Can Neurological Evidence Help Courts Assess Criminal Responsibility? Lessons from Law and Neuroscience

Eyal Aharoni,[a] Chadd Funk,[a] Walter Sinnott-Armstrong,[b] and Michael Gazzaniga[a]

[a]*University of California, Santa Barbara, Santa Barbara, California, USA*

[b]*Dartmouth College, Hanover, New Hampshire, USA*

Can neurological evidence help courts assess criminal responsibility? To answer this question, we must first specify legal criteria for criminal responsibility and then ask how neurological findings can be used to determine whether particular defendants meet those criteria. Cognitive neuroscience may speak to at least two familiar conditions of criminal responsibility: intention and sanity. Functional neuroimaging studies in motor planning, awareness of actions, agency, social contract reasoning, and theory of mind, among others, have recently targeted a small assortment of brain networks thought to be instrumental in such determinations. Advances in each of these areas bring specificity to the problems underlying the application of neuroscience to criminal law.

Key words: **neuroscience; neuroimaging; fMRI; law; ethics; responsibility; excuse; defense; insanity; diminished capacity; mens rea; free will; determinism**

By the time the battery acid had circulated through his bloodstream, Pablo Ortiz was quickly slipping away. The perpetrators of this vicious act, Simon and Heriberto Pirela, then instructed their accomplices to "finish him off" or face the same fate (Weichselbaum 2004). In 1982, a Pennsylvania judge sentenced Simon Pirela to death for murder in the first degree.

Twenty-one years later, an appellate court reduced Pirela's punishment to life in prison on account of new evidence. What kind of evidence could possibly mitigate this killer's crime? Pictures of Pirela's brain. Neuroimaging data successfully convinced the judge that Pirela was not eligible for the death penalty because he suffered from aberrations in his frontal lobes, diminishing his ability to function normally (*Commonwealth of Pennsylvania v. Pirela*, 2007). It seems that when the mind is on trial, pictures of brains are worth a thousand words.

In the time since Pirela's victim took his final breath, unprecedented advances have been made in cognitive neuroscience. As the science has become available, defense attorneys have inevitably tried to use this science to help their clients. The best-known use of neuroscience in criminal trials may be found in the Supreme Court case of *Roper v. Simmons* (2005), which ruled out the death penalty for crimes committed by adolescents younger than 18 years. Since then, the use of imaging data as a tool to reduce responsibility or completely exculpate criminal offenders has become a familiar target for hopes, jokes, and contention.

Can the best neuroscientific evidence available today reduce or even rule out criminal responsibility? Many scientists, lawyers, defendants, and media representatives are already eagerly answering "yes." Others are enthusiastic that such evidence could be used by prosecutors to *establish* criminal responsibility. However, the utmost caution must be taken in making these claims because false conclusions are likely to cost real lives and livelihoods. As some of us have argued previously, the worth of neuroscience in criminal decisions is far from obvious, in part because there is not, and will never be, a brain correlate of responsibility (Gazzaniga & Steven 2005; Grafton et al. 2006–2007). Rather than being an independent neutral property of an individual, legal responsibility also requires a normative judgment that depends on the social purposes of those who ascribe it. Neuroscience can offer us only descriptive models of brain organization and function; ascriptions of responsibility, on the other hand, are unequivocally

Address for correspondence: Eyal Aharoni, M.A., Department of Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106. Voice: 805-893-2791; fax: 805-893-4303.
aharoni@psych.ucsb.edu

Ann. N.Y. Acad. Sci. 1124: 145–160 (2008). © 2008 New York Academy of Sciences.
doi: 10.1196/annals.1440.007

prescriptive. This is one reason why to explain, by itself, is not to excuse.

To examine whether neuroscience can inform determinations of responsibility, we need to begin by examining how neuroscience fits into a larger philosophical debate about responsibility in general and then identify the legal criteria for criminal responsibility in particular. Finally we must ask whether and how current neuroscientific findings can be used to determine whether particular individuals meet these criteria.

This task might seem relatively straightforward, but the language of law is vastly different from the language of neuroscience. Matching neurological data to legal criteria can be much like performing a chemical analysis of a cheesecake to find out whether it was baked with love. To span the divide between the way neuroscientists describe mental states and the way the law applies them, we must develop a set of rules for evaluating when a defendant's neurological profile meets or fails a particular legal requirement. Too liberal, and the guilty run free; too strict, and the ill and innocent suffer imprisonment or death. These rules would have to reconcile probabilistic findings about continuous mental states with categorical legal decisions about guilt and punishment. These rules also need to apply findings about groups to particular defendants. Though difficult to devise, such a system of rules could reveal when, if ever, neuroscientific techniques will become of adequate use for criminal trials.

To accomplish this feat in our lifetime would take nothing short of a miracle. A humbler goal, which we adopt here, is to (1) outline key philosophical issues, (2) identify how the law determines when a defendant is considered responsible, and (3) apply rigorous exemplars of modern neuroscience to these criteria in hopes of providing useful models for future scientific research and legal decision making. We have no guarantee that neuroscientific models, in all their detail, will make responsibility determinations easier rather than harder. Hence, another of our goals is to evaluate how well this technology can contribute in the coming years to the ongoing challenge of improving the criminal justice system.

Philosophical Background

Our task would be relatively easy if responsibility could be ruled out simply by finding any old neural cause of action. The resulting temptation is to proclaim, "*I* didn't do it—my *brain* made me do it." This move is supported by a classic philosophical argument:

(1) Every act is determined.
(2) If an act is determined, then its agent is not responsible for the act.
(3) Therefore, no agent is responsible for any act.

If a cause of an action does not just make that action likely but determines that the action definitely will be done, then any action that is caused is also determined. Neural causes would determine their effects just as much as any other causes. Then, if we can trace an act to a neural cause, its agent is not responsible, according to this argument.

This argument seems to have special force when we can trace an action to causes beyond its agent's control. The neural connections that affect actions developed long before the actions (perhaps during childhood) and were caused by external circumstances or prior events that were beyond the agent's control. Moreover, most agents do not know what is going on in their brains, so they cannot choose certain neural events rather than others with any specificity. In that way, the neural causes of an action are beyond the agent's control. Such considerations, among others, lead some philosophers to deny that agents are responsible for anything they do (see, e.g., Greene & Cohen 2004; Pereboom 2001).

Most philosophers, however, are not hard determinists. Many reject the conclusion that agents should not be held responsible and contest one of the argument's premises. The result is an array of positions (see Table 1).

Some indeterminists defend responsibility by denying determinism and claiming instead that many human actions are uncaused. However, critics charge that uncaused actions, if there were any, would be random, and random actions do not merit responsibility. Many philosophers have argued that randomness removes responsibility. By analogy, a robot programmed to shoot a gun as an output of a random outcome generator is no more free than one programmed by a fixed-outcome generator. Both robots' actions are unfit for responsibility. And, of course, robots that appear random can really have causes that we don't detect. Likewise, human volition that is truly random would not lead to responsible action, and human volition that appears random might really be caused.

Instead of claiming that human actions are not caused at all, most libertarians claim that human actions are self-caused or caused by the agent rather than by any prior event. In the jargon, they deny event causation and then invoke agent causation to avoid randomness (Kane 1996; van Inwagen 1983). There are several problems with this view. First, the denial of

TABLE 1. Philosophical positions on responsibility

	Accept premise (1): determinism	**Deny premise (1): indeterminism**
Accept premise (2): incompatibilism	Accept conclusion (3): hard determinism	Deny conclusion (3): libertarianism
Deny premise (2): compatibilism	Deny conclusion (3): soft determinism	Deny conclusion (3): soft indeterminism

event causation becomes less and less defensible in light of contemporary neuroscience, genetics, behaviorism, and every other science that models antecedents of human behavior. Second, it is difficult to make sense of the notion of agent causation (Pereboom 2001, chaps. 2–3). The most basic problem is that the agent exists equally before the action is done, while it is done, and after it is done, so citing the agent as a cause cannot explain why the act was done at the particular time when it was done. Only prior events can explain that.

