Neuroimaging in the Courts of Law

Bernard Baertschi

University of Geneva, Switzerland

Abstract

Lie detection has recently become a topic of discussion once more. Courts of law were interested in it for a long time, but the unreliability of the polygraph prevented any serious use of it. Now a new technology of mind-reading has been developed, using different devices that are deemed to be able to detect deception, in particular Functional Magnetic Resonance Imaging (fMRI). Is fMRI more reliable than the polygraph? It meets at least with various kinds of obstacles: technical, methodological, conceptual and legal. Technical obstacles are linked with the state of the technique, methodological ones with epistemological difficulties, conceptual ones with problems tied to what lying consists of, and legal ones with the effects of brain imaging on lawsuits. I examine several of these and conclude that at present mind-reading using fMRI is not ready for use in the courts. The obstacles examined may not be insuperable, but a lot more research is needed.

Keywords: brain imaging, mind-reading, lie detection, courts

1. Introduction

For a long time, human beings have tried to decipher the mental states of their fellows without relying on what they say or might say about themselves. There is very good reason for this: in particular, if we could read the minds of other people, some drawbacks in our social and moral life could be avoided. Importantly, we could detect liars and cheats, a very crucial matter to assess others and to improve human cooperation (Cosmides and Tooby, 2008). In this context, it is not surprising that judges are interested in mind reading. In the United States, the polygraph has been used (and is still used) in particular States, but it has never been acknowledged as a reliable tool. However, for some time now a new tool has been available – neuroimaging. Is it more reliable? This is the question that I will address in my paper.

Several devices have been developed to look non-invasively inside the human body beyond traditional X-rays. They are often referred to by the generic term 'scanners'. The more important are Electroencephalography (EEG), Positron emission tomography (PET) and Magnetic resonance imaging (MRI). Physicians are using them for a lot of medical purposes, but as scanners can also 'look' under the skull they can be used to look into the brain and, through it, into the mind. Functional magnetic resonance imaging (fMRI) is now widely used by neuroscientists and psychologists to study the human mind.

In the courts, two subjects are linked to brain imaging; the detection of liars and the insanity defence. I give 'insanity defence' a wide meaning: it includes every claim to mitigate responsibility on the basis of some structural or functional brain abnormality. It is a wider subject than the first, but, in this paper, I will speak exclusively of the detection of liars, for two reasons. First, it has been hotly debated recently in the United States, and on many grounds we can expect that forensic use of scanners will spread in other countries. Second, the questions raised by the use of neuroimaging for detecting liars are relevant for the general reliability of brain imaging in neuroscience.

In the United States, some companies already offer lie detection through EEG (Brain Fingerprinting) or fMRI (Cephos, No Lie MRI). Walter Sinnott-Armstrong and colleagues claim that 'EEG data were admitted as evidence against lying in 2001 by Iowa District Court Judge Tim O'Grady in the case of Terry Harrington' (2008, 360). Terry Harrington was convicted of murder in 1978, and freed in 2001. But if the EEG technique was accepted by the court, contrary to what is very often said, it was not tested on this occasion, because while the district court has been ordered to make a new trial for reasons unrelated to lie detection, it gave up: 'The local prosecutors declined to pursue the case and Harrington was freed' (Wolpe et al., 2005, 43). In the last few years, it has been successfully introduced several times in suits, but more recent attempts, in 2010, have failed (Saenz, 2010). In India, the technology is now being considered for use (Racine, 2010, 3).

How is lying detected by these devices? Two techniques are used: the Control question test (CQT) and the Guilty knowledge test (GKT). As Paul Wolpe and colleagues explain (2005, 40), in CQT the subject is asked to answer three kinds of yes-no questions: relevant questions (e.g. 'Did you kill your wife?'), control questions (e.g. 'Did you ever steal something?') and irrelevant questions (e.g. 'Are you sitting in a chair?'). Everybody is expected to react more strongly to control than to relevant questions, except the culprit. Therefore, if the defendant denies having killed his wife but reacts more strongly to relevant questions than to control ones, then it is a clue that he is lying.

In GKT too, a series of questions are asked, some relevant, some irrelevant or neutral relative to knowledge that only the guilty person could possess concerning the place, time and details of the crime. 'For example, in a crime investigation involving a stolen red car, a sequence of questions could be: "Was the car yellow? Was the car red? Was the car green?" The questions are chosen so that subjects with knowledge of the crime (but not other individuals) would have an amplified physiological response to the relevant question – that the car was red – which is dubbed "guilty knowledge"' (Wolpe et al., 2005, 40). GKT was the technique used in Terry Harrington's case.

