

# BRAIN IMAGES AS LEGAL EVIDENCE

by Walter Sinnott-Armstrong<sup>1,3</sup>, Adina Roskies<sup>1,3</sup>,  
Teneille Brown<sup>2,3</sup>, and Emily Murphy<sup>2,3</sup>

<sup>1</sup> Dartmouth College

<sup>2</sup> Stanford University

<sup>3</sup> MacArthur Law and Neuroscience Project

Brain images are becoming more and more common in courts. Feigenson (2006) found 130 reported opinions involving PET and/or SPECT evidence but only 2 reported opinions citing fMRI evidence. Helen Mayberg, however, has served as an expert witness in over 50 trials in recent years, many of them involving fMRI evidence, and a number of judges have informally told us that evidence from neuroscience, including fMRI, has become standard in capital sentencing.

Some lawyers and neuroscientists, however, are critical of this trend. A few have even suggested in conversation a temporary moratorium on brain images as legal evidence in criminal trials, except possibly in capital sentencing.

This paper will explore the prospects for some uses of data from brain imaging in the courts. Whether brain images should be admitted into trials depends, of course, on how probative they are for specific legal issues and on whether they are likely to mislead fact-finders in trials. We will address these topics in turn after illustrating the variety of uses of brain images in law.

## 1 – What could brain images be legal evidence for?

Brain images could conceivably be used for many different purposes within the legal system. Neuroscientific studies involving structural or functional brain images might, for example, be used to argue for or against certain legislation or prison policies. They might also be used to justify predictions of misbehavior in parole hearings. We will, however, focus on the use of functional brain images in courts.

One proposed use of functional brain images in trials is for mind-reading. When brain scans are used to detect lies or deception, for example, they are supposed to detect whether a person has a mental state of belief in what they say. A few companies (Cephos and NoLieMRI) already offer methods of lie detection using fMRI. Although their results are unconvincing to date, EEG data were admitted as evidence against lying in 2001 by Iowa District Court Judge Tim O'Grady in the case of Terry Harrington. Brain images might also be introduced as evidence of mental states other than deception, such as bias in jurors, consciousness in cases of end of life issues, and pain and suffering in tort plaintiffs or applicants for disability benefits.

Another possibility is to introduce brain images as evidence not of temporary mental states but of more stable mental traits or capacities. Mental capacities might be relevant to competence to stand trial or to be executed or they might be relevant to criminal responsibility if they involve incapacities to gain requisite knowledge or to form requisite intentions that are part of the *mens rea* of most crimes. Neuroscience might also become relevant to whether adolescents and people with brain damage lack substantial capacity to conform their conduct to the law or whether psychopaths, for example, lack substantial capacity to appreciate wrongfulness. Evidence of such capacities might be relevant to an insanity defense, in the guilt phase, or in the sentencing phase of a criminal trial.

This list of potential uses is incomplete, but it gives some sense of the wide range of possible uses of neuroscientific evidence in legal trials. Many of these uses are speculative, and they need not all be treated alike. Each proposed use of neuroscientific evidence needs to be assessed carefully in context and on its own.

For the sake of simplicity, we will focus here on functional brain images used by the defense in a criminal trial to reduce responsibility. These uses might occur in the guilt phase to challenge an element of the crime charged, in an insanity defense, or in the sentencing phase to argue for a lighter sentence.

## 2 – Legal standards of evidence

Whatever legal issue is at stake, to introduce a brain image as evidence in a legal trial, either side needs to meet standards for demonstrative evidence (such as exhibits) as well as standards for scientific expert testimony. The testimony is needed in order to interpret the images during the trial. Although it would also be important to ask when neuroscientists (especially cognitive neuroscientists) meet requirements for expert testimony, we will restrict our discussion here to whether or when functional brain images may be admitted under the rules governing demonstrative evidence.

Most courts follow something like the Federal Rules of Evidence:

FRE 401: “relevant evidence” means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

FRE 403: Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

To apply FRE 403 to brain images in criminal trials, courts must answer three central questions: (1) How probative for criminal responsibility is the brain image? (2) How dangerous (that is, prejudicial, confusing, misleading, or needless) is the brain image? (3) Does its danger substantially outweigh its probative value?