Problems like these lead many philosophers toward compatibilism. They claim that determinism can be reconciled with freedom and responsibility. But how? One popular approach interprets freedom in terms of responsiveness to reasons, and agents can respond to reasons even if they are determined to do so (Fischer & Ravizza 1998; Wolf 1990).

Compatibilism has also been embraced by various legal scholars. Morse and Hoffman, for example, warn on both logical and practical grounds that we should not infer from the bare descriptive fact that an act is caused to the prescriptive claim that its agent ought to be free of responsibility (Morse & Hoffman 2007, 80–81). According to these authors, "My genes made me do it," "My upbringing made me do it," and "My Twinkie made me do it" are not popularly compelling excuses in modern jurisdictions, so "My brain made me do it" should not be exculpatory either (Morse & Hoffman 2007, 82). In their view, causal explanations by themselves provide insufficient grounds to excuse.

Moreover, neuroscience cannot demonstrate that all acts are determined. One reason is that most neuroscientific studies reveal only correlations rather than causation. Even studies that find neural causes do not prove that those causes are deterministic, and they clearly do not generalize to all actions of all sorts. To the confusion of many, neuroscience might then be used to both support and undermine determinism. Neuroscience is not independently qualified to prove either that all actions are caused or that any actions are entirely spontaneous. These are ancient and enduring debates that science or philosophy will not soon solve.

So let's leave determinism behind. Neuroscience still might raise separate problems for freedom and responsibility. Regardless of what, if anything, causes our wills, we also need to ask what, if anything, our wills cause. To see why, imagine that someone plans to kill a rival by running him over as the rival jogs in the park. As the driver backs out of his driveway on the way to commit the murder, the jogger unexpectedly appears and is run over and killed by accident. The driver consciously and did freely will to kill the jogger and had that will at the time when he killed the jogger. Nonetheless, the driver's will did not cause the accident or the death in a normal or expected way. Hence, the driver might not be guilty of either reckless driving or attempted murder. Even if the driver were reckless, this particular act of killing was not done *from* free will, and the driver is not responsible for murder. What this case shows, then, is that freedom and responsibility for an act require more than just a will or intention to do the act. They seem to require that the act results from the conscious will in a normal way.

Moreover, many views of freedom and responsibility focus on conscious will. Libertarians who invoke agent causation often cite the agent's conscious reasons or will as the crucial part of the agent that makes the agent responsible. Compatibilists who analyze freedom in terms of responsiveness to reasons usually refer to conscious reasons, because it is not clear how we could be expected to respond to reasons that we are not aware of. Some laws explicitly require conscious intentions as elements of crimes. This focus on consciousness is supposed to seem plausible in examples. Imagine that someone cooks some soup for a friend, and the cook's only conscious goal is to make the friend happy. Unfortunately, the soup contains peanuts—to which the friend is allergic. If the cook is not negligent, then he does not seem responsible for the friend's allergic reaction. But suppose that a prosecutor or an enemy argues that the cook has some kind of unconscious desire, plan, or will to hurt his friend. Such an unconscious will does not seem enough to make the cook responsible on several common views. After all, if the cook is not conscious of deciding to hurt his friend, how can he control whether he does hurt his friend? Without such control, how can he be responsible for hurting his friend?

Such examples and lines of thought give some initial plausibility to the principle that an agent is free and responsible only for acts that result from that agent's conscious will. This view does not deny that our actions often result from unconscious processes that guide our actions better than conscious thought or decision

would. The claim is only that we are not responsible when our acts result from unconscious processes with no input from conscious will. This thesis has been questioned, but the point here is only that it is popular and persuasive.

Neuroscience raises doubts about this common assumption. In classic experiments, Libet (2004) and his collaborators developed a clever method for determining precisely when participants become conscious of choosing to flex their wrists or fingers. They then used electroencephalograms to determine the onset of activity in the supplementary motor area (SMA), called the "readiness potential," that begins the process that leads to flexing. Surprisingly, this neural activity in the motor strip starts, on average, about 350 ms *earlier* than the consciousness of making a choice. But if the conscious choice to flex one's finger comes later than the neural initiation of action, then the conscious choice cannot independently cause the action in the way that most people assume.

Libet denied that his results rule out free will or responsibility because he thought that, through conscious processes, we still have time to stop the neural activation from causing motion in the finger. Critics respond, however, that the decision to stop that process is itself a consequence of unconscious neural processes. If so, it is hard to see how the decision to stop the action could be any more free or effective than the apparent conscious choice to start it. It is then unclear how free will could play a causal role in action (Wegner 2002).

Of course, defenders of freedom and responsibility have many responses. One common objection is that subjects consciously chose to cooperate in the study, so they willed to flex their fingers long before the brain potential that Libet measured. Nonetheless, this earlier general intention to flex a finger at one time or another did not cause subjects to flex a finger at the precise time that they did, so that intention did not cause the particular action that they performed. The above example of driving over the jogger shows that particular proximate intentions rather than such distal intentions determine responsibility. Hence, Libet's results challenge traditional views of freedom and responsibility, even if our acts are preceded by some general intentions.

Another common objection is that Libet's experiments used simple actions, so more complex actions still might result from conscious intentions. When criminals rob banks according to their plans, their conscious intentions do seem to cause them to do what they do. However, such more complex actions are made up of smaller actions like flexing body parts, and it is not clear how the whole can be free if its parts are not. It is possible for freedom to apply only at the higher level of generality, but it is a challenge to explain how this works. Besides, acts like raising a hand are exactly the kind of case that defenders of free will use to show that they are free: "See. I can raise it or lower it, as I wish." If Libet's findings show that these supposed exemplars of freedom do not result from conscious intentions, that would seem to cast doubt on a larger range of cases. And even if Libet's claim applied only to simple actions, it would still be surprising and important that those actions do not result from conscious will, since they seem to.

More technical objections are also raised. Some critics chide Libet for sometimes describing readiness potentials as intentions or decisions when they are more like urges (Mele 2006). Others question the precision of subjects' post-hoc reports about when they made decisions. Still others point out that similar neural activations can occur when the subject does not intend to act or when the subject only watches someone else (Kilner et al. 2004). All these objections need to be taken seriously. Although Libet's results are not conclusive, they do point toward one direction in which neuroscience might challenge our traditional views of freedom and responsibility without even mentioning determinism.

Ultimately, a keen knowledge of why people break the law might gain leverage from understanding not how free agents make choices but how causal brains influence people to follow some rules and not others. However, we are a long way away from predictive models of how people adhere to legal rules, so until that time, it is important that neuroscience still be compatible with the subjective appearance of freedom. A human action can be both determined, possibly by readiness potentials before conscious willing, and still subjectively free at the same time (see Gazzaniga 2005, and Gazzaniga & Steven 2005 for similar arguments). Choosing to raise one's hand results from physical causes, but it also *feels* free. Even if this feeling of free will is a mere construction of the brain, this does not mean that such a construction is useless or ineffectual. The felt experience of free will, and the ability to attribute free will to others, seems to be integral for human beings to navigate our complex social landscape. Our intuitions about free will enable us to make predictions about human action that are as good as or better than our best neuroscientific models. They cannot then be dismissed merely as misleading illusions.

In these ways, although neuroscience is not equipped to resolve the ancient philosophical debates, the application of neuroscience can illuminate those debates. Moreover, although modern neuroscience is

far from becoming a direct mechanism of exculpation, neuroscientific studies may potentially help to clarify when agents are responsible by testing accusations and excuses against systematic observations of real human behavior. This possibility becomes especially important when neuroscience is applied to legal decisions.

