be a plausible one—i.e., they were not to say they had spent their last vacation “on Mars” or with a nonsense response like “purple bookmark”). The spontaneous lies showed the greatest number of other activated areas, and the prepared fabrication showed greater activation in areas associated with the retrieval of episodic memory. No individual investigation or classification of subjects was done, but the researchers were well aware of this issue, stating that “whether fMRI can become a useful tool for the detection of deception...depends on whether reliable neural signatures of deception can be identified in single participants and in single trials.”

5. The Spence (2001) Study

Subjects in the Spence (2001) Study completed a questionnaire before entering the MRI as to whether or not they had done certain common activities that day. They then essentially “retook” the questionnaire twice inside the MRI, one being asked by using an auditory cue and once by visual query. At random, they were cued to lie

---

84. See id.
85. See Ganis et al., supra note 23, at 833 (areas showing greater activation in both deceit paradigms: 7, 10, 36, and 37).
86. Id. at 835.
on some of these responses. The subjects responded by pressing yes or no on a keypad. These results were broadly consistent across the two types of questioning, showing activation in areas responsible for the inhibition of motor and other responses. 87 This was consistent with a short comparative delay in responding when the individuals were required to lie. Individual subjects and possible heterogeneity were not examined.

6. The Nunez (2005) Study

Broadly speaking, the Nunez (2005) Study was similar in format to the Spence (2001) Study, in having the subjects initially fill out a “yes/no” questionnaire, and then redoing the questionnaire while inside the MRI. 88 In this research, however, the subjects actually did the questionnaire twice while in the scanner, answering truthfully to all questions during the first scan and falsely to all questions during the second scan. The results were generally consistent with the research considered above by finding “there is increased neural activity within the anterior cingulate [and] dorsolateral prefrontal cortex . . . when individuals answer falsely as compared to truthfully.” 89 Again, this was a generalization, although individual results were not reported except to note that reaction times were longer for all subjects when answering falsely, it is apparent there was considerable variability in the nature of individual brain activation. Interestingly, some of this variability was found to be correlated with the results of a personality profile the subjects took. This suggests that much of the variability among people is part of the emotional “noise” that is associated with lying, and is reflected in activation of different areas of the brain. The well-known fact that people vary in their emotionality about falsehood—the very reason that measures of emotionality such as the polygraph are so limiting—is therefore implicated in interpreting the fMRI in order to isolate those cognitively necessary components of falsification that are invariant (or at least less variant).

7. Current Research

Second-generation research appears to be focused on overcoming the problem of individual variation by developing a computer-assisted classification of a set of markers (only some of which will apply in any

87. See Spence et al., supra note 79, at 2851 (areas showing greater activation in both deceit paradigms: 6, 8, 32, 40, and 47). Spence and his collaborators discuss their research paradigm and results in Spence et al., supra note 22, at 1757–58, and suggest that the brain activation appears to be less related to the “work” required to generate a false statement, than with the “work” required to suppress or shunt the true statement away from its normal translation into a verbal response, activation of the speech centers, and utterance by the motor pathways, even though the response in their research was in fact non-verbal.

88. See Nunez et al., supra note 80, at 271 (areas showing greater activation: 6–10, 32, 44–47).

89. Id. at 276.
one case) as indicative of deception. Subjects are motivated by a reward if they are successful in deceiving. This classification system is then validated on a subsequent or separate sample to see whether or not it can tell when these people are lying. That is, the data developed in the initial phase of the study provides a guide for a prediction in the next phase of the study. As mentioned, developing this sort of "litmus test" that can be applied to a future novel subject is a necessary precondition for the use of neuroimaging to evaluate testimony for legal evidence. The primary areas of brain activation that go into constructing the models are broadly similar to earlier studies.90

Langleben's model using the clusters of the brain associated with lying or telling the truth, produced an ability to correctly classify 153/200 of the responses of four novel subjects, who repeatedly denied or admitted possession of a "guilty" item. Obviously, this is far better than chance, but also far from perfect. It produced a Type I error of false "accusations" of lying 17/104 times or 16%, and it missed 33/96 lies told, leading to a Type II error of 31%. When sophisticated machine-learning methods were used to develop and weight the predictors in this study, accuracy was considerably improved: when the computer developed a model on 21 of the 22 original participants and used it to analyzed the remaining person (doing this 22 times), the Type I error remained approximately the same (14%), but the number of lies caught rose to 90% (a Type II error of 10%).91

In the extension of Kozel's research, the individual subjects all "stole" either a watch or a ring and then denied stealing each one, thus uttering a truthful denial in one instance and a false denial in another instance. The purpose of the scanning was to determine which item they stole, which necessarily entailed an ability to discriminate between false denials and true denials. Kozel and his collaborators prototyped the system on thirty subjects, and then, using the predictors derived from this phase, attempted to tell which item had been stolen by a subsequent-group of thirty different subjects. Although wide individual differences in the pattern and extent of activation were present, all thirty subjects

90. See generally Langleben et. al., supra note 81; Kozel et al., supra note 81. Langleben, who uses denied knowledge, unlike other studies, finds areas of the brain activated by truth-telling over a baseline condition. One possible reason for this is that truth-telling in this context involves the "confession" of a relevant, ostensibly guilty fact, and therefore can logically be supposed to involve more decision making and anxiety than normal truthful responses where the truth is not made problematic for the subject.

91. See Davatzikos et al., supra note 81, at 665. Effectively, this meant that the left-out participant was "novel" to the model built to predict their responses. It is notable, however, that since this individual was in actuality part of the same experimental set-up, some sources of variation were absent that presumably would have lowered the accuracy rate if a truly novel individual had been tested and evaluated using the predictors derived from a wholly separate study and study population of twenty-one people.
showed at least some differential activation while lying in one or more of three anatomical clusters. When the relative activation of all three clusters was combined to assess thirty-one novel subjects, it was able to correctly predict which item had been "stolen" by them in twenty-eight cases, or slightly above ninety percent accuracy.

Unless this technology is improved, we could not, however, expect equivalent performance in a forensic setting, because in this clinical test, the system "knows" the individual is lying—it simply has to guess which of two statements is most likely to be the lie, and it is right ninety percent of the time. These results are promising, but picking out whether there is a lie and identifying the lie or lies within a long series of testimonial statements represents different circumstances requiring the use of DDD. Presumably, the Kozel system could tell which of an individual's statements had the least neurobiological indicia of veracity, but this would only be a measure relative to the person's other statements rather than to an objective measure of truth and falsity. The research assumes away the most important element as far as court evidence is concerned: whether or not the testimony in fact contains a lie.

8. Summary

Although the specific areas of the brain noted as indicative of a lie differed substantially, there was some agreement among all investigations to date that: (1) some form of "executive" function that deals with conflicting pressures, generally the anterior cingulate gyrus, was used to handle the "choice" of whether and when to lie, and (2) this often acted with some form of inhibitory mechanism to suppress the truthful response, whether that response was verbal or motor. These results are in accord with the theoretical prediction that witnesses will have a relatively greater amount of difficulty fabricating testimony than telling the truth.