To answer these questions, we need to understand, first, how brain images are constructed. Only then can we determine their probative value and whether they confuse or mislead.

## 3 – What is a brain image?

There are many types of brain images, but usually the term refers to images derived from noninvasive techniques for measuring structural or functional properties of the brain. A number of such techniques exist, including PET, SPECT, MEG, DTI, structural MRI, and functional MRI (fMRI). We will focus here on fMRI, though our main points will apply as well to other functional brain imaging techniques.

The most commonly used fMRI techniques measure changes in the ratio of oxygenated to deoxygenated blood (the BOLD signal). This signal is closely related to blood flow and bears a complicated relation to neural activity. These relations are well-documented although not yet completely understood.

Inferences about brain activity are typically made by designing experiments that contrast the MR signal measured during two different tasks. Ideally, the tasks differ in one respect, and the location and magnitude of the difference in measured signal is attributed to brain activity involved in the difference in task performance. For instance,

one task might involve processes A-E, and another may involve processes A-D but not E. The difference in signal is thus interpreted to be involved in process E. In practice, there are almost always a number of differences among the tasks. With enough psychological sophistication, these can be modeled, although they are not always easily assessed. There are also many minor differences across trials while performing the same task, such as differences in processing individual stimuli; and the signal itself is noisy. When enough stimuli are presented, these minor differences will wash out in the statistics. The difference in MR signal between task conditions is usually quite small, often less than 1%, but with enough data even such small differences can be statistically significant.

The discovered difference in MR signal is often presented as a brain image. These images are usually constructed by superimposing colored pixels on a grey-scale picture of a standard brain in order to indicate where signal was higher (usually red or yellow) or lower (usually blue) than in a contrast state. The resulting fMRI images look something like photographs, but they are *not* photographs. Instead, they are constructions from highly abstract numerical data about magnetic properties. Brain activity and blood flow are not brightly colored, and the brain does not really “light up” when active. It is important to bear in mind that brain images are simply a vivid way to represent the location and magnitude of statistical differences in signal across large data sets. (For more detail, see Roskies 2007, 2008.)

#### 4 – Are brain scans *probative* of criminal responsibility?

The Federal Rules of Evidence do not explicitly define probative value. However, one standard, respected textbook defines probative value as degree of relevance:

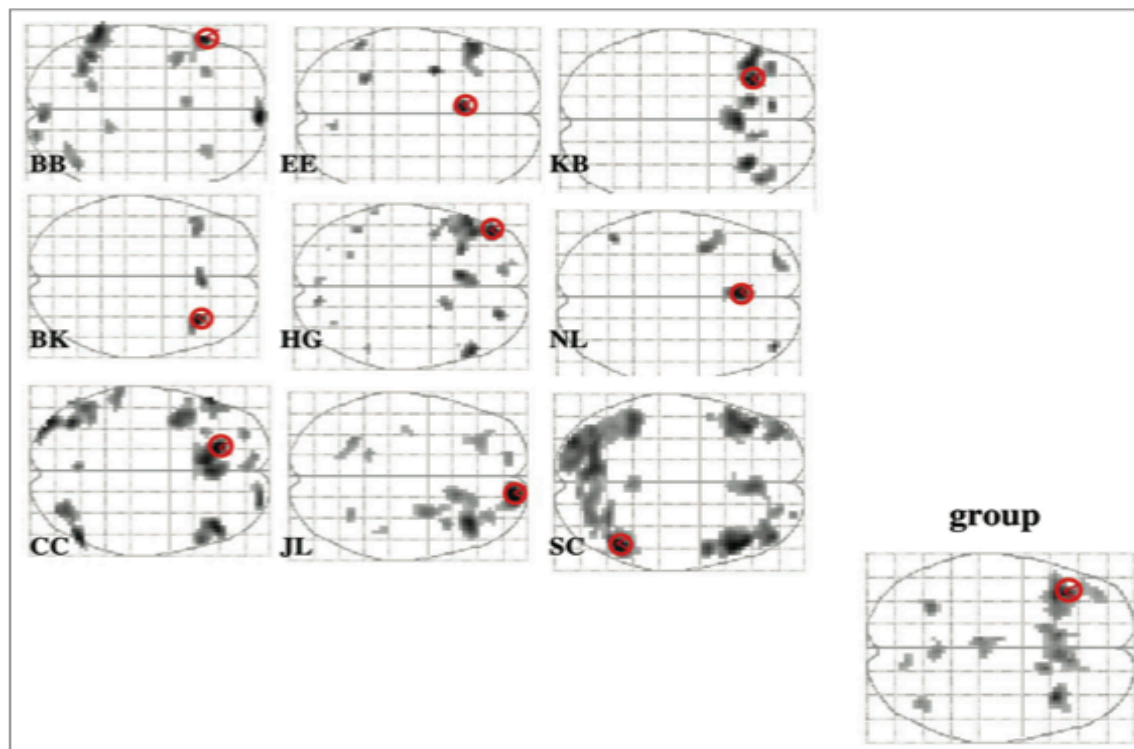
Remember that evidence is “relevant” if it has “any” tendency to make the fact of consequence more or less probable; probative value measures the strength of the effect on the probabilities, even if only in general terms like “highly,” “somewhat,” or “minimally” probative. (Allen *et al.* 2006, 135)