How the Law Determines Whether a Defendant Is Responsible

In determining who is responsible, the law is where the rubber meets the road. However, it is not easy to define how the law determines responsibility. One reason is that different legal jurisdictions have different criteria for responsibility. This variation itself is an indication of the inherent difficulties in defining criteria of responsibility. Furthermore, determinations of responsibility can play many roles in a trial.

The relevance of neuroscientific evidence has been implicated in at least two of these roles: defenses that deny intentions and affirmative defenses, such as insanity. Other variants of mens rea, such as recklessness and negligence, and other excuses, such as duress, coercion, irresistible impulse, and automatism, also bear on criminal responsibility and are intriguing candidates for neuroscientific investigation. However, several recent neuroscientific studies seem directly relevant to intention and insanity. Examining these studies and concepts will provide substantial room for our discussion on the future of criminal responsibility.

We will focus on the extent to which neuroscience could be used to *reduce* responsibility, not to *establish* it. This is because (1) establishing responsibility seems to require the ability to decipher the content of particular mental states, which is a much harder problem for neuroscience to solve than ruling these states out by furnishing evidence that a defendant lacked the capacity for a certain mental state, and (2) establishing responsibility is likely to rely on mandatory neuroimaging, which might violate the defendant's right to privacy, right against search and seizure, and right against self-incrimination—all complex debates that have been reviewed elsewhere (Tovino 2007).

Hence, our topic will be how denials of intentions and of sanity can be used to reduce or remove responsibility. We will begin by outlining the legal concepts of intention and insanity. Then we will discuss some relevant scientific studies.

Mens Rea

A criminal act is standardly divided into the actus reus and the mens rea, which are, respectively, roughly the physical act and the mental element. Theft, for example, might require taking property (the actus reus) with the knowledge that it belongs to someone else and the intention to deprive that person of it (the mens rea). Both elements of a crime must be proven beyond reasonable doubt for the prosecution to convict a defendant of that crime.

Different crimes require different mental states, or mens rea, for conviction. The crime of first-degree murder usually requires an intention to kill (except in cases of felony murder, depraved indifference, and Pinkerton liability), whereas manslaughter can often be committed without any intention to kill. The mens rea required for a given crime can also vary between jurisdictions and over time.

Mens rea commonly includes at least four variants: intention or purpose, knowledge that the act is done, recklessness, and negligence (Model Penal Code § 2.01 1962). These refer to specific mental states required for an act to qualify as an instance of a particular crime. Some crimes require intent, such as an intention to kill, harm, or deprive of property. Other crimes do not require intention but do require knowledge, such as knowledge that the harm will occur. Still other crimes do not require knowledge that a harm or offense will definitely occur but, instead, require only knowledge of a risk of the harm or offense. Unjustified disregard of that known risk then constitutes recklessness, as in some cases of drunk driving. Finally, some crimes do not require actual knowledge even of risk but are committed when an agent should have known about the harm or the risk. The act or agent is then called negligent. A successful defense against any of these variants results in a reduced charge to a less serious offense or, sometimes, an acquittal.

Neuroscience might in principle be used to determine when any of these mental conditions is met. However, current neuroscience speaks most strongly to matters of intention, so we will focus on that form of mens rea.

Intention is commonly defined as a commitment to a plan of action. In normal cases, an act is done intentionally when the actor commits to a plan that includes that action as an essential part. For an act to be intentional, the actor must also know that he is planning and performing it. People kill intentionally in this sense when they know that death is a likely consequence of their acts and when the resulting death is also part of what they need to accomplish to fulfill their plans (Bratman 1987; Perkins 1969, 747). Although long-range premeditation sometimes may be present, it is not necessary to establish intention.

When a crime requires intention of this kind, the prosecution needs to show that the defendant had sufficient knowledge and relevant plans. The defense can respond that the defendant lacked one of these mental states. One way to show this lack might be to show that the defendant has abnormalities in the brain that prevent the defendant from forming or committing to plans of action. Many jurisdictions recognize this as a defense of diminished capacity (e.g., *Kansas v. Wilburn* 1991). These jurisdictions usually regard the diminished capacity defense as a challenge to intention or other culpable states, different from an insanity defense, which we will discuss in the next section.

If it can be empirically demonstrated that a defendant has difficulties forming intentional actions, there might be some probability that he lacked the capacity to act intentionally at the time of the offense. If this case can be made, he might be acquitted of the greater crime and charged with a lesser crime if one is available. In this way, neuroscience could become relevant to the criminal charge. The burden is then placed on neuroscientists to identify and measure abnormalities associated with these functions and dysfunctions. This is a tall order that we will examine below.

Insanity

Neuroscientific evidence also seems applicable to the insanity defense. A successful insanity defense acquits the defendant "by reason of insanity" and usually results in commitment to a mental hospital as long as the defendant is a danger to himself or others.

How insanity is defined could have important implications for how neuroscience evidence is used. The definition varies quite a bit among the United States. It also takes different forms in case law from that in statutes. Still, some formulations are common. Some states still adhere to the M'Naghten Rule, according to which a person is legally insane if

> . . .at the time of the committing of the act, the party accused was laboring under such a defect of reason, from disease of mind, as not to know the nature and quality of the act he was doing; or if he did know it, that he did not know what he was doing was wrong. (1843)

This test is purely cognitive insofar as only lack of knowledge can support a verdict of not guilty by reason of insanity.

What counts as knowledge is not clear, so some later laws required defendants not only to know but also to "appreciate" what they were doing and that it was wrong in order to be found guilty. Some jurisdictions also added a volitional prong, so that defendants could be found not guilty by reason of insanity if they had a mental illness that either removed their capacity to resist impulses or created an impulse so strong that nobody could resist it. Either of these conditions would destroy the defendant's capacity to abide by law.

These developments culminated in the insanity defense of the Model Penal Code of the American Law Institute:

> A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the wrongfulness of his conduct or to conform his conduct to the requirements of the law.

So that psychopaths and sociopaths would not be excused, another clause in the Model Penal Code added, "the terms 'mental disease or defect' do not include an abnormality manifested only by repeated criminal or otherwise antisocial conduct."

The Model Penal Code test of insanity was accepted in most U.S. jurisdictions until John Hinckley was found not guilty by reason of insanity after trying to assassinate President Ronald Reagan (*U.S. v. Hinckley* 1982). Many commentators criticized that decision and blamed it on the broad insanity defense in that jurisdiction. As a result, most states removed the conformity prong of the Model Penal Code test and adopted purely cognitive insanity defenses closer to the old M'Naghten rule. Some jurisdictions also shifted the burden of proof, so that the defense would have to prove insanity and the prosecution would not have to prove sanity.

After these changes, defendants could still be found not guilty by reason of insanity if they showed that they could not know what they were doing or that they could not know that what they were doing was wrong. Such lack of knowledge could result from delusion, retardation, and possibly sleepwalking or automatism in some jurisdictions.

Neuroscientific evidence then becomes useful if it can establish such lack of capacity to know. As with challenges to the existence of mens rea, neuroscientists who want their research to be legally relevant face the challenge of designing a method of measuring an individual's capacity to know. In the next section, we discuss the advances and difficulties in operationalizing these constructs at a neural level.

Comparing Current Neurological Data to Legal Standards

As suggested, an effective neuroscience of responsibility might begin by showing a neurological basis for a reduced capacity to fulfill the mens rea and sanity requirements. We begin by evaluating some recent

evidence related to intention formation and then the two chief conditions of legal insanity.

Defense against Mens Rea

How might neuroscience determine whether a particular person performed a particular act intentionally? Unfortunately, modern neuroscience is in no position to demonstrate a lack of intention for the particular crime charged because of course there is no known way to retroactively observe the state of the brain as it was during the commission of the offense.

It still might be possible to answer this question indirectly and probabilistically by showing that the individual probably lacked the ability to form plans or intentions. To support such a defense, neuroscientists would need (1) to identify some network in the brain that is necessary for forming "intentions" and then (2) to show that this network in the defendant's brain is dysfunctional during attempts at planned action, preventing him from reliably forming intentions but leaving intact the ability to perform the prohibited action (the actus reus).