92. Kozel et al., supra note 81. These areas were the right anterior cingulate (Brodmann areas 6, 8 and 32), the right orbitofrontal and inferior frontal (Brodmann areas 38 and 47), and the right middle frontal (Brodmann area 46).

93. In a parallel to the early history of the polygraph, Kozel and his collaborators have apparently licensed their technology for commercial forensic purposes. See http://www.cephoscorp.com. We can only hope that the lessons of the polygraph's commercialization can be a guide to avoid the pitfalls of technology stagnation.

94. Cf. Spence et al., supra note 22 (coming to a similar conclusion regarding the presence of executive differences and inhibition of truthful responses as a common feature of early studies, because truthfulness is the "baseline" and the brain must make extra efforts to respond in some other way, as partially evidenced by the failure of most studies to show areas that are more activated by truth-telling than by telling lies). But see Langleben et al., supra note 90; Kozel et al., supra note 90.

95. This comparative difficulty in increased "cognitive load" is directly and quantitatively measured in an MRI context. However, it also provides the basis for the successful use of cross-examination against an untruthful witness, who will generally lack the cognitive capacity to generate false memories quickly and with the same detail and consistency as true memories, a limitation the skillful cross-examiner exposes. Chris William Sanchirico, Evidence, Procedure, and the Upside of
In a very recent study, Raine and his colleagues have provided another line of confirmatory evidence in this regard. In line with traditional clinical neuropsychology, they attempted to find biological differences between a normal control group and individuals known prior to the study to be "liars." Both liars and controls were recruited from "temp" agencies in Southern California, and twelve liars were identified by psychological tests that indicated regular use of "conning" others, use of aliases, psychopathic personality characteristics, and false claims of illness to acquire disability benefits. These liars showed an average increase of twenty-two percent in the amount of "white matter" in their prefrontal cortex and a decrease in the amount of grey matter, which the researchers hypothesize as involving respectively, an increased capacity to confabulate items and a decreased inhibition about doing so. In any event, this anatomical differentiation among persons (as opposed to differentiation among behaviors) is supportive of both the biological locus of lying and the potential for identifying it.

Despite the theoretical soundness of MRI deception research generally, the current divergence between the tasks posed and the results obtained suggests that no single overarching marker of deception is likely to be soon identified. However, even with regard to simple lying, a unique marker has not been identified. This in itself appears to be remediable by use of statistical methods that can incorporate multiple predictors, but it does not solve the problem of whether a particular person, who might have a reasonably unique pattern of activation during deception, can be evaluated for truthfulness. Most tellingly, in the only study to break apart individuals the individuals all showed distinct patterns of activation during deception. It is therefore probable that the other studies also contained a large amount of individual variation hidden within their group reports. Because all of the above were in effect pilot studies, it will be interesting to see whether replication of the experiment will show—at least on the group level—the same areas of activation during deception.

Of particular importance to legal application would be a study so far apparently not performed. That is, if a particular individual was retested, it would be of interest whether on an individual basis the differential levels of activation could be replicated, or if in effect people lie in


97. See id. at 323.

98. An interesting study melding the two approaches, in order to deal with individual variation, would seek to find out if habitual liars lie in a different way than "controls" do. This seems likely as a theoretical matter, and would allow for the elimination of some of the noise in fMRI studies, since the set of markers would be tailored for those classified as habitual liars and for the rest of the population.
different ways on different days. The studies so far demonstrate that lies are distinguishable—in gross—from honest responses using neuroimaging. However, if the results are not reproducible, despite holding the subject and the experimental paradigm constant, there is serious doubt that the technique will soon be ready to operate as a lie detector. On the other hand, because the Kozel (2005) Study in particular shows the predictive power of a previous study in a subsequent one, there is some evidence consistent results can be achieved under those controls, and a challenge to the credibility of a crucial witness using fMRI should be scientifically feasible and legally cognizable, perhaps after some level of calibration for that individual's variation.

II. A MODEL FOR ADMITTING DETECTION OF DECEPTION EVIDENCE

What fMRI does, at least with regard to issues of memory, is simply strip off further the layers of assumption and non-specificity that attach to the “guilty knowledge” or feigned ignorance testing paradigm. Rather than relying on merely an empirically-derived construct such as the “orienting response,” it is now possible to see the brain react to new information in a way distinctive from the reaction to material that is familiar. As with the polygraph and the EEG used to detect concealed knowledge, there will be an argument that the cases are rare where this type of information will be useful, or where the examined party will not already have been given access to the material details and become familiar with them through means of discovery. However, the duty to disclose adverse information can be regulated by the court in the following way: if a party discovers material evidence, and an opposing party witness’ familiarity with it is probative of an issue in the case, then under certain circumstances, the witness could be examined for this familiarity before the truth about these facts is revealed. Criminal defendants, of course, would be protected against such examination by their right against self-incrimination, but in circumstances where they might be innocent, they should share an incentive to take the test with the prosecution (which clearly has a greater belief in their guilt).

Once a particular neural pattern is validated as indicative of memory or deception, then the recognition of this pattern on a series of neuroimages can occur by computer, or by someone independent and

99. There is considerable evidence that the brain will react differentially to familiar information even if it has been consciously “forgotten” and it is not retrievable. This relates to the challenges of having neuroimaging overcome deception that is bolstered by self-deception, which, for cognitive science purposes, means simply that little or no cognitive effort is required to suppress the “true” answer to the question.

100. But see FAIGMAN ET AL., supra note 11, § 19-3.4.2 (arguing that once a guilty knowledge test “breaks through” as courtroom evidence, that law enforcement and others will begin to gather material at the outset of the case that can be used to implement it).
unfamiliar with the purpose of the test. The time stamp of the image is then synchronized with the time sequence of the examination and questions and responses are correlated. Expert testimony would be used to explain the nature of the technique, the criteria used for scoring, and the background assumptions. Cross-examination or rival expert testimony would then be introduced focusing on issues such as the error rate, possible contamination, invalid assumptions and so forth. All of these types of challenges occurred (and continue to occur, although usually with more precision) on the uses of DNA technology. Nor is it difficult to believe that this process, as absurd as certain of the challenges to DNA now seem in retrospect,\textsuperscript{101} served to refine the process and presentation of DNA evidence in court. Whether evidentiary challenges impeded or improved forensic use of DNA, they certainly did not halt it. Over the last decade, this process along with technical advances in DNA sequencing have converted this technology to the "gold standard" for identification,\textsuperscript{102} even displacing the traditional bias toward eyewitnesses, which the courts now clearly consider to be less reliable.

In a certain manner, the neuroimaging of a subject's information state is merely the converse of the collection of DNA traces left behind. Just as it is difficult for an individual to interact with the external world without affecting it, and leaving behind his distinctive genetic pattern in the form of spent surface cells, each distinctive part of the external world leaves its mark on the human body, acting on the cellular level to imprint a trace of itself. Formerly, only those comparatively rare parts of the external world possessing the capacity to adhere to the surface of the subject (fibers, mud, semen, etc.) were able to imprint themselves as evidence of the subject-environment interaction. The distinct auditory, visual, olfactory or tactile properties of an environment, although sensed and recorded by the subject, could not be accessed without accessibility to those cells (brain cells) acted on by these features.