This account makes probative value equivalent to a relative conditional probability. We doubt that the issue is this simple, because values enter into the equation in ways that we will see. Still, a good starting point for assessing the probative value of any evidence is to ask how much the evidence increases the probability of some fact that matters.

To apply this standard to brain images, we need to consider the precise nature of the information that is presented in the image and also which fact the image is supposed to be evidence for. This is an immensely complex topic. Here we can only run through five main problems that arise when trying to use brain images as evidence of facts that are relevant to criminal responsibility.

##### 4.1 – Normality

First, brain images are sometimes offered to show that a particular defendant is abnormal in some way that is claimed to remove or reduce criminal responsibility. The notion of normality, however, is dubious when based on functional brain images. Most neuroscientific studies using fMRI report group averages, but individual functional profiles can vary so much that it is not unusual for most individuals to differ from the group average. This point is made graphically in this diagram from Miller *et al.* 2002:



In this study, subjects SC, JL, and BK look very far from the group average, and subjects EE and NL are pretty far as well. Thus, 5 of 9 subjects whose scans were averaged to create the “normal” group average seem “abnormal.” In cases like these, to say that an individual is abnormal does not mean much at all. This demonstration indicates that the source of any comparison scans must also be scrutinized as critically as the individual scan that is compared to the average or norm.

These different patterns of activation in normal subjects are not a statistical glitch: activations can be fairly stable across time for a given individual, even when that individual varies far from the group average (Miller *et al.* 2002). Different people may simply process the same information differently, even when none is “abnormal” in any way that would be relevant to criminal responsibility. The range of individual variability in functional architecture poses a real problem for using functional scans to determine abnormalities in individuals that would be relevant to legal issues.

#### 4.2 – Base rates and false alarms

Even if we can identify certain functional patterns as abnormal, we still need to determine which individual defendants display that abnormal pattern. The problem here is that functional abnormalities that remove criminal responsibility are likely to be rare. When the base rate is low in this way, even a fairly high specificity (that is, a low rate of false alarms) will yield a high number of false alarms. To illustrate this problem, consider a population of 10,000 with a 1% base rate of a functional abnormality that leads to murder. (Luckily, this is an overestimate.) That means that 100 people in the population have the relevant functional abnormality and 9,900 do not. If an fMRI test for this functional abnormality has 95% specificity, then it will still test positive in 5% of the 9900 who lack that abnormality, which is 495 false alarms — 495 people who test positive but do not really have the relevant functional abnormality. Thus, even if the fMRI test is 100% sensitive, a positive test result still has a predictive value of only  $100/595$ , which is less than 17%.

This low predictive value of a single positive test for a functional abnormality can be improved by additional tests, but only if the defense allows (or the court

requires) those additional tests to be conducted. If the defense gets a positive result that they think supports their claim that the defendant is not fully responsible, then they might be unwilling to subject the defendant to further testing. That will make it difficult in practice to improve upon the low predictive value.