To localize an "intention" mechanism in the brain, an ideal approach would be to measure the neural activation of an intended act as well as the neural activation of the same act performed unintentionally and then to subtract the latter from the former. Unfortunately, it is no easy task to systematically elicit specific unintentional actions, especially within the confines of a brain scanning apparatus. That might be why this direct approach has not been used.

An alternative approach could be to have participants intentionally perform simple motor behaviors such as a button press and to then instruct them to attend either to the experience of the act or to the experience of intending itself, and subtract the activation in the action condition from the activation in the intention condition. The fundamental assumption underlying this design is that attention directly modulates the activity of brain networks responsible for intention formation. Specifically, by attending to intention formation, there should be enhanced activity in the networks responsible for intention formation relative to conditions in which intention is formed less attentively.

Lau et al. (2004) did precisely this in a penetrating article in *Science*. They identified unique activation in the pre-SMA when participants attended to their intentions. This finding led the authors to conclude that this area of the brain generates intentions for motor behavior. They also observed an interaction with the dorsolateral prefrontal cortex, leading them to believe that this area was responsible for the ability to attend to these intentions. They concluded that this "attention to intention may be one mechanism by which effective conscious control of actions becomes possible" (1210).

Data like those of Lau et al. might seem to suggest that intention has been found in the brain and that observable abnormality in this area is proof of impairment in forming an intention. However, as with any study, such tempting inferences must be qualified. One limitation arises from the initial assumption that attention positively modulates activity in areas implicated in intention. A plausible alternative, for instance, might be that when attending to intention, this metacognitive task produces cognitive load for the attention processor, perhaps even *reducing* intentional capacities, and that the resulting increased activation is actually evidence of this attentional load, not of intention per se.

More evidence for the involvement of the medial frontal cortex in intention is reported by Haynes and colleagues (2007). They used a clever experimental design in which they had subjects freely decide whether they would add or subtract a subsequently presented pair of numbers. However, before these numbers were displayed, there was a time lag. This required subjects to maintain their intention for the length of the lag (which varied across trials). The numbers were displayed, and then four answers were displayed in random positions. This prevented subjects from forming habitual motor responses, enabling the researchers to isolate intention formation from motor planning. The authors found areas in the anterior medial frontal cortex that could be used to decode which intention the subjects were maintaining. Each intention was predicted by a separate pattern of activity within this region. Also, the authors noted activity in the posterior medial frontal cortex that was associated with action execution during the response phase. Like Lau et al., the investigators seem to have dissociated intention formation and maintenance from action execution, providing more evidence for the involvement of the medial frontal cortex in intention.

Others have attempted to isolate intention in the brain by comparing an intentional action task, such as a button press, to a task in which the same action has been previously conditioned as an almost automatic response to a stimulus such as an aural tone (Cunnington et al. 2002). The logic is that, if one subtracts the more "exogenous," reflexive action from the more "endogenous," effortful one, an activation profile for intention will manifest. However, this design suffers from the criticism that conditioned responses do not necessarily lack intentional regulation and conversely that the intention to act was not entirely endogenous but was prompted by instruction. Thus, it may come as no surprise that this design has yielded mixed results.

Additional evidence suggests that pre-SMA activity decays when subjects repeat a simple motor response (Sakai et al. 1999), but it is difficult to determine that this initial activation actually reflects intention and not effort. It is also difficult to show that performing practiced actions under experimental conditions is driven by the intention to perform a specific repeated behavior and not simply by the intention to follow the experimenter's instructions. It is certainly possible that these two operations are neurally distinct. For instance, work by Hoshi and Tanji indicates that the pre-SMA may be involved with specific motor intentions (Hoshi and Tanji 2004), whereas the cingulate motor area may be involved in the more general decision to cooperate with the experimenter (Hoshi, Sawamura, & Tanji 2005). This supposition is far from established but helps us to appreciate the subtle controls required for making general claims about intention.

A more central problem to overcome in studies of intentionality, and the application of neuroscience to law in general, is the homunculus problem. Indeed, one of the most fundamental asymmetries between the law and cognitive neuroscience is that while the law seeks to analyze the guilt of an *agent*, cognitive neuroscience has largely exiled the existence of a "homunculus," or an agent who resides at the top of top-down processes and makes decisions and exerts free will. Typical cognitive neuroscience models, and even the rare few that defend the notion of a homunculus with strictly limited powers (Roepstorff & Frith 2004), approach brain function in terms of a series of interacting mechanisms forming various computations in convergent, divergent, and parallel fashions. To understand how voluntary or intentional actions emerge from neural activity, one provocative theory posits the notion of competition between neural systems.

Recent efforts have had success applying this notion to activity of the pre-SMA (Nachev et al. 2005; Sumner et al. 2007). In this view, the neural correlates of self-generated, voluntary actions arise from automatic action plans originating in the pre-SMA. But even as intentions are formed, there are many competing action plans being formed in various cortical and subcortical brain areas preconsciously, in response to the environment. These include the SMA (Grèzes & Decety 2002). Competition between these conflicting plans is thought to determine which intentions become selected. Consequently, the enhanced activity of the pre-SMA may represent the neural events necessary for the selection and execution of "intentional" rather than "unintentional" action plans.

It is possible that this competition, regulated by cortical regions like the pre-SMA, takes place largely subcortically. The pre-SMA has projections to the putamen (Johansen-Berg et al. 2004; Wolbers et al. 2006) and the subthalamic nucleus (Aron et al. 2007) that may be important for motor selection and inhibition, respectively. Without the pre-SMA to provide well-orchestrated positive and negative bias to these subcortical areas, action plans that normally would have been outcompeted would be victorious. This consequence of the mechanism is crucial for assessing the selection of actions.

As the conflict hypothesis might predict, patients with an inability to perform intentional actions *are not* paralyzed. They can still perform motor actions based on environmental cues and other nonintentional sources of action. For example, patient A.G., who has damage to the pre-SMA but not the SMA proper, has difficulty voluntarily initiating action or voluntarily making online changes in the middle of an act. Nonetheless, she can perform cued behaviors with response times comparable to those of control subjects (Nachev et al. 2007).

Thus, there are two important points to glean from the neural framework outlined here: (1) There is an emerging case to be made that the pre-SMA reflects a neural basis of intention and that it displays the functional connectivity necessary for cognitive influence on intention formation and thereby on the execution of action; (2) when the neural areas responsible for intention are dysfunctional, an imbalance in competition between various automatic action plans allows complex actions to be performed in the absence of intention.

Once the neural mechanisms of intention have been implicated, a defendant's brain would have to yield evidence of dysfunction in these networks during attempts at planned action along with behavioral indicators of difficulty forming such plans. The above-mentioned research suggests that this goal might be achieved by searching for pre-SMA dysfunction in qualified defendants during intention formation.

However, even if we are looking at "intention" in the brain, it does not necessarily follow that abnormalities in this region imply an inability to form intentions. In general, abnormal activation could manifest as hypoactivation, hyperactivation, positive or negative activation, or some erratic pattern. It is not now known which of these patterns indicates dysfunction, but we do know that it will almost certainly depend on the neural area in question. This is one reason why examining the behavioral correlates of brain dysfunction is so essential. After all, the law cares about the brain only insofar as it can tell us something about its effect on actual behavior.

Another reason that abnormalities in the pre-SMA area do not imply difficulty forming intentions lies in the ample evidence of brain redundancy and distributed processing. It is plausible that when the pre-SMA area is dysfunctional, other areas "take up the slack." Other as yet undetected brain networks might supplant the computations underlying intention formation so that the procedures required for intention formation can still be normally executed. Thus, whether a particular brain network is useful does not imply that it is necessary (Halgren & Marinkovic 1996; Price et al. 1999). Second, for hypoactivity, maybe reduced activity in the intention areas simply indicates that this individual's pre-SMA requires less activation to function normally. Reduced activation in the pre-SMA, although encouraging and important, leaves many questions about intention formation unanswered.