As to deception, or misleading \textit{per se}, it is less clear how or when neuroimaging will prove its worth. A thoroughgoing evolutionary approach, which sees deception as basic to the adaptations underlying language and thought, would be expected to predict a "module" for deceit recognizable when activated. The difficulty in finding this module is that we do not turn deceit off and on the way we might the reflex to sneeze. Rather, the potential for multifarious deception may be so deeply

\textsuperscript{101} See, e.g., United States v. Chischilly, 30 F.3d 1144, 1154 (9th Cir. 1994) (rejecting claims made by academic commentators that DNA matching failed tests of general acceptance and relevancy and had unknown rates of error).

\textsuperscript{102} See \textsc{Faigman et al., supra} note 11, at § 25-1.2.1 (recounting the furious evidentiary debates that surrounded DNA through the early 1990s and concluding that "in little more than a decade, DNA has made the transition from a novel set of methods for identification to a relatively mature and well studied forensic technology").
embedded in our social decision-making that it is simply not possible to
distinguish it from the ordinary way in which we communicate where we
"modulate" what we say, how we say it, and what we do not say, in order
to convey the desired impression. A recent meta-analysis of 1,338 cues of
deception appears to indicate this is indeed the case, and there is a
significant psychological gray area between truth and deceit. Therefore,
any global measure of deceit detectable in the brain would be like the
behavioral indicators of deceit and be better assessed as a continuous
rather than discrete variable. As a practical matter for neuroimaging,
therefore, it is quite unlikely on the current facts that we could reliably
detect when someone is "shading the truth" or "misleading" because
those states could not be objectively specified as processing or
transmitting information in a way distinct from ordinary social
communication.

Somewhere in the middle of plausibility is lie detection, and this is
where the focus of current research is likely to be. As there have only
been exploratory studies, only of few of them replicated, and all of them
with significant limitations, nobody is likely to be able to walk into a
courtroom tomorrow and challenge an opposing witness to spend an
afternoon in a magnetic resonance chamber. On the other hand, at the
basic level of distinguishing a group of lies from a group of honest
answers, these tests do work. Moreover, there is a scientific rationale in
the form of an argument for why they should. Lies as informational
products require more processing steps or inputs, in most instances, than
truths. If a lie is called for, it appears that some executive or "conflict-
resolving" portions of the brain must hold in working memory both the
honest answer and the proposed lie (whether retrieved or made up on

103. Bella M. DePaolo et al., Cues to Deception, 129 PSYCHOL. BULL. 74, 105 (2003).
104. No one familiar with developments in the area of cognitive technology can be unaware that it
is likely to pose significant privacy issues in the future. When neuroimaging becomes more accurate in
detecting deception than the polygraph, one non-libertarian justification restraining the use of the
deception detection will be abrogated, thus requiring a more direct confrontation between individual
rights and institutional desire to know. The theme of this Article is that testing be initially developed
in the limited and more controlled circumstances of court cases where the technology is critical to
accurate adjudication. For reasons of cost and the substantial burden imposed on employees, quite
apart from the novelty of the technology, it is unlikely that most employers would resort to functional
neuroimaging in order to investigate or screen their employees. In particular, there would be no
neuroimaging unreliable, the rationale of the Act (that employers not be allowed to use a bogus lie-
detector as an intimidation tactic) remains valid. However, if courts rule—as I believe is inevitable—that some types of this evidence are sufficiently reliable, the need for independent employer use
remains dubious, since many deterrent ends would be achieved simply by the credible threat of a
potential legal action (i.e., for conversion where embezzlement is suspected) and a court-supervised
form of evidence would be available. Hence, it is unclear to me that a slippery slope argument directed
against my proposal has great force. On the other hand, I reiterate that potential uses of neuroimaging
for investigatory purposes by state authority might well be more prone to abuse.

105. See Spence et al., supra note 22, at 1760.
the spot) and compare the two. In addition, the true answer and the associated motor activity that would communicate it must be inhibited. Based on the limited results of the recent studies, it appears these tasks are accomplished differently depending on precisely how they are framed, and it may well be that different individuals accomplish them using different neuroanatomical structures.

Nevertheless, there is some hope that, for instance, an individual subject’s pattern of honesty or deception could be reliably identified. As in the “control-question” method in polygraphy, the subject would be asked to generate a certain number of known lies. The neuroanatomical correlates to these events would then be compared (in a blind test) with those corresponding to his known truthful answers. The statements as to which his credibility is in doubt would then be classified to see into which category they would fall. This avoids many of the pitfalls inherent to the “control question” technique as now practiced by polygraphers, because: (1) it does not depend on arousal but on the similarity of cognitive demands, therefore making it less susceptible to confounding influences: and (2) the outcome is a complex pattern of activation—if there is some irrelevant psychological influence such as stress that causes the target questions to be different than the known truths, this influence would be far more likely to generate a third distinctive pattern than it would the pattern corresponding to the known lie, meaning a minimum of false positives. I will speculate regarding the admissibility of this type of evidence in the future, assuming a modest development of the technology along the lines I have just outlined.

A. **The Frye Standard and the Contrast with Polygraphy**

Many states continue to adjudicate the admissibility of scientific evidence under the principle of “general acceptance” in the relevant field. In this respect, at least as to methodology, functional neuroimaging has significant advantages over the polygraph. Although what constituted the appropriate reference group for polygraphers was always a matter of acrimonious debate, the fact that the majority of psychologists never accepted it as a valid instrument was one of the primary bases for its exclusion under the *Frye* standard.

By contrast, fMRI research not only is common throughout what is undoubtedly the relevant field, cognitive neuroscience, it is closely related to the general use of MRI in many different disciplines. Thus the fact that an fMRI measures how a brain reacts to an event would be “generally accepted” under the *Frye* standard.\(^6\) In addition, the ability

---

106. See Kulynych, *supra* note 30, at 1265 ("For neuroimaging, the general scientific consensus is that most methods of PET and MRI scanning consistently produce useful information about brain status, within the known limits of temporal and spatial resolution for a given technique.").
to infer the differing effects on a brain between a novel and a familiar stimulus would likewise be generally accepted, as issues of learning and memory are the central focus of cognitive neuroscience, and the match between behavior and structure activation is relatively well developed in this area.

What would be more questionable would be any general acceptance of whether a particular neural pattern corresponds to a lie. It would have been theoretically possible for any of the past studies on this question to show a distinct pattern that, based on validity of the technology and the statistics used, would be generally accepted. However, this did not occur, and the initial researchers did not claim this. In my speculative example where a single individual is tested and retested, yielding consistent differences between lies and truthful answers, the question would be a close one, but assuming that a distinctive pattern was shown for statements known to be untrue, this research—considered apart from its conclusions—does not deviate from the normal practice in clinical cognitive neuroscience that measures the responsiveness of single subjects to stimuli such as pharmaceuticals. It would therefore also be likely to find “general acceptance,” in part because a cognitive neuroscientist would a priori expect that the cognitive operation of lying would produce some distinctiveness on a neuroimage, even if it might be too subtle or variable to be of use (this latter problem being overcome by the rigor of the research).