As we can see, lie detection is more direct with CQT, but both techniques can be used to test the truth. The underlying hypothesis is that brain activity changes in a controlled manner when somebody tells a lie. Is this true? To answer this question, it is necessary to review thoroughly the tools and techniques used.

2. Some Technical Obstacles

I will limit my investigation to fMRI, because it is the device that has been mostly scrutinised in peer-review papers. In order to use fMRI, be it for experimental or for forensic purposes, the task or procedure must first be calibrated to be adapted to the subject or the defendant. The subject can then engage in the mental task under investigation (e.g. lying). This still cannot be recorded directly: as we have seen, such a measurement always rests on a comparison of two tasks (or more), as Eric Racine and colleagues explain:

> An assessment of brain activity during the mental process of interest then resides in statistical comparisons across the entire brain (separated into cubic elements or voxels), of the signal level between the control task and the experimental task. Those brain regions, or voxels, which meet statistical significance at a set threshold level, differentiate areas where the signal is statistically greater in the active versus control task. These regions are labeled in a colour coded map to represent active brain areas, and are generally overlaid onto a structural image. (2010, 246)

Briefly said, coloured spots in brain pictures represent variations in blood oxygen, variations that are recorded only if they are above a threshold and characteristic for the experimental task investigated.

As we can see, using fMRI is not an easy matter: brain imaging has nothing to do with photographs. Therefore, it is not surprising that the technique encounters some specific obstacles. I will examine three of them: (i) replicability of results, (ii) BOLD factors and cognition, and (iii) method of subtraction. There exist other technical obstacles, i.e., obstacles arising from the limitations of the technique used, the most well-known being that brain activity is much faster that fMRI measurements, but I will not dwell on them.

By *replicability of results*, I do not refer to the applicability of laboratory experiments to real-world scenarios. This applicability is a real problem, but I will ponder on it when I will attend to legal obstacles. What I mean by replicability of results is the problem created by what has been called 'uniqueness of neural signature' (VanMeter, 2010, 237). Every brain has its singularity and is different from other brains; consequently, there exists an individual variability that could prevent the acquisition of universal inductive knowledge on lies applicable to all individuals. As we can see, this obstacle is technical *and* methodological or epistemological; therefore it is not certain that technical progress will be able to overcome it.

This variability is individual *and* general: we have observed some structural differences between the brain of psychopaths and of normal persons; the effect of these differences could be that the neural signature of lies is not the same in a psychopath and in a normal person.

The problem of replicability is particularly important for brain imaging, because fMRI pictures representative of a certain mental task and used as standards report not individual brains, but group-average brains. Sinnott-Armstrong and colleagues notice: 'Individual functional profiles can vary so much that it is not unusual for most individuals to differ from the group average' (2008, 362). It means that it will sometimes happen that the brain image of a particular liar will be sufficiently different from the 'typical' brain image of a liar, that he will be considered as a truth teller. The reverse is possible, too.

MRI measures *BOLD factors* – i.e., blood oxygenation level dependent factors – in short, the amount of oxygen present in blood. But, as Valerie Hardcastle and Matthew Stewart say: BOLD factors are not the same thing as cognition (2009, 187); it is not even identical with brain activity. Between BOLD factors and cognition, there are two intermediate levels; brain activity and coloured voxels (the 3D pixels visible on brain pictures). To pass from one level to another requires interpretation – some authors even speak of 'manipulation' –, and interpretations can lead us astray.

Interpretation of data is difficult, but obtaining relevant data is not easier. To gather the data, we must first set the threshold of the MRI device in order to avoid false positives *and* obtain useful results. If the threshold is too low, we are overwhelmed by a wealth of uninterpretable voxels or false positives, like the dead salmon perceiving human feelings! Bennett put a dead salmon in an MRI machine and observed some activity characteristic for the perception of emotions in third parties in the fish's brain, although it is absurd to claim that a dead salmon can perceive anything. However the problem cannot be solved simply in setting the threshold much higher, because, as Bennett points out: 'We could set our threshold so high that we have no false positives, but we have no legitimate results' (Madrigal, 2009.2, 2). A threshold must be determined, but there is no easy way to do it.