Moreover, even if normal pre-SMA activation is necessary to form intentions, the fact that a defendant shows abnormal activation in the pre-SMA after being arrested does not show that the same area was dysfunctional when the crime was committed. It is also not clear that dysfunction in the pre-SMA for one kind of action, which is tested in the lab after arrest, shows that the defendant's pre-SMA would also be dysfunctional for other kinds of actions in real circumstances. Such questions about whether lab findings during trials establish mental states during real crimes are a recurring problem when applying neuroscience in the law.

Insanity Defense

Can neuroscience help us determine whether a defendant is insane? That depends on how insanity is defined. According to the laws discussed above, insanity seems to depend at least in part on an inability to know the nature and quality of the act and to know that the act is wrong. Neuroscience is potentially relevant to each of these conditions.

Knowing the Nature and Quality of an Act

What constitutes "knowing" something? How is the "nature" of an act different from its "quality"? Legal theorists have examined these questions in detail (Robinson, 2 Crim. L. Def. § 173, West, 2007). However, legal concepts almost always rely on intuitive explanations without clear operational scientific definitions. To apply neuroscience to legal issues, we must eventually jump the divide between legal concepts and scientific ones.

The "nature and quality of an act" (in the M'Naghten rule) includes the consequences and circumstances of the act. To know that an act has the nature and quality of killing, its agent must know that the act will have death as a consequence. To know that an act is theft, its agent must know, or at least truly believe, that the taken object belongs to someone else. If an agent cannot know such essential consequences or circumstances, that agent cannot know the nature and quality of the act. The agent then fails this cognitive part of the insanity defense.

It is still not clear what counts as *knowing*. One operational way to define "know[ing] the nature and quality of the act" is having an explicit, declarative representation of an instrumental action, including its consequences, circumstances, and means. By this definition, a defense relying on neuroscience would have to show that an individual lacks the capacity to build such representations because of dysfunction in the brain networks responsible for this operation.

At least two kinds of information are used by the brain to gain knowledge of action: efferent information governed by motor planning and afferent information such as somatosensory and proprioceptive feedback. Interestingly, afferent feedback is not required for successful execution of actions (Farrer et al. 2003). Hence, selective damage to these feedback mechanisms might not provide sufficient reason to excuse offenders with this type of damage. In contrast, motor planning and initiation do seem to be necessary for successful execution of actions. Above, we discussed the automatic, nonreflective processes underlying intention formation, but the law is concerned not just with intention formation but also with awareness, or knowledge, of planned actions, as the M'Naghten rule illustrates. Indeed, there is some evidence that these two operations are neurally distinct.

Research has shown that the angular gyrus, an area within the parietal cortex, may house mechanisms that allow us to reflect on the intentions being formed in the frontal cortex (Sirigu et al. 2004). Subjective awareness of intentions may rely on predictive models that project what the execution and consequences of the action will be like—a known function of the parietal cortex (Desmurget & Grafton 2000). These predictions are formed before sensory feedback about the action arrives. If indeed these predictive models are the correlates of awareness of motor intention, and because this awareness precedes action initiation in control subjects (Sirigu et al. 2004), one function of this awareness may be inhibition of automatically generated intentions. Evidence from at least one study suggests that it is possible to inhibit intended actions when subjective awareness of intention precedes the execution of the action (Brass & Haggard 2007). This view relegates

awareness to the indirect role of moderating the relationship between forming intentions and inhibiting them.

If the angular gyrus is dysfunctional, we may not become subjectively aware of an intention until we become aware of the resulting action through sensory feedback, after the action has already commenced (Sirigu et al. 2004). Consequently, for quick actions, such as a simple pull of a trigger, it should be possible for people with damage to this area to complete the action before becoming aware of it. However, such damage may be of little exculpatory use for someone guilty of more complex actions such as loading and aiming a gun or robbing a bank because, in such complex cases, enough time exists for sensory feedback to reach awareness before the action is complete.

The sequence of intention awareness followed by perceptual awareness of action is also important in another way. The ability to become aware of one's intentions before they have resulted in action seems necessary for a sense of agency—the experience of being the causal source of one's actions. To "know the nature and quality of [his] act," an individual would presumably have to recognize that these actions were caused by him- or herself rather than someone else.

Recent developments by Farrer and colleagues (in press) have advanced our understanding of the neural correlates of a sense of agency. In one study, participants were asked to perform a simple motor task that was visually recorded and played back to them after a short delay. Participants were led to believe that half of the delayed-feedback videos were of acts authored by someone else and that their presentation order was random. Participants then had to decide on which trials the feedback was self-authored, without the benefit of direct feedback from intentional systems. The tendency to attribute actions to an external agent correlated with increased bilateral angular gyrus activity relative to self-attribution. These results indicate that the angular gyrus may control the subjective experience of agency and that it is modulated by the synchrony between predicted (intentional) and actual (sensory) feedback. (See Moore and Haggard [2007] for a compelling theory of how the experience of agency arises from the dynamic interactions between predictive and inferential processes.)

Translating the results from these studies into the courtroom would require structural and functional imaging of the angular gyrus. If it is dysfunctional, the defendant may have diminished awareness of his motor intentions before they are actually realized. This supposition could be verified with behavioral tasks involving temporal judgment of the experiences of intention and the initiation of actions. If the angular gyrus is abnormally active at times when it should not be or if it is not modulated by delayed-feedback conditions, then it is possible that the defendant misattributes his own actions to other agents. In both cases, action awareness is diminished in such a way that the defendant could not, or would not, attempt to prevent the actions from occurring.

Once again, it is dangerously tempting to conclude that damage to this brain area necessarily implies that an individual cannot consciously represent his own actions. As we mentioned, there may be conditions under which this is possible, but we have also indicated that "awareness" may not be a unitary phenomenon and its degree of plasticity as well as its functional significance remain largely unknown. Consequently, the angular gyrus should be a starting point, not a destination, in our understanding of the different aspects of awareness and how they interact.

Knowing that an Act Is Wrong

Knowledge of rules against the act. Even if a defendant is aware of "the nature and quality of the act," he still might not be aware that it was wrong (either legally or morally). Much research has been done recently on the neural basis of moral judgments (see Greene & Haidt 2002 and Moll et al. 2005; see also Sinnott-Armstrong 2008). If neuroscientists could determine which brain circuits are necessary to form moral judgments, then dysfunctions in those circuits might be used as evidence that defendants cannot know that their acts are wrong.

To illustrate some problems for this strategy, consider a recent study (Koenigs et al. 2007). This group presented several types of moral dilemmas to six subjects with damage to the ventromedial prefrontal cortex (VMPFC). In each case, subjects chose whether to perform a hypothetical act that saves more lives by killing fewer. These dilemmas varied both in the personal involvement demanded of the subject (such as either pulling a switch or pushing someone off a bridge) and the aggregate utility of the outcome (the proportion of people saved). The investigators compared their moral judgments to those of typical subjects as well as subjects with brain damage outside the VMPFC. Subjects without VMPFC damage were relatively unwilling to endorse highly personal acts even when those acts resulted in high aggregate utility. The VMPFC subjects, in contrast, showed a relatively increased willingness to endorse such acts. This finding suggests that an intact VMPFC weighs personal involvement as a factor in moral reasoning. The authors are the first to point out that the specificity of this

finding does not suggest that these subjects lack a general capacity to judge moral wrongness, because their judgments were normal in the other conditions. Thus, such studies, while sometimes illuminating, do not yet have clear implications for the legal issue of whether defendants can know that their acts are wrong under general conditions.

Other brain studies become relevant if we assume that knowing whether something is wrong requires an ability to understand and apply social rules. Research into the neural correlates of reasoning about rules has been particularly fruitful in recent years.