Of course, the predictive value will be greater if the base rate is greater, so defenders of fMRI evidence will point out that the base rate of relevant abnormalities among criminal defendants could be much higher than the base rate in the population as a whole. Bayesian calculations will then yield a higher predictive value for the higher base rate. However, we do not know any base rate for criminal defendants to have a functional abnormality that leads to murder. It is hard to see how we could reasonably guess the base rate, especially since we cannot assume the guilt of defendants in criminal trials. That makes it hard to overcome this problem of false alarms given low base rates.

### 4.3 – Probability of behavior

Even if the defense can prove a functional abnormality in a particular defendant, what matters to law is not brain function but behavior, and abnormal brain function does not necessarily make abnormal behavior likely. Numerous studies reveal that brain damage, particularly in the frontal lobe, can be associated with increased aggressive or antisocial behavior, but the overall prevalence of actual violent crime in these studies is still small. For example, in a large study (including brain scans) of Vietnam Veterans with head injury, 14% of subjects with injury to the frontal lobe engaged in fights or damaged property, compared with about 4% of controls without head injury. Increased aggression was more likely to be present if there was evidence of damage to the medial or orbital areas of the frontal cortex (Grafman *et al.*, 1996). This and other studies establish that damage to the frontal lobes, particularly the medial or orbitofrontal cortex, can lead to abnormal executive function, particularly dyscontrol that could increase the chances for future impulsive behavior or aggression. However, these studies show only an increase above baseline. The increased probability is still low. No study to date has reliably demonstrated a characteristic pattern of frontal lobe dysfunction based on behavioral measures or brain scanning that is predictive of a loss of control or the emergence of violent crime that is applicable to an individual case.

In addition to all of the abnormal brains without violent behavior, there are also, of course, many people with normal brains who commit violent crimes. Thus, we cannot infer the conclusion that a defendant's brain is probably abnormal from the premise that he committed a violent crime any more than we can infer the conclusion that a defendant will probably commit a violent crime from the premise that his brain is abnormal.

### 4.4 – Causation

Even if the defendant has a functional abnormality that is correlated with violent crime, correlation does not prove causation. Functional brain scans reveal only correlations, so we need additional evidence before we can conclude that any functional abnormality played a causal role in the production of criminal behavior.

Causation can reasonably be inferred in rare cases of brain abnormality. Burns & Swerdlow (2003) describe a 40-year old male with little previous use of pornography and no prior sexual deviance. In 2000, he started to use pornography, then child pornography, and eventually he molested his step-daughter. He was arrested, convicted, and required to choose between prison and a twelve-step in-patient program for sex offenders. Inside the program, he propositioned staff members, so he had to be removed. While awaiting sentencing to prison, he experienced headaches, dysgraphia, and loss of coordination, so his brain was scanned, and an egg-sized tumor was found. The tumor was removed, at

which point he lost his symptoms, including his deviant sexual desires. Ten months later, however, he started to collect pornography again, and it was discovered that the tumor had grown back. When a behavior comes and goes with the presence of a tumor in this way, then it is reasonable to infer causation.

What is striking about this case is how unusual it is. Almost no other reported cases explicitly relate sexual deviance to frontal lobe damage. Moreover, we rarely get to observe the behavior come and go with the tumor. When all the evidence we have is functional brain scans, then we do not have enough evidence to infer causation.

#### 4.5 – Control

Finally, even if the defendant has a functional abnormality that causes a violent crime, causation still might not be relevant to the precise legal issue at stake. For example, suppose a defendant asks to introduce a brain image as evidence that he lacked the kind of control that is necessary for criminal responsibility according to one clause in the Model Penal Code version of the insanity defense. To be relevant to this particular legal issue, it is not enough for the brain image to reveal an abnormality that caused the criminal behavior. Instead, that brain abnormality would need to indicate a lack of control that is relevant to this specific defense. Even if a brain abnormality creates a desire that causes the defendant to act, the defendant still might be able to stop himself from acting on that desire. In this case, the defendant would have control over his conduct and, therefore, would meet the pertinent control condition for criminal responsibility.