When the threshold is set, the usual way to obtain relevant data is the *method of subtraction*. As we already know, the subject performs two tasks, the data obtained are subtracted and the remaining data is specific for the task the researcher wants to investigate. Marcus Raichle illustrates this method with an example: 'For example, to "isolate" areas of the brain concerned with reading words aloud, one might select as the control task passively viewing words. Having eliminated areas of the brain concerned with visual word perception, the resulting 'difference image' would contain only those areas concerned with reading aloud' (2009, 5).

Comparing results is ubiquitous in brain imaging. We compare because we are unable to observe directly. But it is not without dangers, as Adina Roskies states: 'The very same raw data from the main task can give rise to a very different image depending on which comparison task is employed in the analysis, and very different tasks can give rise to very similar images if appropriate comparison tasks are chosen. Neuroimaging, unlike photography, is essentially contrastive' (Roskies, 2007, 870). The result of a subtraction depends on the numbers used in the calculation; if one number changes, the result will of course not be the same.

3. Some Methodological Obstacles

Let us imagine that the technical obstacles are overridden, thanks to technological progress (a super-fMRI is available), and suppose that we observe constant correlations between brain patterns and mental tasks, delivering replicable results. Would all the obstacles be suppressed and could the judges confidently rely on brain imaging? Unfortunately, it will not be the case, because there will still be obstacles of a more pervasive nature. These obstacles are methodological or epistemological and three of them are prominent in my mind: (i) correlation and explanation, (ii) correlation and causation, (iii) reverse inferences.

Correlation is not explanation. From the fact that *A* correlates with *B*, it does not follow that *A* explains *B*. We observe correlations between brain events (BOLD factors) and mental events, and even constant correlations. Therefore we are tempted to conclude that brain events explain mental events. For instance, from the fact that amygdala activation is strongly correlated with anger and fear, we jump to the conclusion that fear and anger are explained by amygdala activity. This conclusion is too hasty for two reasons. First, it draws on a metaphysical thesis (like brain/mind identity, epiphenomenalism or supervenience of mind on brain) without arguing in its favour. Second, and more importantly – because we have in fact a lot of independent arguments in favour of some materialist metaphysical thesis (Kim, 1998) –, other neuroscientific explanations are possible, as Ellis states:

> It remained for a few maverick neuroscientists, such as Panksepp, to keep insisting that amygdala activation was not the substrate of anger, but instead might correlate with anger only because it played an important role in learning and remembering which stimuli should elicit anger. Panksepp's view is that the periaqueductal grey area deep in the subcortex is the most crucial part of a complex anger circuit involving many brain areas other than the amygdala. (2010, 69)

Explanations need more than correlations.

Correlation is not causation. We have known for long time that succession is not causation (*pace* Hume); constant succession is a kind of correlation, i.e., a relation that is not law-like, contrary to causality.

Therefore, if constant succession is not causation, it is because correlation is not causation. It is easy to show why. If *A* correlates with *B*, it does not follow that *A* causes *B*. *A* and *B* can for instance be the joint effect of *C*. The hypothesis of Panksepp could illustrate this argument, too. In his view, amygdala activation is not the cause of fear, but both are something like joint effects of events taking place in the periaqueductal grey area deep in the subcortex.

Technical obstacles did show that we do not observe a bi-univocal correspondence (a one-to-one relation) between types of brain events or brain pictures and types of mental events. The methodological obstacles examined so far add epistemological reasons to the technical ones. It does not mean that it is logically impossible to draw a bi-univocal correspondence between both types of events: when we discover laws in nature we sometimes succeed at that precisely; but for now we are unable to ascertain that it will be possible for the brain. The difficulty is made worse by the fact that mental events are private and therefore hidden; consequently, to know them, we are obliged to make *reverse inferences*. What does it mean? In establishing correlations between the mental and the cerebral, we give the subject some mental task to perform and we observe the correlated activation pattern through fMRI. When we want to know which mental task another subject is performing, we must predict it on the basis on the brain pattern observed. This prediction consists in a reverse inference (inference from brain to mind, grounded on inferences from mind to brain), but such an inference is not reliable as a bi-univocal correspondence between brain and mind has not been established. Kamila Sip and colleagues apply it to lie detection:

> When using functional magnetic resonance imaging (fMRI) (or any other physiological measurement) to detect deception, we are confronted with the problem of making reverse inferences about cognitive processes from patterns of brain activity. Even if deception reliably activates a certain brain region, we cannot logically conclude that, if that brain region is activated, deception is taking place. (2007, 50)

4. Some Conceptual Obstacles

The technical and methodological obstacles I have mentioned and examined concern brain imaging using fMRI in general, not only its forensic use. If neuroscientists and neuropsychologists are confronted with these difficulties, judges and lawyers will be, too. In a sense, the problem is greater in the courts, because lawyers are not so well aware of the scope and limits of the technology. In another sense, however, the problem is perhaps more manageable: in the courts, we are interested in lies and in nothing else, a very narrow topic. Consequently it would not be necessary to identify the correct brain explanation of lying or its real cerebral cause to detect lies: correlations like those observed in CQT and GKT tests could suffice. Would it not be possible then to reach an agreement concerning the neural signature of lies, even if in certain respects this signature is manifold?

To reach an affirmative answer to this question, two conceptual difficulties must still be solved concerning: (i) the nature of lying, and (ii) the importance of intentional context.

We wish to detect liars, but what does it mean, to lie? What is *the nature of lying*? At first sight, the answer seems unproblematic: to lie is to conceal the truth. But there are many ways to conceal the truth. What is the difference between a spontaneous and a prepared lie? A temptation to lie and an effective lie? An exaggeration and an omission? Another precision is in order: to detect someone as a liar presupposes that the liar knows that he is lying; if he does not know, his brain will not show the right activation pattern. What to do then with self-denial and self-deception? Often, an ingrained liar is persuaded that he is sincere, and if he thinks he is sincere, is he lying? As for everyday life, for the law too there exist several kinds of lies. Besides, some lies are innocent lies and an important ingredient in our life. Marcel Proust said that without lies, life would be impossible to live (Proust, 1999, 2063); for instance, we frequently lie in order to protect the peace of mind of ourselves and of our family, and to preserve good relationships with our fellows. What does neuroimaging have to say concerning this diversity of lies? Nothing at all, says Justice Jed Rakoff:

> A little white lie is altogether different, in the eyes of the law and of common sense, from an intentional scheme to defraud. Nothing in the brain-scan approach to lie detection even attempts to make such distinctions. And what might a brain scan be predicted to show in the case of a lie by omission […]? In my experience, these are the most common kinds of lies in court. (2009, 44-45)

This presents a very bad situation: how can we detect liars if we do not know what we are looking for? Such a problem is not linked with the technique used or with the methodology, it is conceptual or semantic: we must know what it means to lie if we want to test lying with fMRI or any other tool. Nevertheless, neuroscientists and neuroethicists are aware of the problem, and they have proposed a precise concept of lie to be tested. Kamila Sip and colleagues summarise it in the following way: ‘A useful characterisation is provided by Vrij, who defined deception as follows: "A deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue"' (2007, 48). In their comment, the authors underline that two things are important in this definition. First, the crucial point is not the truth value of what is said, but the intentional and deliberate attempt to deceive (if it not the case that *X* and I believe that *X*, then I do not lie when I say that *X*; I just make a mistake – if my interlocutor listens badly to what I say and understands not-*X*, I do not lie either). This point was established long ago by Augustine in his seminal treatise on lying (*De Mendacio*). Second, the liar is not instructed to lie (the lie occurs 'without forewarning'), but he decides freely and voluntary to deceive.

This last point prompts a new methodological problem because, in the vast majority of studies, subjects are instructed to lie, therefore they know that they ought to lie. Consequently, what fMRI detects has little to do with authentic lies, as Nancy Kanwisher states: 'Making a false response when you are instructed to do so isn't a lie, and it's not deception. It's simply doing what you are told' (2009, 12). The methodological problem is the following: to fulfil a task, the subject must know what this task consists in, therefore he must be instructed to lie if lying should be measured. But to be instructed to lie is not to lie. Consequently it is impossible to test authentic lies with fMRI. Nevertheless, the situation is different in the courts: defendants are not instructed to lie, therefore it seems that this methodological problem could be bypassed. Unfortunately, this is not the case, because lie detection standards must be established before this technique could be employed in the courts, and these standards must be set in experiments, i.e., in situations where authentic lies cannot be tested.