In a functional magnetic resonance imaging (fMRI) experiment, Fiddick et al. (2005) had participants use both precautionary reasoning about hazardous situations and social contract reasoning about obligations to others. Participants were presented with various rules from both categories, followed by brief descriptions about people who may or may not have followed these rules. The precautionary category included items like "if you go hang gliding, then you must stay away from power lines." The social contract category included items like "if you order the buffet dinner, then you must eat the food yourself." Using only the information provided, participants had to decide whether the person could have broken the rule. Social contract reasoning but not precautionary reasoning was associated with increased activity in the bilateral ventrolateral prefrontal cortex and the medial frontal gyrus, as well as the left angular gyrus and the left orbitofrontal cortex. Interestingly, two of these areas, the medial frontal gyrus and the angular gyrus, are thought to be involved in generating emotional responses that inform some kinds of moral judgments among other things (Greene et al. 2004; Greene et al. 2001).

If dysfunction in areas associated with social contract reasoning reduces the capacity to reason about social contract rules, then this reduced capacity may in turn hinder one's ability to judge any particular social contract violation as wrong. If so, defendants with such deficits might become eligible for the insanity defense under some common formulations.

A lack of moral knowledge also might seem to excuse psychopaths (Fine & Kennett 2004). Though this conclusion might sound troubling, there is some evidence that psychopaths do not know or at least appreciate that their acts are morally wrong. First, psychopaths show reduced startle and skin-conductance responses to pictures of people harmed by violent assaults (Blair et al. 1997; Kiehl 2008; Levenston et al. 2000). This finding seems to underscore their notorious lack of empathy and may plausibly hinder their ability to appreciate the wrongness of immoral acts. Second, when psychopaths talk about moral wrongness, they often show confusion about what makes acts wrong and what it means for acts to be wrong (Kennett & Fine 2008). Third, at least one study (Blair et al. 1995) found that psychopaths fail to distinguish moral from conventional violations, oddly overclassifying conventional violations as moral ones. It is, of course, still controversial to claim that psychopaths do not know that their acts are wrong. Indeed, the ability to persuade and deceive others may profit from an ability to entertain others' notions of moral wrongness. However, even if psychopathic offenders tend to have a poor understanding of the wrongness of their acts, this would not necessarily imply that they *lack the ability* to understand wrongness. Only if this latter criterion can be met could neuroscience become legally relevant to psychopathy.

Some studies have suggested that psychopaths display a distinctive pattern in electroencephalograms (Kiehl 2008) and event-related potentials tests (Raine 1989a, 1989b). Those patterns, if they prove to be highly predictive of psychopathic behavior, might then be used to determine which defendants are psychopaths. This diagnosis could then be used to argue that these defendants are eligible for the insanity defense if psychopaths in general were shown to be incapable of appreciating wrongfulness and if that incapacity suffices for the insanity defense in the relevant jurisdiction. Although neuroscientific methods have never been considered necessary for psychopathy diagnoses, these methods might carry the potential to influence how we interpret psychopathy by showing that psychopathic brains are physically abnormal in ways relevant to crime and responsibility.

Of course, acquitted psychopaths would not be let back on the streets to commit more crimes. They would, instead, be institutionalized in a secure mental hospital rather than a prison, possibly for longer than the time they would have spent in prison. As mentioned above, the Model Penal Code added a special clause to its insanity test to avoid acquitting or freeing psychopaths. However, that special exclusion would no longer apply if the diagnosis of psychopathy can be established by neuroscientific evidence rather than by repeated criminal behavior. Although this concern remains entirely conjectural, the question of whether psychopaths are responsible needs to be faced, and it is one area where neuroscience might become relevant to legal decisions regarding insanity.

Knowledge of others' mental states. A sense of moral wrongness might also depend on the ability to anticipate that other people may be averse to the consequences of one's actions. When an act offends other people or makes them suffer or distrust the agent,

those consequences provide some reason to judge the act wrong. Defendants may be unable to know that such acts are wrong if they cannot reason normally about others' mental states—that is, if they do not have an intact Theory of Mind.

Recent cognitive neuroscience literature has implicated a network of brain areas thought to be required for a Theory of Mind (Saxe 2006). One particularly important area is the right temporoparietal junction (RTPJ). Saxe and Wexler (2005) provide strong evidence that the RTPJ is involved in belief attribution, a fundamental component of Theory of Mind. They had participants read statements about an individual's background, desires, and the outcome of a related story, and they then asked participants to decide if the character would be pleased with the outcome. To respond to this question, participants had to develop a model of the character's desires and predict the preferred outcome on the basis of that model. They found that relative to reading the background statement, reading the desire statement induced increased activity in the RTPJ. Thus, the RTPJ was active during the segment of the experiment in which participants had to attribute beliefs and desires to another person. This role of the RTPJ has been replicated in a recent study of moral judgments (Young et al. 2007).

Interestingly, the RTPJ has been identified as an area that is hypoactive relative to that in control subjects when high-functioning autistic and Asperger syndrome patients perform tasks that require a Theory of Mind (Castelli et al. 2002). A leading theory regarding the nature of autism and Asperger syndrome is that affected individuals cannot form an adequate Theory of Mind. Taken together, the results from the above studies indicate that hypoactivity in the RTPJ could result in inadequate belief attribution and, therefore, Theory of Mind. This finding lends itself to a related hypothesis, that other abnormal forms of activation in the RTPJ may result in overattributions of others' mental states—perhaps the frightening delusion that others want to harm you (a common symptom of paranoid schizophrenia). Some studies have already shown a correspondence between positive symptoms of schizophrenia and hyperactivation in other brain areas (Dierks et al. 1999). Psychotic delusions could potentially motivate an actor to deploy a defensive assault that is entirely justified within the logic of that actor's delusion.

If a defendant could be shown to have dysfunction in the RTPJ and to lack the ability to make judgments about others' beliefs and desires, a case could be made that this individual has an improper understanding of others' mental states and, consequently, lacks sufficient capacity to judge the wrongness of his own acts.

Importantly, however, even if social contract reasoning and Theory of Mind are shown to be functionally impaired, there easily could be other cues to assessments of wrongness that typical brains compute for which we simply have not accounted. It is widely accepted, for instance, that people model their behavior after other social agents (Bandura 1977). It could be that individuals can extract moral information from the behavior of others, even when social contract reasoning and Theory of Mind are dysfunctional. This is just one of many alternative hypotheses that awaits thorough psychological and neuroscientific investigation.

Future of the Neuroscience of Criminal Responsibility

Thus far we have argued that the significant advances made in the neuroscience of mental states do not yet provide compelling evidence that associated brain regions are necessary or sufficient for normal functioning of these mental states. Even our most thorough descriptions of abnormal brain activity do not necessarily imply dysfunction (Pinker 2002, 184). Moreover, if strong evidence of specific brain dysfunction is found, this alone does not necessarily imply innocence or impunity because, after all, most individuals with similar dysfunction never commit crimes (Gazzaniga & Steven 2005; Grafton et al. 2006–2007). As we discussed above, there are several missing links in the connection between neuroscience and responsibility.

How can these problems be solved? Where do we go from here? The field of law and neuroscience is changing quickly. It is hard to predict what will come next or where the field is headed. Still, we can say a little about what is needed and likely in the law and in neuroscience.

What Do We Need from Law?

The legal issues outlined above are all filled with uncertainty. Even the best neuroscientific evidence will leave us unsure whether some particular defendants meet the conditions for criminal responsibility. Mistakes can have devastating effects in criminal justice. If the scientific interpretations of neurological results are accurate, say, 90% of the time, these interpretations will be misleading 10 of 100 times. Defendants then face a considerable risk to liberty or life if they (or prosecutors) rely on neuroscientific evidence. Scientific

claims must attain impeccable accuracy when lives and livelihoods are on the line.