B. **Scientific Status of fMRI Under Rule 702**

Under the federal rules, a cognitive neuroscientist could testify as to either memory or deception, provided that “(1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.” Again, with 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, although the principles and methods would be deemed reliable, and presumably they would be applied reliably, the evidence would fail on the first criteria, because there are not currently any reliable “deception signatures” against which the individual’s neuroimaging result could be matched to say whether it was

or was not deceptive. (As noted above, current research is focused in part on identifying multipart signatures by computer-aided pattern recognition). Moreover, the particular brain structures implementing deception, or alternative sets of structures varying among persons, have not yet been identified in more than the most tentative fashion, undermining any inference that activation of such structures in a particular subject is indeed "deception." As under Frye, at least once the foregoing scientific groundwork has been established, a researcher could develop reliable data for the case on what counted as deception (for instance, a set of data on known lies showing a distinct pattern with little variation), and this application of generally accepted procedures would be admissible for Rule 702 purposes.

Under the Daubert standard, now incorporated into Rule 702, a court would inquire into testability, error rate, peer review and publication, and general acceptance.\textsuperscript{108} Clearly any general hypothesis about how a particular part of the brain generates behavior, or the consistency of the neuroimaged response to a particular task, is testable, and if posed as a general rule, falsifiable. To the extent the particular technique has been applied for purposes of classification, the best error rates from studies currently in press approach ten percent, although they have, as mentioned, not dealt directly with the situation of an individual being tested for veracity, and their error rate in this context is therefore unknown. At the current time, the estimated error rate for any particular pattern activation as indicative of lying, without calibration on the individual, would be unacceptably high for admissibility. Because there are literally thousands of event-related fMRI studies in the peer-reviewed literature (although very few on deception), this would be a relatively easy test, since this factor is generally seen as referring largely to the methodology rather than the particular application.\textsuperscript{109}

The particular application to the assessment of deception, and more precisely, to the case at hand, would instead be governed under the rules laid out in Joiner, which requires that the expert testimony be able to close the "analytical gap between the data and the opinion proffered."\textsuperscript{110} Applying Joiner, the proponent would face several challenges. First, for the admissibility of an opinion regarding the credibility of a specific subject, the court should be shown that neuroimaging can in principle identify lies, and second, a causal theory of why the particular structures implicated as involving lies are in fact involved. As mentioned, any such causal theory remains only in development, awaiting identification of


\textsuperscript{109} See FAIGMAN ET AL., supra note 11, at § 1-3.4.

those areas that are consistently shown to activate during lying. However, there is at least the potential of making such a connection between the data (an fMRI) and the process of forming a lie.

By contrast, it is generally considered unlikely that the polygraph, at least outside the guilty-knowledge test, will ever be able to bridge this gap. The most common technique of the polygraph, put simply, attempts to detect deception by noting the difference in physiological response between "control questions" and the "test questions" designed to elicit legally relevant information. At the most fundamental level, the difficulty is one of validity—in Joiner terms, there is too great a gap between the method chosen and the inferences it attempts to make. Its invalidity stems from a variety of sources, including what is often termed "construct" validity—that there is no necessary connection between differential response under the "control question" technique and deception—the response may stem from stress, fear, surprise or other sources activating the sympathetic nervous system, such activation being the only thing the polygraph measures.¹¹¹

Even if satisfying the scientific standard, it is likely neuroimaging evidence will confront the attitude of the courts, alluded to earlier, that the jury is to be the judge of credibility, and the use of experts in this area is "invading their province." Procedurally, this is implemented as a bar on the use of expert evidence on a witness's credibility.¹¹² Not all courts adhere to this rule, which, as one might expect, is related to barring polygraph evidence, and those that do have a number of exceptions. The main purpose of this rule, apart from keeping out the polygraph, appears to be related to barring expert judgments about character, rather than specific testimony about why a brain scan during a particular statement has indicia of deception.¹¹³ The evidence presented by the expert is about the scan, or perhaps at most the statement, and not directly about the witness. Consequently, evidence of this sort allows a jury to make its own assessment of the witness.


¹¹². See David H. Kaye et al., The New Wigmore: A Treatise on Evidence, Expert Evidence § 1.5 (2004) (discussing this rule and its rationales, and noting that "when sufficiently valid and reliable methods for diagnosing truthfulness in specific instances are developed,... the blanket rule should be reexamined").

¹¹³. An exception along these lines might be comparable to that sometimes employed to admit expert testimony about "false confessions" explaining why a confession made in police custody might be untrue, although made. Considering indicia in confessions, or in the circumstances surrounding them that casts doubt on their veracity, see id. at § 1.55 (discussing the trend to admit this sort of testimony).
C. LIMITED ADMISSIBILITY MAY BE PROPER FOR PURPOSES OF RULES 402 AND 403

Scientific validity under Rule 702, as elucidated by Daubert/Joiner, is only the threshold measure of judicially-determined reliability that might permit such evidence to be lawfully used. It still remains to be considered whether neuroimaging evidence can be projected to satisfy the remaining strictures on presentation of material at trial. Because of the expense involved, the amount of court time required, and the potential to confuse the jury, Rule 403 would admit this evidence only in circumstances where the evidence of credibility was particularly relevant (under Rule 402) and probative of key issues. Three particular circumstances come to mind. Because functional neuroimaging has achieved a greater level of validation with regard to malingering, it may be called for in civil cases where insurance fraud is suspected, for instance, if the plaintiff is suing on a claim whose proof is founded on essentially his own testimony regarding pain or impairment. Another class of cases that turns on credibility is where evidence of the criminal defendant’s guilt derives primarily from an alleged accomplice or other potentially untrustworthy source. In the latter circumstances, the impeccable witness could be subject to a potential feigned ignorance test as to the details of the crime scene he claims to have viewed or the context of the confession he claims to have heard. A third type of case where current technology might be obviously relevant would be in trade secret or patent actions, in which the defendant may have to disclaim familiarity with the work he is alleged to have misappropriated. Particularly when such actions verge on industrial espionage, in which blueprints, designs, or prototypes are taken and misused, a party’s credibility about unfamiliarity is both material and potentially testable.

Moreover, in order to avoid neurological evidence overwhelming the case, it could be limited in two other ways that some courts already employ for polygraph testimony and therefore, a fortiori, would be appropriate for a developed neuroimaging test of deception. First, the testimony should be limited to impeachment only, rather than for its substantive truth. In other words, neuroimaging during witness testimony would call statements into question, but they would not

---

114. This is true whether or not the action actually proceeds to trial or is disposed of by pre-trial summary judgment. See Fed. R. Civ. P. 56(e) (requiring facts presented for adjudication by motion to be “as would be admissible in evidence”).

115. “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 . . . .” Fed. R. Evid. 403.