Some neural abnormalities do remove control. An epileptic seizure, for example, can cause someone's hand to hit someone else in the face, but this case would not even pass a voluntary act requirement. In contrast, other neural abnormalities are compatible with control over conduct that they cause. Consider thrill seekers who jump out of airplanes (with parachutes), jump off bridges (with bungee cords), ski down triple black diamond slopes or drive racing cars. There is some evidence that thrill seekers have a common neural abnormality (in the statistical sense) that creates the desires that cause them to do acts that most other people would never consider doing. Assume that is true. These thrill seekers still do not lack control. They can stop themselves from acting on their impulses when it is dangerous to themselves or others.

Analogously, when neural abnormalities create desires to do illegal acts, people with those abnormalities still might be able to stop themselves. All of us have some desires to do acts that we know we ought not to do, and we stop ourselves. We are excused only when we cannot stop ourselves. However, brain scans alone cannot reveal an inability to control oneself. To show that, we would need to know a great deal more about the nature of control systems and how they interact with systems involved in motivation and desire. We would also need to know retrospectively how control systems were related to the particular defendant's behavior at the time of the crime.

#### 4.6 – Conclusions on probative value

None of these problems show that fMRI brain images cannot ever be relevant to issues of criminal responsibility. Indeed, we can easily imagine circumstances where they would become relevant as markers of some cognitive or behavioral disability. Moreover, as we have said, how probative brain images are will vary with the legal facts that they are supposed to be probative of, as well as with the generation and interpretation of the images and the accuracy of comparisons and inferences.

At present, however, the probative value of brain images for behavioral control as a condition of criminal responsibility seems minimal on Allen *et al.*'s scale of highly/somewhat/minimally probative, because, as we argued, (1) functional normality is dubious in light of individual differences, (2) false alarms are numerous because of low base rates, (3) criminal behavior is unlikely even with functional abnormalities, (4)

correlations cannot show that abnormalities cause particular criminal acts, and (5) even causation by a brain abnormality does not prove any lack of control that would remove criminal responsibility. The probative value of functional brain images for criminal responsibility might be higher for other uses or at some later time, but the problems that we listed will have to be overcome.

## 5 – Are brain images *dangerous*?

According to FRE 403, if brain images are only minimally probative, then they should not be admitted as legal evidence if their probative value is substantially outweighed by any of the dangers listed in FRE 403: "... [1] unfair *prejudice*, [2] *confusion* of the issues, or [3] *misleading* the jury, or [4] by considerations of undue delay, waste of time, or *needless* presentation of cumulative evidence." (our emphasis) Evidence that creates any of these four dangers is sometimes described as prejudicial. This terminology can be confusing, however, because prejudice is itself one of the four dangers. Hence, we will instead use the term "dangerous" to describe evidence that creates any of the listed dangers.

The four specific dangers are not defined in the Federal Rules of Evidence, but Allen *et al.* (2006, pp. 138-141) explain them in this way: Evidence is *prejudicial* to the extent that it leads the jury to decide a case on an improper basis. (The Advisory Committee note to FRE 403 says that this improper basis is "commonly, though not necessarily, an emotional one", such as dislike or pity for the defendant.) Evidence is *confusing* to the extent that it focuses the jury's attention on collateral or inessential issues (such as how the defendant's brain damage occurred). Evidence is *misleading* to the extent that it leads the jury to draw a mistaken inference (such as that the defendant felt an irresistible impulse). Evidence is *needless* if other evidence is easier or cheaper to present and reliable enough by itself.

Do brain images create these dangers? The answer depends, of course, on their particular context and use, so we will continue to focus on uses of functional brain images in criminal trials to reduce responsibility. Even for this particular context and use, the answer is not clear at this time. Some studies do suggest some dangers of brain images, but none of the existing studies is directly on point, because none of them tests the impact of brain images in anything like the environment of a real trial. Still, it is worth running through a few of these studies in order to clarify what would be needed to evaluate the possible dangers of brain images.