The importance of such claims means, first, that we need to determine the error rates of various methods in neuroscience. It is not clear how we can determine error rates to begin with. What are the average rates of misses and of false alarms for fMRI detection of various conditions? Whatever they are, these error rates are only compounded when legal officials, most of whom know little neuroscience, need to draw conclusions from technical data. The release of noisy, unreliable neuroscientific evidence into the courtroom could actually serve to increase error rates in convictions, whereby judges and juries acquit the guilty and convict the innocent.

Will neuroscience bring more harm than good to criminal trials? Only time and careful analysis will tell. Still, some steps might help to minimize its destructive power and develop a constructive trajectory for its use.

One way to reduce the worst kinds of errors is to properly distribute the burden of proof or persuasion. If society is most concerned not to convict the innocent, then it can reduce that kind of error by placing a heavy burden of proof on prosecutors. The uncertainties in neuroscientific data will make it hard for prosecutors to use such data to prove beyond a reasonable doubt that the conditions of responsibility, including intention and sanity, are met. In contrast, if society is most concerned not to acquit and release the guilty and dangerous, then it can reduce that kind of error by shifting the burden of proof onto the defense. If the defense is required to prove, even to a preponderance of the evidence, that the defendant is insane in order to be found not guilty by reason of insanity, then it will be hard to carry that burden with uncertain evidence from neuroscience. It is, therefore, crucial for the law to develop appropriate rules governing the burden of proof to be able to handle the new evidence from neuroscience.

Finally, even if we tailor law to handle error asymmetries, we still face problems related to the admissibility of neuroscientific evidence, a domain in which procedural law might shape the way we do neuroscience. It is not clear when neuroscience findings should qualify as relevant, material, or competent, or reliable, as defined by the rules of evidence. It is also not obvious under what classification it should fall: real, demonstrative, documentary, or testimonial evidence? Finally, for evidence affirming responsibility, no such evidence can be admitted that violates the defendant's basic rights, as we noted above. Such questions of admissibility are generating increasing consideration (Tovino 2007).

What Do We Need from Neuroscience?

Chomsky (1975) once distinguished puzzles from mysteries. Puzzles are closed-ended problems to which solutions can be systematically approached and obtained. Mysteries are open-ended problems to which we are bewildered at the prospect of how to go about approaching a solution. Certainly the prospect of finding a simple "understanding" or "free will" center in the brain does seem doubtful to most scientists as well as philosophers. But the challenges we have charged to neuroscience—to make accurate probabilistic inferences about whether a brain is adequately equipped to deploy coordinated, goal-oriented plans of action, to reasonably anticipate consequences of these plans, and to be amenable to veto power by other brain processes—become a smaller puzzle every day. Extending our earlier analogy: Chemists perhaps cannot determine if a cake was baked with love, but they can determine if it was baked with cyanide, which in turn provides circumstantial evidence against the love hypothesis. Likewise, although neuroscience cannot locate responsibility in the brain, perhaps it can identify maladies that provide at least circumstantial evidence against guilt or liability. Several pivotal advancements lend support to our optimism.

Immersive Technology

Our findings can be only as rich as the environments in which we test them. But how could we possibly test complex, ecological environments in a crowded fMRI chamber? One exciting possibility lies in digital immersive virtual environment technology (IVET).

Digital IVET typically powers an interface among individuals and a computer-generated, three-dimensional world. The interface can take the form of a pair of stereoscopic display goggles and headphones. The resolution can approximate perfect photorealism, and environments can be as interactive as designers choose. In fMRI, users could traverse this environment with a simple joystick.

Armed with this technology, researchers could test models of cognition in environments that are as ecologically rich as the physical and social environments in which our cognitive mechanisms evolved. For instance, researchers seeking to understand intention formation may assess not just intentions to press a button but intentions to defend against a looming attacker. These different plans of action may or may not be formed in the same area of the brain, but advanced imaging technology combined with IVET can help us find out. IVET can enrich our ability to map sophisticated

models of cognition onto the functional organization of the brain.

Advances in Temporal Resolution

A notorious drawback of research in fMRI scanning technology is its meager temporal resolution. A tremendous amount of activity can occur in the brain in a matter of seconds. The most cutting-edge fMRI scanners peak at a resolution of about 2000 ms—the length of one scan. The brain's blood oxygen level depletion response has an even longer duration. These two constraints make it difficult to know precisely when a predicted neural signal has occurred. Event-related designs can limit these problems to some extent, but these designs are expensive and taxing. Another solution is to supplement fMRI with higher temporal-resolution measures, such as event-related potentials. Synchronization between these technologies will provide the needed leverage to precisely localize brain events in both time and space. This triangulation tactic will enable neuroscientists to draw stronger inferences about causality between brain events, as well as between the brain and the body.

It's Not All about fMRI: Diffusion Tensor Imaging

A third technology with high hopes of building more sophisticated models of cognition is diffusion tensor imaging (DTI). DTI is a relatively new *in vivo* MRI technique used to measure the integrity, coherence, and directionality of white-matter fibers, which connect distal structures in the brain. The technique relies on the diffusion of water molecules within myelinated axons and measures the direction of the diffusion.

DTI can provide information about tissue microstructure and architecture for each voxel in the brain. It can also provide information related to the presence and coherence of the brain's white matter. Also, because the main direction of diffusion is linked to the orientation of structures in space, it allows for reconstruction of fiber pathways.

Information provided by DTI has recently been combined with cognitive–behavioral data and fMRI data to explore how the integrity and orientation of white-matter pathways relate to brain activation patterns and cognition (Baird et al. 2005). Knowledge about anatomical connections of distal cortical areas might even provide information about the temporal capacity of connecting fibers between activated foci from an fMRI experiment. This information can indirectly provide clues to the timing of the activation of each node in a cortical network. Whereas fMRI is limited to observations about localized brain activity, DTI can be particularly useful for exploring variations between cortical regions. Other techniques such as independent component analysis and dynamic causal modeling show similar prospects. A more complete understanding of functional connectivity could ultimately make possible better analysis of how brain abnormalities affect brain function and thus human behavior.

A Cautious Step Forward

As we brace for the future, the ideal test of the validity of neuroscience technology, it might be argued, is to predict in advance whether a person with observable abnormalities in legally relevant brain areas is at a specifiable risk of criminal behavior or can conform to the law. This inevitable goal may raise an entirely new set of moral and philosophical questions about how the law ought to regulate the behavior of innocent people. Therefore, this prospect should be approached by the scientific community with judicious reserve.

Neuroscience is much more limited in the kinds of conclusions it can support than the public, the legal system, and many neuroscientists would like to acknowledge. As we progress, many scientists and lawyers will undoubtedly make claims that are not warranted by the neuroscientific data. Like any new science, neuroscience is vulnerable to abuse. For these reasons, neuroscientists will, and ought to be, burdened with the responsibility not only of generating data but also of criticizing and thwarting those abuses. This dual role for neuroscientists is imperative if neuroscience is to have a positive effect on law.

Neuroscience has copious challenges to undertake before becoming a reliable benefit to courtroom procedures. As a case in point, this review has only touched on how neuroscience can inform determinations of intention and insanity, but a thorough understanding of the brain might also one day help us to inform legal determinations of the many other mental states relevant to criminal law, such as recklessness, negligence, duress, and automatism. There are also many areas within civil law procedures in which neuroscience will undoubtedly play an increasing role (Tovino 2007). Until such a thorough understanding is reached, stepwise advance in our theoretical models, experimental manipulations, and measurement technology should ultimately contribute to the overarching goal of bringing specificity to the problems that result from the application of neuroscience to questions of legal responsibility and exculpation.

Acknowledgments

We thank Jim Blascovich, Adina Roskies, Larry Crocker, Karl Doron, Emily Murphy, Suzanne Gazzaniga, Scott Grafton, Annie Wertz, Galia Aharoni, Matthew Green, Teneille Brown, Craig Bennett, and our reviewers for their invaluable feedback, as well as the MacArthur Foundation for its generous support.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience, 27*, 3743–3752.