116. Or international espionage for that matter, which would presumably explain why the Defense Advanced Research Projects Agency is listed as providing funds for Langleben et al., supra note 81.

117. See United States v. Piccinonna, 885 F.2d 1529, 1535-37 (11th Cir. 1989) (allowing polygraph evidence only for impeachment or corroboration).
establish the converse of the statements (that the witness did in fact know something of which he claims to be ignorant, or was ignorant of that which he claims to know); his statements would simply be nullified. Second, in order to admit this form of evidence to impeach, the proponent of the evidence would have to show not only that credibility generally was of key importance to the case, but also that the credibility of the particular witness who he proposes to examine has been put in question by at least some other form of evidence such as contradiction by others or a reputation for untruthfulness. Given the foregoing constraints, the evidence should be admissible under the basic hurdles of Rules 402 and 403 or their state equivalents.

D. Admissibility of Neuroimpeachment Should Proceed Under Rule 607

Assuming neuroimaging evidence is introduced for purposes of impeachment, one might legitimately ask how to further classify it in conventional terms, since the Federal Rules limit the type of material acceptable for these purposes. If, as I propose, the proponent of the evidence is seeking to use brain imaging to impeach, this is clearly challenging the direct testimony of an opposing party with extrinsic evidence going to the credibility of this testimony. In particular, it is introducing an event, recorded by fMRI in which the witness made a statement consistent with his direct testimony, but simultaneously with making this statement, his brain state recorded that this statement was not true (i.e., it showed novelty or familiarity where this was denied, or alternatively, indicated the individual was lying).

Obviously, the rule on prior inconsistent statements does not apply here, unless one stretches this term to mean that while the witness was saying one thing, his brain was "saying" another.\(^{118}\) Nor can one easily argue for the use of Federal Rule of Evidence 608, as the test does not bear on a witness's character for untruthfulness.\(^{119}\) No one is able to offer a competent opinion on the individual's character based on the previous examination alone. Under 608(b) it might be possible to inquire into the fMRI on cross-examination of the witness himself, as a "specific instance" where the witness was untruthful, save that this rules bars extrinsic evidence, which in this case would include the fMRI itself.

The way out of this legal puzzle for fMRI evidence is to admit the evidence on the general ground of Federal Rule of Evidence 607, because whatever else it might be, it is evidence tending to contradict a statement of a witness.\(^{120}\) The evidentiary bar of "Rule 608(b) does not

\(^{118}\) Fed. R. Evid. 613(b).

\(^{119}\) Admissible "evidence may refer only to character for truthfulness or untruthfulness...." Fed. R. Evid. 608(a).

\(^{120}\) Impeachment by contradiction is properly considered under Rule 607, not Rule 608(b).
apply . . . when extrinsic evidence is used to show that a statement made by a [witness] on direct examination is false, even if the statement is about a collateral issue."121 Under this view, admissibility is present because the verbal statement at trial is consistent or identical with that given during pretrial examination. The extrinsic evidence of the fMRI is admissible, as this contradicts the truth of the equivalent assertion repeated by the witness at trial. At the time of the prior statement, the contemporaneous fMRI had a tendency to show that the statement was false when uttered. If the same statement is made, its veracity is put in doubt by its relationship to the questionable prior statement. Since the witness knows of this threat, he would presumably refrain from repeating the deceptive testimony, but this places him in another bind, because he can then be impeached with what he actually said under the prior inconsistent statement rule for hearsay, even if he tries to remain silent on the matter.122

E. INCENTIVES FOR TRUTHFULNESS AND PROGRESS BY STRUCTURING ADMISSIBILITY

Assuming that the very early stages of research reported here eventually produce a more reliable method of assessing witness credibility, its gradual introduction will have to take account of the incentives of the litigants and, ideally, encourage them to choose tests that will be more, rather than less, accurate. According to the assumptions of the model, the adjudicator will prefer truth to falsehood, as will at least one party in the dispute. The other party will usually prefer falsehood,123 whenever his expected chance of winning is less than fifty percent. If there is a certain amount of uncertainty about the test, then the parties’ willingness to have it performed will depend on their individual assessments of the test outcome. If those assessments are equally optimistic or are risk-prone, then there will be a possibility of a joint stipulation regarding the results. If the parties are risk-neutral or risk-prone, however, one of them will have no interest in pursuing the DDD option without further legal adjustments. Because of the nature of techniques such as fMRI, particularly in a feigned ignorance paradigm,

122. See United States v. Gajo, 290 F.3d 922 (7th Cir. 2002) (concluding that inconsistency under Fed. R. Evid. 801(d)(1)(A) “may be found in evasive answers, . . . silence, or changes in positions” (citation omitted)).
123. Governmental representatives, of course, have been given incentives to favor truthful outcomes, and other legal and social norms regulate the extent to which even lawyers in civil matters can encourage false beliefs.
the test could in fact be ordered—and should be ordered—to prepare the expert evidence prior to the trial. The initial evidentiary showing described above should be made in pre-trial motions, specifically arguing that the proposed method of testing credibility satisfies the rules for the introduction of scientific evidence, that the credibility of a witness to be so tested has already been put at issue, and that the credibility of this witness is sufficiently material to the outcome of the case as to warrant the use of an examination by MRI.

For convenience, I will call the source of the substantive evidence "W," who offers up statement "S." When S is uttered by W, it will be denominated "Sw." The opposing party—and potential proponent of impeaching fMRI evidence—will be "X," the cross-examiner, whose goal in the evidentiary context is to rebut or minimize anything offered by W. In part this occurs by attacking the general credibility of W, his character, background and reputation. In part this also occurs by focusing on the inherent likelihood of S, apart from the fact that it is coming from the mouth of W. Finally, and this is where DDD can be employed, X will attack the specific credibility of Sw, by attempting to show that a particular statement coming from that particular speaker should be suspect. Under the polygraph paradigm, the proponent of DDD evidence would usually have been W, testing himself, and it would be a test-negative result that would be offered as either substantive or corroborative evidence of the truth of Sw. Hence this system favors any test that is highly insensitive to deception or to the untruth of Sw, one that would produce a large number of Type II errors.