Perhaps it would be possible to take an indirect path. When someone is sincere or lies intentionally, it necessarily has an effect on his cognitive and emotional states. In particular, it is more difficult to lie than to tell the truth, because the liar must make an effort to invent some falsehood, and even some plausible falsehood. In consequence he will be more stressed, and sometimes embarrassed or anxious. The polygraph was invented to measure the physiological correlates of those states; in a parallel manner, fMRI could be used to measure their brain correlates. As Victoria Holderied-Milis says: 'Even when liars manage to stay undetected, their integrity is damaged nevertheless. Because of the difference between what they hold to be true and what they articulate, they endure an internal tension, which also has to be hidden from their fellows' (2010, 110-1). Many psychologists agree and think that the tension evinced in lying denotes an inhibition on telling the truth. Now, it is possible to test inhibitions with fMRI, in particular with the help of the Sternberg Proactive Interference Paradigm. Elizabeth Phelps explains it in the following way:

> In a typical version of this paradigm, a participant is shown a set of stimuli and told to remember it. For example, the set might include three letters, such as B, D, F. After a short delay the participant is presented a letter and asked, 'Was this letter in the target set?' If the letter is D, the participant should answer 'yes.' In the next trial the participant is given another target set, such as K, E, H. At this point, if the participant is shown the letter P, she or he should say 'no.' If the participant is shown the letter B, the correct answer is also 'no.' However, for most participants it will take longer to correctly respond 'no' to B than P. This is because B was a member of the immediately preceding target set (B, D, F), but it is not a member of the current target set (K, E, H). […] To correctly respond 'no' to B on the current trial requires the participant to inhibit this potential 'yes' response and focus only on the current target set. (2009, 19)

Brain imaging has shown that the inferior frontal gyrus plays an important role in this type of inhibition, and even if the above experiment has nothing to do with lying, but with our access to truth, it is thought that this region could be involved in lying.

Consequently, there seems to be a possible way out of this new methodological problem, if we adopt the following psychological thesis: lying requires inhibition of telling the truth and this inhibition can be correctly detected. The inhibition hypothesis looks plausible, but is it true? In other words, can we confidently say that all or most intentional lies are linked with such an inhibition? If that is the case, it would mean that the liar would want spontaneously to tell the truth, but he blocks this reaction, which requires an effort. Unfortunately, a recent study casts some doubts on this hypothesis.

Joshua Greene and Joseph Paxton have conducted a study on moral decision. Subjects were instructed to predict the outcomes of coin-flips. When their predictions were correct, they gained some money, but when they were not correct they were financially punished. Subjects made their prediction privately and checked themselves the result, without any supervision. Therefore it was possible for them to report falsely, i.e., to cheat. As the coin-flips were randomly made by a computer, the probability of an accurate prediction was 0.5. Therefore all subjects who claimed that they had been able to predict correctly the outcomes of coin-flips with a probability higher than 0.7 made a false report (i.e., they lied) on several occasions.

Did they have to refrain from telling the truth, i.e., from reporting the correct result each time they did not? No. The additional brain activity, deemed relevant for deception, was actually observed in fMRI, but especially in dishonest subjects (i.e. subjects who claimed that they had been able to predict with a probability higher than 0.7) who *renounced* a dishonest gain, that is, those who made an honest report of their failure to predict accurately the outcome of the flip.

Therefore, even if lying implies an effort not to tell the truth for occasional liars, telling the truth, that is not to lie, implies a more important effort for usual liars. The authors conclude:

> We find that control network activity is most robustly associated, not with lying per se, but with the limited honesty of individuals who are willing to lie in the present context. It is unlikely that control network activity associated with limited honesty is related to overcoming a default honesty response because such responses are themselves honest. (2009, 12509)

This study forces us to conclude that lying is not regularly or *per se* correlated with an inhibition on telling the truth.

This study is particularly important for the reason that the courts are frequently confronted with dishonest individuals. This drives us to *the importance of the intentional context*, the second conceptual obstacle. Discriminating between the brain reactions of honest and dishonest people already takes intentional context into account. This context is wider, however, and consists in the desires, beliefs and mental attitudes of the individuals who are lying. Kamila Sip and colleagues underline this situation with the striking example of psychopaths: 'The lack of emotional response observed in psychopaths, when they tell a lie, stems from their attitude to the victim of the deception. They have no empathy for the potential suffering that their actions might cause.' (2007, 52). This lack of emotional response is the psychological counterpart of the structural brain differences I have already alluded to, so that the neural signature of lies should be different in a psychopath and in a normal person. This is an extreme example, but even in the life of normal individuals, the intentional context has obviously an impact on the way we behave, verbally and non-verbally. A law court is a peculiar place, generating its own intentional context (think of a defendant who tries to exonerate oneself). With this remark, we come to the legal obstacles.