### 5.1 – Pictures

In the first study, Bright & Goodman-Delahunty (2006) had subjects read and issue verdicts in fictional criminal cases. Some of their subjects saw no photographs, but other subjects saw either gruesome photographs or neutral photographs, such as photographs of scratches on the door that had been jimmed open. Bright & Goodman-Delahunty found that the conviction rate with neutral photographs (38%) was almost as high as with gruesome photographs (41%) and much higher than with no photographs (8.8%), even though the neutral photographs added no new information that would have justified conviction. This result suggests that photographs as a form of presentation might influence jurors more than the content of the photographs, such as whether the photographs were gruesome or neutral. The photographs in this study had nothing to do with the brain, but this study raises suspicions about whether brain images might have similar effects if they are confused with photographs. One of us has suggested that people's misapprehension

of brain images as photographs could affect the epistemic status they attribute to brain images as a source of evidence (Roskies, 2007; cf. also Dumit 2004).

## 5.2 – Neurobabble

A second study tested the effects of information about the brain without accompanying visual images. Weisberg *et al.* (2008) had subjects assess good and bad explanations with and without neurobabble (our term for nonsensical or irrelevant brain information). In one example, subjects read this:

Researchers created a list of facts that about 50% of people knew. Subjects in this experiment read the list of facts and had to say which ones they knew. Then they had to judge what percentage of other people would know those facts. Researchers found that the subjects responded differently about other people's knowledge of a fact when the subjects themselves knew that fact. If the subjects did know a fact, they said that an inaccurately large percentage of others would know it, too. For example, if a subject already knew that Hartford was the capital of Connecticut, that subject might say that 80% of people would know this, even though the correct answer is 50%. The researchers call this finding "the curse of knowledge."

Subjects then read explanations of this reported phenomenon. Some explanations included no neuroscience, like this one:

The researchers claim this curse happens because subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.

Other subjects received the same explanation with added neurobabble (italicized here):

*Brain scans indicate that this curse happens because of the frontal lobe brain circuitry known to be involved in self-knowledge.* Subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.

Both of these explanations, like the other "bad" explanations in the study, were supposed to be "circular restatements of the phenomenon, hence, not explanatory" (471). Crucially, the added neurobabble does not make the explanation any better. Nonetheless, Weisberg *et al.* found that bad explanations were rated more "satisfying" when accompanied by irrelevant neurobabble (means = 0.16 and 0.2) than without neurobabble (means = -0.73 and -1.1) in their first two experiments (with novices and neuroscience students, respectively). This finding suggests that irrelevant and nonsensical neurobabble confuses and misleads in the senses defined above.

Of course, this does not show that accurate and useful neuroscience confuses or misleads. Moreover, this study is not about brain images, since only words about neuroscience were added (though sometimes the words referred to brain images). Still, this study raises suspicions that jurors might be confused and misled by expert witnesses who add irrelevant neuroscience to their testimony.

## 5.3 – Brain images

The next study is more directly relevant to our topic because it adds neural information and brain images. McCabe and Castel (2008) had subjects read articles with bad arguments, such as “Watching TV helps with math ability because both activate the temporal lobe.” These articles included either brain images, bar graphs, or neither. McCabe and Castel found that subjects rated the articles as making more sense when accompanied by brain images (2.9) than when accompanied by only a bar graph (2.7) or by neither (2.7). This effect might not be large, and it does not involve legal cases, but it is statistically significant, and it suggests that brain images can confuse and mislead people about the value of arguments. This suggestion receives further support from McCabe and Castel’s second study, which found similar results with brain images compared to topographical maps of brain activation that were just as visually complex as the brain images but did not look like pictures of brains.

#### 5.4 – NGRI verdicts

The final study discussed here introduces brain images in legal cases. Gurley and Marcus (2008) asked subjects to read about a violent crime and then decide whether the defendant should be found not guilty by reason of insanity (NGRI). Some subjects read about expert testimony that defendant had a psychosis, whereas others read about expert testimony that the defendant had psychopathy. Some subjects read expert testimony about traumatic brain injury, but others did not. Some subjects were shown brain images suggesting damage in the frontal lobes, whereas others were not shown any brain images. Gurley and Marcus found that the percentage of subjects who found the defendant NGRI after reading expert testimony on mental disorder (psychopathy / psychosis) was higher when accompanied by a brain image (19 / 37%), by testimony about traumatic brain injury (27 / 43%), or by both (44 / 50%) than when subjects received neither (11 / 22%). Thus, the introduction of both testimony about traumatic brain injury and images of brain damage increased the NGRI rate from 11% to 44% in the case of psychopathy. That is a big effect, so brain images and neuroscience do seem to affect legal decisions.