Baird, A. A., Colvin, M. K., VanHorn, J. D., Inati, S., & Gazzaniga, M. S. (2005). Functional connectivity: integrating behavioral, diffusion tensor imaging and functional magnetic resonance imaging datasets. *Journal of Cognitive Neuroscience, 17*, 687–693.

Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Blair, R. J., Jones, L., Clark, F., & Smith, M. (1995). Is the psychopath 'morally insane'? *Personality and Individual Differences, 19*, 741–752.

Blair, R. J., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: a lack of responsiveness to distress cues? *Psychophysiology, 34*, 192–198.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *Journal of Neuroscience, 27*(34), 9141–9145.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839–1849.

Chomsky, N. (1975). *Reflections on Language*. New York, Pantheon.

Commonwealth v. Pirela, 929 A.2d 629 (Pa. 2007).

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. *Neuroimage, 15*(2), 373–385.

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4*, 423–431.

Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., et al. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron, 22*, 615–621.

Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. *Consciousness and Cognition, 12*, 609–619.

Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S. et al. (in press). The angular gyrus computes action awareness representations. *Cerebral Cortex*.

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: an fMRI investigation of deontic reasoning. *NeuroImage, 28*, 778–786.

Fine, C., & Kennett, J. (2004). Mental impairment, moral understanding and criminal responsibility: psychopathy and the purposes of punishment. *International Journal of Law and Psychiatry, 27*, 425–443.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Gazzaniga, M. S. (2005). *The Ethical Brain*. New York, Dana Press.

Gazzaniga, M. S., & Steven, M. S. (April 2005). Neuroscience and the law. *Scientific American Mind*, 42–49.

Grafton, S., Sinnott-Armstrong, W. P., Gazzaniga, S. I., & Gazzaniga, M. S. (December 2006/January 2007). Brain scans go legal. *Scientific American Mind*, 30–37.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science, 6*, 517–523.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 359*, 1775–1785. Also in S. Zeki & O. Goodenough, ed., *Law and the Brain*. New York: Oxford University Press.

Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia, 40*, 212–222.

Halgren, E., & Marinkovic, K. (1996). General principles for the physiology of cognition as suggested by intracranial ERPs. In C. Ogura, Y. Koga & M. Shimokochi (Eds.), *Recent Advances in Event-Related Brain Potential Research* (pp. 1072–1084), Amsterdam, New York: Elsevier.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, E. E. (2007). Reading hidden intentions in the human brain. *Current Biology, 17*(4), 323–328.

Hoshi, E., Sawamura, H., & Tanji, J. (2005). Neurons in the rostral cingulate motor area monitor multiple phases of visuomotor behavior with modest parametric selectivity. *Journal of Neurophysiology, 94*, 640–656.

Hoshi, E., & Tanji, J. (2004). Differential roles of neuronal activity in the supplementary and presupplementary motor areas: from information retrieval to motor planning and execution. *Journal of Neurophysiology, 92*, 3482–3499.

Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, J. M., et al. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences USA, 101*, 13335–13340.

Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.

Kansas v. Wilburn, 822 P.2d 609, 615 (1991).

Kennett J., & Fine, C. (2008). Internalism and the evidence from psychopaths and "acquired sociopaths." In W.

Sinnott-Armstrong (ed.) *Moral Psychology, Volume 3: The Neuroscience of Morality* (pp. 173–190). Cambridge, MA: MIT Press.

Kiehl, K. (2008). Without morals: The cognitive neuroscience of criminal psychopaths. In W. Sinnott-Armstrong (ed.) *Moral Psychology, Volume 3: The Neuroscience of Morality* (pp. 119–149). Cambridge, MA: MIT Press.

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience, 7*, 1299–1301.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*, 908–911.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science, 303*, 1208–1210.

Levenston, G. K., Patrick, C. J., Bradley, M. M., & Lang, P. J. (2000). The psychopath as observer: emotion and attention in picture processing. *Journal of Abnormal Psychology, 109*, 373–385.

Libet, B. (2004). *Mind Time*. Cambridge, MA: Harvard University Press.

Mele, A. (2006). *Free Will and Luck*. New York: Oxford University Press.

M'Naghten's Case, 10 Clark, & F 200, 8 Reprint 718 (1843).

Model Penal code § 2.01 (1962).

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (October 2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience, 6*, 799–809.

Moore, J. & Haggard, P. (in press). Awareness of action: Inference and prediction. *Consciousness and Cognition*.

Morse, S., & Hoffman, M. (2007). The uneasy entente between insanity and mens rea: beyond *Clark v. Arizona*. *University of Pennsylvania Law School. Scholarship at Penn Law*. Paper 143.

Nachev, P., Rees, G., Parton, A., Kennard, C., & Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Current Biology, 15*, 122–128.

Nachev, P., Wydell, H., O'Neil, K., Husain, M., & Kennard, C. (2007). The role of pre-supplementary motor area in the control of action. *NeuroImage, 36*, T155–T163.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge, UK: Cambridge University Press.

Perkins, R. N. (1969). *Perkins on Criminal Law* (2nd ed.). New York: Foundation Press.

Pinker, S. (2002). *The Blank Slate*. New York: Penguin Putnam, Inc.

Price, C. J., Mummery, C. J., Moore, C. J., Frakowiak, R. S. J., & Friston, K. J. (1999). Delineating necessary and sufficient neural systems with functional imaging studies of neuropsychological patients. *Journal of Cognitive Neuroscience, 11*, 371–382.

Raine, A. (1989a). Evoked potential models of psychopathy: a critical evaluation. *International Journal of Psychophysiology, 8*, 29–34.

Raine, A. (1989b). Evoked potentials and psychopathy. *International Journal of Psychophysiology, 8*, 1–16.

Robinson, P. H. 2 Criminal Law Defenses § 173 (West, 2007)

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and "top-top" control of action in cognitive experiments. *Psychological Research, 68*, 189–198.

Roper v. Simmons, 543 U.S. 551, 125 S.Ct. 1183, 1200, 161 L.Ed.2d 1, 28 (Mo. 2005).

Sakai, K., Hikosaka, O., Miyauchi, S., Sasaki, N., Fujimaki, N., & Putz, B. (1999). Presupplementary motor area activation during sequence learning reflects visuo-motor association. *Journal of Neuroscience, 19*, 1–6.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*, 235–239.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43*, 1391–1399.

Sinnott-Armstrong, W. (Ed.) (2008). *Moral Psychology, Volume 3: The Neuroscience of Morality*. Cambridge, MA: MIT Press.

Sirigu, A., Daprati, E., Sophie, C., Giraux, P., Nighoghossian, N., Posada, A., et al. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience, 7*, 80–84.

Sumner, P., Nachev, P., Morris, P., Peters, A. M., Jackson, S. R., Kennard, C., et al. (2007). Human medial frontal cortex mediates unconscious inhibition of voluntary action. *Neuron, 54*, 697–711.

Tovino, S. A. (2007). Functional neuroimaging and the law: trends and directions for future scholarship. *American Journal of Bioethics, 7*, 44–56.

U.S. v. Hinckley, 525 F. Supp. 1342 (D.D.C. 1981), clarified, 529 F. Supp. 520 (D.D.C. 1982), aff'd, 672 F.2d 115 (D.C. Cir. 1982).

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford, UK: Oxford University Press.

Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Weichselbaum, S. (2004, May 1). Killer again beats death sentence. *Philadelphia Daily News*.

Wolbers, T., Schoell, E. D., Verleger, R., Kraft, S., McNamara, A., Jaskowski, P., et al. (2006). Changes in connectivity profiles as a mechanism for strategic control over interfering subliminal information. *Cerebral Cortex, 16*, 857–864.

Wolf, S. R. (1990). *Freedom Within Reason*. New York: Oxford University Press.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences USA, 104*, 8235–8240.