5. Some Legal Obstacles

Beyond the use of a common device, fMRI, and the several common obstacles it encounters, the domain where lie detection takes place (i.e. the court) generates its own problems: the domain of experimentation is not identical with the domain of trials. Trials obey specific rules: legal and procedural ones. These rules create some new obstacles for the detection of liars by means of technological devices like fMRI. I will examine four of them: (i) a trial is not an experiment, (ii) effects on juries, (iii) the role of juries, and (iv) the importance of behaviour.

A trial is not an experiment, therefore lying in a trial is not the same thing as lying in an experiment. More generally, it is doubtful that the results obtained in experiments are transferable to a litigious environment. There are two reasons for that. First, as we have already seen, lies in an experimental setting are either not authentic lies or no lies at all. Second, even if they were authentic intentional lies, the emotional, intentional and institutional contexts are so different that the brain images will be unreliable. The persons investigated are diverse, too; students on one side, defendants on the other.

The difference in context has another impact. Researchers claim that they have been able to identify lies with an accuracy ranging from 76 to 90% (Madrigal, 2009.1, 2). Commercial enterprises give even more favourable figures, ranging from 90 to 95% (look at their websites, the numbers are changing with time). This represents good results in the context of experiments, but not so good in a trial; 76% accuracy means 24% error, that is, almost one quarter. If 90% is better, it leaves nonetheless 10% error. Of course, not every error would result in a miscarriage of justice, because fMRI data would be only one piece of evidence among many. However, the nature of brain pictures and their psychological effects are a source of concern; this constitutes the second legal obstacle.

Brain pictures could have a devastating *effect on juries* in that they could influence their minds widely beyond the evidence the images can afford. This has been studied by Gurley and Marcus not in cases of lie detection, but of the insanity defence or NGRI (not guilty by reason of insanity), a study reported by Sinnott-Armstrong and colleagues in these terms:

> Gurley and Marcus (2008) 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. (2008, 369-70)

However, as the authors comment, it does not prove that jurors are unduly influenced; maybe MRI gives them the information they need to decide correctly.

This optimistic stance is nevertheless not warranted, in the sense that we know that brain images have a strong influence on people (Racine, Bar-Illan and Illes, 2005). Several studies have demonstrated that when a brain picture is added to an argument or a thesis, it appears more convincing. The same is true if brain information is added. For instance, Deena Skolnick Weisberg and colleagues (2008) have shown that, for students, the addition of neuroscience information moderately increases their confidence in good explanations, but worse it blocks lucidity and transforms bad explanations into good ones in their minds. The authority of pictures and of science can be misleading for an uninformed public. As jurors belong to the public, they will be prone to be misled, as Paul Wolpe and colleagues state: 'Brain scan images might influence juries even when the images add no reliable or additional information to the case' (2005, 47).

This conclusion is corroborated by a very recent study led by David McCabe and colleagues, the first made on fMRI and lie detection: 'Results showed that presenting fMRI evidence suggesting the defendant was lying about having committed a crime was more influential than any other evidence condition' (2011, 574). This is not surprising, but a source of true concern. Nevertheless, there exists some hope: when subjects are informed of the limitations of fMRI, the confidence aroused by it decreases to the same level as that of other evidence conditions.

For some authors, brain imaging can have a second bad effect on juries. The *role of juries* is to assess what witnesses say and to weigh the arguments presented. fMRI would change that and could nullify the role of juries. With this argument, a defence attorney has successfully pleaded against the admission of brain imaging in a trial: 'Defence attorney Jessica Cortes of the firm Davis and Gilbert won her motion to exclude the evidence without getting into the science behind brain scans. Juries are supposed to decide the credibility of the witness, she argued, and fMRI lie detection, even if it could be proven completely accurate, infringes on that right' (Madrigal, 2010, 1). A success for the attorney, but grounded on a weak argument: scientific data has been admitted by courts for a long time. Fingerprints and DNA are daily invoked as evidence, and without any substantial change in the role of juries. The admission of fMRI data would probably not introduce any further change. And even if it did, would it be a great loss for justice if the role of juries was modified or even nullified? In many countries, juries have lost importance, due to the growing complexity of cases. This could well be an improvement, but that is another debate. For the time being, we have a lot of other reasons to be wary about the use of brain imaging in the courts. This will be my conclusion below. But I must still address one last legal obstacle: the *importance of behaviour*.