In general, the impeachment value of a test-positive result on a DDD in some particular instance, "D'," is determined by both the rates of Type I and Type II errors, such that the probability of the Sw's truth, "T,'" given the result of the test, is defined by the following Bayesian updating:

\[
\text{Prob}(T|D') = \frac{(D'|T)}{((D'|T) + (D'|\text{not } T))}
\]

The probability of the numerator shows that a result indicating deception occurs during truth telling and is the false positive rate, while the probability that it occurs with falsehood is 1, which is the rate of false negatives. This converts the above equation to a more easily

124. Bayesian updating involves the adjustment of perceived probabilities by the addition of a new piece of information. Here, the critical question is the probability that the statement S is true. If S bears indicia of truth when spoken by W while under examination by a DDD, this increases the probability of the truth of S. How much this probability increases depends on the prior probability of the truth of S, as well as the on the accuracy of the DDD.
understandable form:\textsuperscript{125}

\begin{equation}
\text{Prob}(\text{T}|\text{D}') = \frac{\text{false positives}}{\text{false positives} + (1 - \text{false negatives}) (1 - \text{T})}
\end{equation}

Introducing one more piece of algebra, the false positive rate is designated as "E," and the false negative rate as "E'." Ultimately it is the sum of these errors, E + E', that a rational legal system should seek to minimize over both the individual case, and over the progressive development of more accurate techniques of adjudication and fact investigation. The equation in that form becomes:\textsuperscript{126}

\begin{equation}
\text{Prob}(\text{T}|\text{D}+) = \frac{E \cdot T}{E \cdot T + (1 - T) - E} (1 - T)
\end{equation}

The goal of rules of procedure and evidence would therefore be to encourage each party, but in particular the party with most control over the testing procedure, to have an incentive in the form of greater expected utility for the reduction of E and E'. If the utility of the parties can be reasonably approximated by the changing likelihoods of a favorable litigation outcome, I propose that the choice of test be given to X, who has an incentive to minimize at least Type II error (\text{AU}(X)/\text{AE}, < 0).\textsuperscript{127} Under Rule 403, the court is required to prevent "prejudicial" outcomes. Here prejudicial outcomes would be characterized as dependent on the possibility of "false positives," which will fall in relation to both T and E'. Consequently, in order to avoid exclusion on this ground, X will also have an incentive to choose tests with a lower rate of E, and minimize Type I error (\text{AU}(X)/\text{AE}_i < 0), as well as to preferentially apply tests only when T is already well below

\textsuperscript{125} These equations would also play into the initial evaluation of admissibility, as it would be the burden of the testing proponent to show that there was (1) sufficient doubt about T (0<T<1), (2) a likelihood that the marginal effect of a test-positive result would change T, and (3) that the rate of false positives and negatives was sufficiently low that it would not overwhelm any inferences that could be drawn from the test results.

\textsuperscript{126} A numerical example may provide clarification. Suppose the pre-test likelihood of a statement, after an initial showing of credibility problems, is eighty percent. The rate of errors, both false positives and false negatives, E, and E', is ten percent. Then if there is deception detected, the new likelihood is .08/(.08 + (.2 - .02)) or .08/.26; in other words, there is now only approximately a thirty percent chance of the statement being truthful. Thus, if this is the key issue in the case, the preponderance of evidence W would have otherwise provided disappears. Readers can easily derive for themselves that if T=1 or if T=0, there is simply no point in conducting the test. Likewise, if the error rates become high, the probative value quickly vanishes. However, even a test with a twenty percent misclassification rate would be useful here, producing a conditional probability of the initially plausible statement of .16/(.16 + .2 - .04) = 50/50 proposition.

\textsuperscript{127} As mentioned, a Type II error in this context would mean the DDD registers no deception, even though the witness is in fact being deceptive. This is the worst of all possible worlds for the cross-examiner, since his use of the DDD has in effect "bolstered" harmful and fallacious testimony.
one. By contrast, the current method of evaluating the admissibility of DDD evidence has the witness W as the proponent of polygraph evidence. Although W has the right incentives about Type I error (\(\Delta U(W)/\Delta E_i < 0\)), it is difficult to get them aligned so as to minimize Type II error.\(^{128}\) A witness would like to be "validated" by the failure of a DDD to register deception, regardless of whether or not they are telling the truth; hence, if all else were equal, they would prefer relatively insensitive devices with a high type II error.

Fundamentally, any test must be able to pick up the phenomena of interest, and this is an issue of sensitivity. The probability of a false positive is further reduced in the impeachment context if X is required to bring in evidence of credibility problems of W before the test is permitted. This could be a credibility attack of any of the kinds suggested before: on the witness W, on the statement S, or on \(S_w\), since ultimately the impeachment relevance of attacks on the witness or on the statement simpliciter exists only because of the probable connection they have to the statement whose truth value is actually at issue. All else equal, it is reasonable to assume that individuals with independent evidence against their credibility will be more likely to utter an untrue statement. This directly affects the overall rate of false positives (the numerator above), which is determined by the background likelihood of true statements. Call this foundational impeachment "M."

Moreover, the threshold requirement M of attacking W's credibility by conventional means such as past convictions, prior bad acts, character, or reputation will have the effect of—in those cases where the statement tested is in fact true—of encouraging W's cooperation with the detection of deception test. As the rules of procedure and evidence are currently structured, this cooperation made by stipulation as to discovery by neuroimaging would seem to be necessary in most cases, as an MRI would undoubtedly qualify as a "physical or mental examination" under the Federal Rules, which limits such examinations to parties or those, such as children, under the control of parties.\(^{129}\)

---

\(^{128}\) E, is the false positive rate so its reduction reduces false positives by definition. In addition, when Truth is more likely (\(T \approx 1\)), there are more true statements on which an error leading to a false positive can be made. For instance, if something were absolutely true, any register of deception made while uttering it would be a false positive. By contrast, when the prior probability of T is low, a result indicating "deception" is more likely to be associated with a statement that is actually false.

\(^{129}\) Fed. R. Civ. P. 35(a). This rule also allows such court orders only when "good cause" has been shown and the examined party's "condition" is in question. Arguably, this language could be stretched to incorporate the individual's veracity on a particular declaration since the physical basis for such statements in terms of recall—or the lack of such a physical basis, indicating confabulation—can now be potentially examined. The foundational impeachment would then be formalized as constituting the necessary good cause. However, because of the ambiguities involved in this interpretation, and the limited reach of this rule, it seems more practical to focus on the development of incentive structures that encourage stipulation.
Because impeachment information $M$ is as a practical matter very damaging to $W$'s credibility despite its dubious relevance to the truth or falsity of his statement, $W$ might be eager to bring evidence on the statement. This is particularly so if by rule or as a practical effect a test-negative result on a test chosen by his adversary $X$ would disallow $X$ from introducing $M$ (or some elements thereof) at trial. Thus $X$ has to put a certain amount of credibility-attacking material at risk in order to get the benefit of neuroimpeachment. If $W$ refuses to take the test, the proponent of $W$'s evidence might well consider withdrawing $W$ (as they would do, presumably, if a deception positive result was obtained), particularly if the court was able to issue an instruction regarding the refusal to take the test that allowed a negative inference to be drawn from that refusal.

The end result would be that in many circumstances, if the evidence is collected “in the shadow” of a valid DDD, there will in fact be no “battle of the experts” or “trial on how to conduct trials.” There will be less evidence actually introduced—including less bad evidence. The credibility-challenged $W$ may not be presented, and consequently neither the foundational impeachment $M$, nor the neuroimaging, will ever be seen. Alternatively, if $W$ “passes the test,” the foundational impeachment will not be presented, and neither will the test-negative result that nullified it, because the test-negative result “bolstering” $W$'s statement would not be admissible without a prior attempt at the impeachment to which it responds. Since the sort of material used for foundational impeachment is not particularly accurate in any event, its occasional loss should not be mourned—the best use of character and reputation evidence may be in form of leverage, rather than presented as genuinely probative of truth. In neither circumstance will the jury ever hear the term “fMRI.” Magnetic-resonance evidence would only be discussed if the proponent of $W$, despite his failure on the DDD (or an instruction regarding his refusal to cooperate) decides to “fight the test” (or justify his refusal) using expert testimony, and as the tests improve,

---

130. Unless $W$ is a party, the actual persons with an interest in the case are often in the same informational position as $X$—they do not really know whether the crucial statement of $W$ is true or not, and their opponent’s attack on $W$'s credibility will normally influence the beliefs of everyone involved about this fact.