#### 5.5 – Problems

Not so quick! Although these studies are suggestive, they are hardly conclusive. All of these studies face several problems before we can infer anything about the dangers listed in FRE 403. First, brain scans might inform and increase accuracy instead of misleading. After all, nothing in the Gurley and Marcus study shows which rate of NGRI is correct. In that study, brain images and testimony increased the rate of NGRI, but that is *good* if the defendant really deserved to be found NGRI. Evidence is not confusing or misleading if it increases accuracy.

A second problem with concluding that brain scans are dangerous is that the above studies test only certain kinds of people in certain kinds of circumstances. Lab subjects differ from real jurors in important ways: Real jurors are not all college students. Real jurors are not located in labs or classroom. Real jurors hear more details of each case. Real jurors know that real lives are affected. Real jurors hear both sides of an argument including the cross-examination. Real jurors deliberate and know that they are accountable to other jurors insofar as they will probably be asked to give reasons for their beliefs and decisions. Because of such differences, we cannot quickly draw precise conclusions about real jurors from studies about lab subjects.

Some studies (Bornstein 1999) have found that decisions by student and non-student mock jurors did not differ significantly. Still, we cannot know whether the

special circumstances of real jurors reduce or nullify the effects of brain images in particular until we study subjects in circumstances more like those of real juries.

### **5.6 – An attempted solution**

We will probably never be able to obtain useful data from real jurors in real trials. Still, we can try to make the circumstances of actual jurors and experimental subjects as close as possible in respects that are likely to affect the issue at hand. In particular, it would be better to draw subjects from juror pools (rather than just students), re-enact real cross-examination of experts or closing arguments along with jury instructions in detail (rather than short summaries), measure inclinations during the re-enactment (to see how each bit of evidence makes jurors lean one way or the other), ask for verdicts before *and* after deliberation (to see how much deliberation affects verdicts), and ask not just for verdicts but also for knowledge and memory of relevant facts (to see whether images are confusing or misleading by distracting jurors from crucial facts). In this setup, we can compare verdicts and correct answers by subjects who view the re-enactment (a) with brain images as well as expert testimony, (b) with neuroscientific information presented by experts but without brain images, and (c) with neither brain images nor any other neuroscientific information.

This design still has flaws. Subjects still know that no real lives or freedom hangs on their decisions. (It is hard to imagine a solution to this problem that would get past human subjects review committees.) We will also not be able to reproduce all of the details during a real trial, especially lengthy cross-examination, because not enough subjects will volunteer for that long. Despite these problems, our proposed study will, we hope, give more and better information about how juries react to brain images.

## **6 – Conclusions**

Unfortunately, we do not have any data from our study yet. Hence, our conclusions must all be tentative and conditional. Still, we can draw three kinds of conclusions.

### **6.1 – Scientific conclusions**

We argued above that current fMRI brain images are only minimally probative of control as a condition of criminal responsibility because of issues inherent in both the technology and in the assessment of the brain basis of control capacity. Of course, brain images might be more probative of other legal issues, and their probative value might increase as techniques improve. Advances on the horizon include pattern classifiers and new ways to base individual predictions on group data given knowledge of individual differences (cf. Haynes and Rees 2006).

On the other side of the scale, we do not yet know whether or how much brain images are confusing or misleading to jurors. There is some reason to suspect that they are, but the simple studies reported so far do not show whether more detail along with cross-examination will undermine any tendency of brain images to confuse and mislead. We need better studies to determine how dangerous brain images are in various circumstances.