When we discuss technical matters, we often tend to forget what is really at stake. fMRI could be a useful tool or a useless tool, but it will remain just a *tool* for the institution of justice. The aim of this institution is to assess behaviour: courts judge behaviour, not brains. Therefore, a structural or functional brain picture becomes relevant only if a person has broken the law, or is accused of it – i.e., to have behaved unlawfully. Sinnott-Armstrong and colleagues emphasise: 'What matters to law is not brain function but behaviour, and abnormal brain function does not necessarily make abnormal behaviour likely' (2008, 364). Stephen Morse had already insisted on that point when the question of the death penalty for teenagers was debated in the United States. Commenting on *Roper v. Simmons*, he said: 'If the behavioural differences between adolescents and adults are slight, it would not matter if their brains were quite different. Similarly, if the behavioural differences are sufficient for moral and constitutional differential treatment, then it would not matter if the brains were essentially indistinguishable' (2006, 48).

The impact of this last obstacle spreads widely beyond the question of the use of brain imaging in the courts, but it is useful to recall it in order that fMRI stays in its proper place, if ever it is accepted as evidence by courts.

6. Conclusion

In this paper, I have listed and reviewed a large number of obstacles to the use of neuroimaging in the courts. Other obstacles exist, but I have discussed the ones I find philosophically the more interesting. What is the result of this examination? It appears to be widely negative: fMRI and other devices to detect liars should not be used in the courts, because the present obstacles are huge. But I have also said that fMRI should stay in its proper place, if ever it is accepted as evidence by courts. Does fMRI have a proper place? I consider that it has one, or rather that it will have one when its more important obstacles are overcome. I particularly think of technical and conceptual ones. Methodological and legal obstacles are of a more general nature: every piece of scientific evidence invoked in a lawsuit meets with them. Therefore they are reasons to be circumspect, but not to exclude scientific evidence from the courts. Technical obstacles are limitations that could be overcome in the future, at least sufficiently: if measurements become more precise in space and in time, difficulties with the interpretation of BOLD factors and with subtraction method could be lessened, and we will perhaps have a better way to manage the uniqueness of the neural signature. Conceptual obstacles will require conceptual and empirical work to be overcome. We must make efforts to develop a more precise psychological theory of lie and deception, a theory that will inspire neuroimaging experiments. This is not out of our reach at all, but it will take time.

The problem is finally not with the use of fMRI or other devices for lie detection in the courts *per se*, but with a premature adoption of an existing device. And if there is an actual risk in a premature adoption, it is because the brain imaging technology is not well understood by certain of the interested parties. I hope that this paper will throw some light on this.

Acknowledgments

I wish to thank Patrik Vuilleumier for his helpful comments. An earlier version of this paper was presented at the 5th International Conference on Applied Ethics, at Hokkaido Universty, on 6 November 2010.

References

Augustine, *De Mendacio*, http://www.augustinus.it/latino/menzogna/index.htm.

Cosmides, L. and Tooby, J. (2008), ‘Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters’, in Sinnott-Armstrong, W. (ed.), *Moral Psychology*, vol. 1. Cambridge Mass.: MIT Press, 53-119.

Ellis, R. D. (2010), ‘On the Cusp’, in Giordano, J. and Gordjin, B. (eds.), *Scientific and Philosophical Perspectives in Neuroethics*. Cambridge: Cambridge University Press, 66-94.

Greene, J. and Paxton, J. (2009), ‘Patterns of Neural Activity Associated with Honest and Dishonest Moral Decisions’, *Proceeding of the National Academy of Sciences*, 106 (30): 12506-12511.

Gurley, J. R. and Marcus, D. K. (2008), ‘The Effects of Neuroimaging and Brain Injury on Insanity Defenses’, *Behavioral Sciences and the Law*, 26: 85–97.

Hardcastle, V. G. and Stewart, C. M. (2009), ‘fMRI: A Modern Cerebrascope? The Case of Pain’, in Bickle, J. (ed.), *The Oxford Handbook of Philosophy and Neuroscience*, Oxford: Oxford University Press, 179-199.