131. Hence, even if one believed in the adjudicative value of the foundational impeachment, its exclusion also excludes the rehabilitating result on the fMRI, leaving the parties in much the same evidentiary state as they would be if there was no exclusion (at much less cost and confusion). Putting $M$ at risk is thus not only a way to create incentives for the parties, but also approximates what the end result would be if the material $M$, used to open the door to neuroimaging, were responded to—as would seem only just—by the countervailing evidence it had engendered.


133. Recalling also that $W$ will already have two strikes against his credibility in the form of foundational impeachment and a test positive result, unless the credibility of $W$'s testimony is
Putting this situation in an illustrative game-theoretical (extensive) form, with payoffs to W and X equal to the probability of a favorable verdict, the game between the parties might look like the accompanying Figure 1. In the particular example, even if W knows that \( S_w \) is false, he will still marginally favor taking his chances with the test, so long as there is even a minimal error rate, because this cannot be worse than a jury instruction indicating W refused the offer of a test for deception. Of course, if he knows \( S_w \) to be true, he will certainly agree to the test. As for X, he does not know whether \( S_w \) is true or not, but his decision is absolutely crucial (when indeed the entire dispute hinges on this point), it is questionable whether it would be worthwhile to make this fight.
governed by .3T + .9(1 - T) > .5, meaning that so long as he thinks (in this particular example) that there is at least a one in three chance that the statement is false, he will go ahead with the test, despite the risk of the loss of his foundational impeachment.\textsuperscript{134}

More generally, X will propose impeachment by detection of deception where:

\begin{equation}
(4) \ [TE_i + (1 - E_i)(1 - T)] \ast [(\text{prob}(\text{winning})|D') - (\text{prob}(\text{winning})|M)],
\end{equation}

which equals the marginal gain in victory probability if the test shows deception, thus "working out" for X, is greater than

\begin{equation}
(5) \ [T + E_i(1 - T) - TE_i] \ast [(\text{prob}(\text{winning})|M) - (\text{prob}(\text{winning}))],
\end{equation}

which is equivalent to the marginal loss if the results of the test favor W. Taking the probability of X winning and calling it "P," and multiplying through, we are left with:

\begin{equation}
(6) \ (TE_i + TE_i + 1 - E_i)(\text{P}|D') - (\text{P}|M) > (T + E_i - TE_i - TE_i)((\text{P}|M) - P),
\end{equation}

which further reduces to the following decision equation: X will use a DDD if

\begin{equation}
(7) \ (2E_iT - (2E_i(1 - T)) + 1)(\text{AP}|D') > T(\text{AP}|M)
\end{equation}

Equation (7) measures the comparative marginal improvements in success made by the foundational material and the test-positive result, and assesses the potential gains in a test positive result with the losses (reversion to a baseline chance of winning) if a test-negative result occurs. Because $E_i$ and $E_i$ must be less than one in any case where the test would pass \textit{Daubert} standards, and $T$ is always less than one, X would behave approximately as we would want him to, because the previous equation (7) may be reduced to:

\begin{equation}
(8) \ (\text{AP}|D') > T(\text{AP}|M)
\end{equation}

If using the test gives a greater marginal gain in the probability of winning than does using the impeachment evidence (discounted by the background estimate of statement truthfulness), then X will use the test. Taking the right-hand of the equation first, we see that as the background likelihood of veracity ($T$) increases, the motive for examination falls, and likewise, as the power of traditional impeachment ($M$) increases, it becomes increasingly less likely that there will be resort to neuroimpeachment. Now taking the left-hand of the equation, we can see that the potential marginal improvement by DDD will obviously create an incentive. However, as we know from prior equations (2) and (3), the amount of new information one can extract from a DDD result is

\textsuperscript{134} In this equation, $T$ is the prior probability as perceived by X, who puts at risk his foundational impeachment; the expected marginal gain from a confirmatory DDD has to be greater than the possible loss. In the numerical example, the inequality reduces to $.6T < .4$ or $T$ must be less than two-thirds to go ahead with the test.
itself negatively dependent on both veracity and the rate of $E_r$. In addition, in this form, the more common-sense incentive for $X$ to minimize error rates showing no deception ($E_r$ or test-negative results) becomes even clearer, since the net marginal effect of an increase in $E_r$ will make it less likely that the equation will be satisfied, and thus, that neuroimpeachment will be warranted. Because cross-examiners are the proponents of this type of evidence and have been provided with incentives against error, there is some hope that the examinations they propose will satisfy, or come to satisfy, the court. Even in the absence of a compulsory process requiring their adversaries to undergo such examinations, there will be some circumstances where they may voluntarily agree to such investigations, either because they have almost nothing to lose, or because their estimate or knowledge of their own veracity exceeds that of their skeptical opponent.

**CONCLUSION AND PROSPECTIVE**

... Jove, with Indignation moved,  
At last in Anger swore, he'd rid  
The bawling Hive of Fraud, and did.  
The very Moment it departs,  
And Honesty fills all their Hearts;  
There shews 'em, like the Instructive Tree,  
Those Crimes, which they're ashamed to see  
Which now in Silence they confess,  
By Blushing at their Ugliness;  
Like Children, that would hide their Faults,  
And by their Colour own their Thoughts;  
Imag'ning, when they're look'd upon,  
That others see, what they have done.  

The Bar was silent from that Day;  
For now the willing Debtors pay,  
Even what's by Creditors forgot;  
Who quitted them, who had it not.  
Those, that were in the Wrong, stood mute,  
And dropt the patch'd vexatious Suit.  
On which, since nothing less can thrive,  
Than Lawyers in an honest Hive,  
All, except those, that got enough,  
With Ink-horns by their Sides trooped off.  

Bernard de Mandeville, *The Fable of the Bees* (1705)

The commitment of the legal profession to truth, although often expressed, is likely somewhat deceptive. This Article assumes that

---

society possesses an interest in truth, and that the Anglo-American adversary system purports to serve this interest; but common wisdom supports and economic insight shows that the practicing lawyer has an interest adverse to a reality that is transparent to all. The general economic approach to litigation assumes litigation is preferred over settlement only so long as the parties have different expectations of trial outcome; and a corollary is that a lawyer benefits in added fees by any decrease in the rate at which the likelihood of success is revealed. All members of the legal profession extract rents so long as mutual confusion is maintained, and the skill of the litigator is revealed by his capacity to "shift the odds"—that is, to distort the generic probability of success under the law in a particular set of circumstances. The high-quality advocate could not then gain a premium level of compensation, which he fully maximizes by unleashing the acme of his skill only late and by surprise (i.e., at trial), since his total compensation depends on the length of the dispute, along with reputation.