### **6.2 – Legal conclusions**

Because of the scientific uncertainties, our legal conclusion has to be conditional: If brain images are as confusing and misleading in trial contexts as they seem to be in reported experiments, and if they lack much probative value because

they cannot overcome the problems listed above, then their moderate dangers “substantially outweigh” their minimal probative value, so brain images fail the balancing test of FRE 403 and should not be admitted into trials. That is a lot of “if”, but it is all that we can conclude for now.

It is also worth recalling that brain scans still might be admissible in some situations, such as capital sentencing, where the defense is strongly favored and the admissibility standards are much more lax. This is where they seem to be most regularly admitted now, and this practice seems to rest on the traditional value judgment that it is particularly horrible to find someone guilty who is innocent or who has not been given every reasonable chance to defend himself or herself.

### **6.3 – Philosophical conclusions**

Although we have not emphasized these issues, we hope our discussion has made it clear that what counts as evidence and whether evidence is strong enough for beliefs and decisions to be justified depends not just on pure probabilities but also on values, including costs of errors. In addition, gathering and assessing legal evidence from neuroscience is a social task that requires many people from different fields: neuroscientists and lawyers. One lesson for philosophers, then, is that epistemology needs to become both normative and social if it is to become applicable to the ways in which real people go about gaining knowledge, such as in our legal system.

## **ACKNOWLEDGEMENTS**

This essay benefited from enlightening discussions with Scott Grafton, Mike Gazzaniga, and Suzanne Gazzaniga. This material is based on work supported by the John D. and Catherine T. MacArthur Foundation and The Regents of the University of California. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John D. and Catherine T. MacArthur Foundation or of The Regents of the University of California.

## REFERENCES

- Allen, R. J., Kuhns, R. B., Swift, E., Schwartz, D. S. 2006. *Evidence: Texts, Problems, and Cases, 4th Ed.* New York: Aspen.
- Bornstein, B. H. 1999. The ecological validity of jury simulations: is the jury still out? *Law and Human Behavior*, 23, pp. 75-91.
- Bright, D. A. & Goodman-Delahunty, J. 2006. Gruesome evidence and emotion: Anger, blame, and jury decision-making. *Law Hum Behav*, 30, pp. 183-202.
- Brower, M. C., & Price, B. H. 2001. Neuropsychiatry of frontal lobe dysfunction in violent and criminal behavior: A critical review. *J Neurol Neurosurg Psychiatry*, 71, pp. 720-726.
- Bufkin, J. L., & Luttrell, V. R. 2005. Neuroimaging studies of aggressive and violent behavior. Current findings and implications for criminology and criminal justice. *Trauma, Violence and Abuse*, 6, pp. 176-191.
- Burns, J. M., & Swerdlow, R. H. 2003. Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign. *Arch Neurol*, 60, pp. 437-440.
- Dumit, J. 2004. *Picturing Personhood: Brain Scans and Biomedical Identity*. Princeton: Princeton University Press.
- Feigenson, Neal. 2006. Brain imaging and courtroom evidence: on the admissibility and persuasiveness of fMRI. *International Journal of Law in Context*, 2, pp. 233-255.
- Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., & Salazar, A. M. 1996. Frontal lobe injuries, violence, and aggression: A report of the Vietnam head injury study. *Neurology*, 46, pp. 1231-1238.
- Gurley, J. R., & Marcus, D. K. 2008. The effects of neuroimaging and brain injury on insanity defenses. *Behavioral Sciences and the Law*, 26, pp. 85-97.
- Haynes, J-D., and Rees, G. 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7, pp. 523-34.
- McCabe, D. P., & Castel, A. D. 2008. Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, 107, pp. 343-352.
- Miller, M., Van Horn, J. D., Wolford, G., Handy, T. C., Valsangkar-Smyth, M., Inati, S., Grafton, S., and Gazzaniga, M. 2002. Extensive individual differences in brain activations with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, 14, pp. 1200-1214.
- Roskies, A. L. 2007. Are neuroimages like photographs of the brain? *Philosophy of Science*, 74, pp. 860-872.
- Roskies, A. L. 2008. Neuroimaging and inferential distance. *Neuroethics*, 1, pp. 19-30.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. 2008. *Journal of Cognitive Neuroscience* 20, pp. 470-477.