Holderied-Milis, V. (2010), ‘Online Chat – a Hatchery of Lies? Toward an Ethical Analysis of Truth and Lying in Internet Chatrooms’, *Ethical Perspectives*, 1: 95-118.

Kanwisher, N. (2009), ‘The Use of fMRI in Lie Detection: What Has Been Shown and What Has Not’, in Bizzi, E., Hyman, S., Raichle, M., Kanwisher, N., Phelps, E., Morse, S. et al. (eds.), *Using Imaging to Identify Deceit*. Cambridge Mass.: American Academy of Arts and Sciences, 7-13.

Kim, J. (1998), *Mind in a Physical World*, London: MIT Press.

Madrigal, A. (2009.1), ‘MRI Lie Detection to Get First Day in Court’, *Wired Science*, 16 March, wired.com, 1-11.

Madrigal, A. (2009.2), ‘Brain Scan Evidence Rejected by Brooklyn Court’, *Wired Science*, 18 September, wired.com, 1-7.

Madrigal, A. (2010), ‘Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings’, *Wired Science*, 5 May, wired.com, 1-3.

McCabe, D. P., Castel, A. D. and Rhodes, M. G. (2011), ‘The influence of fMRI Lie Detection Evidence on Juror Decision-Making’, *Behavioral Sciences and the Law*, 29: 566-577.

Morse, S. (2006), ‘Moral and Legal Responsibility and the New Neuroscience’, in Illes, J. (ed.) *Neuroethics*. Oxford: Oxford University Press, 33-50.

Phelps, E. (2009), ‘Lying Outside the Laboratory: The Impact of Imagery and Emotion on the Neural Circuitry of Lie Detection’, in Bizzi, E. et al. (eds.), *Using Imaging to Identify Deceit*. American Academy of Arts and Sciences, 14-22.

Proust, M. (1999), *Albertine disparue*, in *A la recherche du temps perdu*, Paris: Gallimard-Quarto.

Raichle, M. (2009), ‘An Introduction to Functional Brain Imaging in the Context of Lie Detection’, in Bizzi, E. et al. (eds.), *Using Imaging to Identify Deceit*. American Academy of Arts and Sciences, 3-6.

Racine, E. (2010), *Pragmatic Neuroethics*, Cambridge Mass.: MIT Press.

Racine, E., Bar-Ilan, O. and Illes, J. (2005), ’fMRI in the Public Eye’, *Nature Review Neuroscience*, 6 (2): 159-164.

Racine, E., Bell, E. and Illes, J. (2010), ’Can We Read Minds?’, in Giordano, J. and Gordjin, B. (eds.), *Scientific and Philosophical Perspectives in Neuroethics*. Cambridge: Cambridge University Press, 244-270.

Rakoff, J. (2009), ‘Lie Detection in the Courts: the Vain Search for the Magic Bullet’, in Bizzi, E. et al. (eds.), *Using Imaging to Identify Deceit*. American Academy of Arts and Sciences, 40-45.

Roskies, A. L. (2007), ‘Are Neuroimages Like Photographs of the Brain?’, *Philosophy of Science*, 74: 860-872.

Saenz, A. (2010), ‘Another Attempt to Use fMRI Lie Detector in US Courts Fails in Brooklyn’, *Singularity Hub*, 6 May.

Sinnott-Armstrong, W., Roskies, A., Brown, T. and Murphy, E., (2008), ‘Brain Images as Legal Evidence’, *Episteme*, 359-373.

Sip, K. E., Roepstorff, A., McGregor, W. and Frith, C. D. (2007), ‘Detecting Deception: The Scope and Limits’, *Trends in Cognitive Neuroscience*, 12 (2): 48-53.

VanMeter, J. (2010). ‘Neuroimaging’, in Giordano, J. and Gordjin, B. (eds.), *Scientific and Philosophical Perspectives in Neuroethics*. Cambridge: Cambridge University Press, 230-243.

Weisberg, D. S., Keil, F. C., Goldstein, J. Rawson, E. and Gray, J. R. (2008), ‘The Seductive Allure of Neuroscience Explanations’, *Journal of Cognitive Neuroscience*, 20 (3): 470-477.

Wolpe, P. R., Foster, K. R. and Langleben, D. (2005), ‘Emerging Neurotechnologies for Lie-Detection: Promises and Perils’, *The American Journal f Bioethics*, 5 (2): 39-49.