Common prejudice often errs, however, in identifying the lawyer's interest as lying, for this is not what analysis suggests; a successful lie would lead to a determination of a state of affairs contrary to fact, but it would lead to a determination. A lawyer's actual interest, all else equal, is in the maintenance of indeterminacy of any state of affairs for as long as possible. It is little wonder then—to the cynic—that lawyers have, at

136. This is by no means an uncontroversial assumption, given the strong philosophical and political tradition supposing an excess of truth to be harmful or fatal to social order, as Mandeville exemplifies. So, in fairness, a note of caution is needed regarding the introduction of innovation designed to increase honesty, although any harm done might be limited if honesty is confined to the courtroom, even if this casts its "shadow" over ordinary discourse by creating disincentives for deceit. Apart from concepts such as Plato's noble lie, there is, for instance, Pascal, who wrote: "Man is therefore nothing but disguise, lies, and hypocrisy, both as individuals and with regard to others. They therefore do not want to be told the truth. They avoid telling it to others. And all these tendencies, so remote from justice and reason, are naturally rooted in their heart." BLAISE PASCAL, PENSEES (Honor Levi trans., Oxford Univ. Press 1995) (1670). Nevertheless, I am taking this point as a given, by presuming that honesty has not yet reached superabundant levels in American society, and that if the technology discussed here (or more likely, its distant progeny) actually threatened to create such an oversupply, I have the perhaps naïve faith that government and business would adapt to remedy this problem of excess candor.

137. See STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 405 (2004) (discussing the "divergence of interests" between lawyer and client as a barrier to disclosure of truth).

138. Of course, the lawyer must know the true expected value in order to abandon the case to pursue other more lucrative opportunities if they present themselves. However, this points out the tactical problem for the lawyer in revealing even information beneficial to his side, except at the end, although a client's fears and possibility of settlement (or professional regulation) might force him to reveal it earlier.

139. An interesting question is therefore raised as to whether the lawyer in the end favors false or true outcomes. A false outcome may be less stable, thus offering greater potential for further gains. However, there are disincentives attached to adjudicative falsity, since one or more participants may be blamed for the error (by contrast, it would be very rare for someone external to the dispute to take action against a lawyer for mere inefficiency). Therefore, I must respectfully disagree with Professor
best, reacted unenthusiastically to any innovation that has threatened their guild control of adjudicative fact. One telling point may be that in the United States, where lawyers represent a more important special interest group, there has been little movement to eliminate the civil jury as has been done in the United Kingdom. Moreover, this system has imposed on the “finder of fact,” over the course of time, a requirement of initial ignorance, which serves to lengthen proceedings and maintain a legal monopoly on the transmission of relevant information into the decision making process,\(^{140}\) whatever laudable goals it might have. The legal profession has been similarly lukewarm toward any mechanical method of detecting witness deception that might compete with the advocate’s use of verbal cross-examination, which has the capacity to upset the witness’ self-presentation so as to induce a belief of untrustworthiness\(^{141}\) in the minds of the jury.

One argument often put forward in defense of the status quo is that our evidentiary system is not solely about maximizing truth (or the rate of truth production). This is most obvious with regard to both the constitutional protection afforded to the criminal defendant, along with the other scattered testimonial privileges, and the related asymmetric risk averseness against Type I errors in adjudicating criminal liability. Perhaps more relevant to the current inquiry, the system affords a certain value to the privacy of information, such that the evidentiary value must be worth the loss of privacy.\(^{142}\) For instance, the rules restricting the use of character evidence generally, or those more specifically to do with

---

McGinnis’ characterization of lawyers as “the enemies of the truth”; there are good reasons for them to prefer truthful outcomes if forced to arrive at something, but they are “handicappers” of the truth, perhaps taking both meanings of this term. Cf. McGinnis, supra note 135.

140. The most obvious exception to the law’s cartelization of information is the criminal defendant’s capacity to testify on his own behalf or to represent himself, although this is a relatively uncommon event and generally ends unfavorably.

141. Ideally, these aspersions are intended to be cast on the witness being examined, rather than upon the questioner or his profession.

142. In addition, the various sets of pretrial investigative restrictions in criminal cases also fall into this category. For this reason, although the focus of polygraphs has been, and continues to be, on the finding of a guilty party in a criminal investigation, I do not address this here. Since the standard of proof is lower for an individual to be indicted than for him to be convicted, it would presumably be appropriate at an earlier stage of fMRI accuracy to use the devices for such purposes and make the results of an adversarial test cognizable by a grand jury—all else being equal. All else, however, is not equal. The level of suspicion required for involuntary questioning by fMRI would presumably be at minimum that required for any form of custodial interrogation. The procedure involved in integrating the technology into this area therefore requires separate treatment. Tentatively, however, I should note that there may be a possible difference here between imaging of memory activation and imaging of false statement activation. It is only the latter that requires a statement and that directly implicates the Fifth Amendment. A defendant may exercise “his right to remain silent” while still being shown items, words or pictures to judge his nonverbal reaction. Could his brain be scanned simultaneously, in the way he is now compelled to surrender a DNA sample? I admit this to be a difficult question, but one I think society will inevitably confront, and one well worth exploring in advance of this confrontation.
victims of sexual assault, can be seen in this light, as can Federal Rule Civil Procedure 35, restricting the mental or physical examination of parties, and the use of information so obtained.

These objections carry little weight with regard to the credibility of already proposed witnesses. Such individuals have already either volunteered their testimony or else the court has been sufficiently satisfied with its potential value that they have been subpoenaed into giving evidence. Therefore, the initial hurdle of their privacy has been overcome before the question of impeachment, by neuroimaging or otherwise, need even be broached. It is in getting the evidence of witnesses to be admitted for its substance that procedure hems in the truth-finding process. Once this point is reached, however, witnesses—according to their oath—possess no right to conceal or to lie. In this local component of the trial, the system does appear to seek a maximization of truth, even if this is not quite so about the legal or trial process more globally. Consequently, using neuroimaging as part of the assessment of the crucial question of oath adherence has seemed to me to be an appropriate entry point for the technology's use by the legal system.

The unconvinced might consider the irony of the current primary restriction on cross-examination. Apart from distortion of the evidence being impeached, the cross-examiner is somewhat constrained in the manner of the examination by the desire to “protect witnesses from harassment or undue embarrassment.” At the same time, it is perfectly obvious that the premise of the display of the impeached witness to the jury, and the jury's supposed role as lie detector, means that impeachment serves two goals, both to reveal further substantive evidence and to embarrass the witness (but not unduly) in order elicit “demeanor” reactions from which the jury can (allegedly) assess credibility. One consequence of the current proposal would be to make less salient this latter aspect of the actual impeachment process at trial. Pre-trial impeachment by neuroimaging, whatever its flaws, would be less intrusive—or at least more private in its intrusion—than current practice.