THE NEW ATLANTIS
A Journal of Technology & Society

Neuroimaging and Capital Punishment

O. Carter Snead

Can brain scans be used to determine whether a person is inclined toward criminality or violent behavior?"

This question, asked by Senator Joseph Biden of Delaware at the hearing considering the nomination of John Roberts to be Chief Justice of the United States, illustrates the extent to which cognitive neuroscience—increasingly augmented by the growing powers of neuroimaging, the use of various technologies to directly or indirectly observe the structure and function of the brain—has captured the imagination of those who make, enforce, interpret, and study the law. Judges, both state and federal, have convened conferences to discuss the legal ramifications of developments in cognitive neuroscience. Numerous scholarly volumes have been devoted to the subject. The President's Council on Bioethics convened several sessions to discuss cognitive neuroscience and its potential impact on theories of moral and legal responsibility. Civil libertarians have expressed suspicion and concern that the United States government is using various neuroimaging techniques in the war on terrorism. Personal injury lawyers have urged the use of functional neuroimaging to make "mild to moderate brain [and nervous] injuries…visible [to] jurors"—and members of the civil defense bar have, not surprisingly, published articles criticizing the reliability of such evidence and arguing that it should be inadmissible. Criminal defense attorneys have likewise expressed a strong interest in using neuroimaging evidence to help their clients.

The attraction of the legal community to cognitive neuroscience is by no means unreciprocated. Cognitive neuroscientists have expressed profound interest in how their work might impact the law. Michael Gazzaniga, one of the field's leading lights—and in fact the man who coined the term "cognitive neuroscience"—predicted in his 2005 book *The Ethical Brain* that advances in neuroscience will someday "dominate the entire legal system."

Practitioners of cognitive neuroscience seem particularly drawn to the criminal law; more specifically, they have evinced an interest in the death

---

O. Carter Snead *is an associate professor of law at the University of Notre Dame and a fellow at the Ethics and Public Policy Center.* *This essay is adapted from an article that appeared in the* New York University Law Review.

penalty. Indeed, a well-formed cognitive neuroscience project to reform capital sentencing has emerged from their work in the courtroom and their arguments in the public square. In the short term, cognitive neuroscientists seek to invoke cutting-edge brain imaging research to bolster defendants' claims that, although legally guilty, they do not deserve to die because brain abnormalities diminish their culpability. In the long term, cognitive neuroscientists aim to draw upon the tools of their discipline to embarrass, discredit, and ultimately overthrow retribution as a distributive justification for punishment. The architects of this effort regard retribution as the root cause of the brutality and inhumanity of the American criminal justice system, generally, and the institution of capital punishment, in particular. To replace retribution, they argue for the adoption of a criminal law regime animated solely by the forward-looking (consequentialist) aim of avoiding social harms. This new framework, they hope, will usher in a new era of what some have referred to as "therapeutic justice" for capital defendants, which is meant to be more humane and compassionate. But in fact, despite these humanitarian intentions, the project's aspirations for capital sentencing reform would more likely exacerbate the draconian and brutal features of the present capital sentencing regime.

The Era of Cognitive Neuroscience

Cognitive neuroscience is an investigational field that seeks to understand how human sensory systems, motor systems, attention, memory, language, higher cognitive functions, emotions, and even consciousness arise from the structure and function of the brain. "The overwhelming question in neurobiology," as Francis Crick and Christof Koch put it in a 1992 *Scientific American* article, is "the relation between the mind and the brain." Cognitive neuroscience has been described as a "bridging discipline"—between biology and neuroscience, on the one hand, and cognitive science and psychology, on the other.

Interest and activity in the field exploded in the early 1990s, when advances in imaging technology made studying the brain far easier and less invasive than before. Prior to such advances, many scientists had focused their inquiries on animal and computational models rather than on live human subjects. The advent of neuroimaging has led to an enormous proliferation of scholarship; over the past five years, according to one estimate, an average of one thousand peer-reviewed scholarly articles based on neuroimaging have been published *each month.*

Broadly speaking, neuroimaging techniques can be divided into two categories. First, "structural" or "anatomical" neuroimaging is limited to the observation of the brain's architecture. Computed tomography (CT) scanning and magnetic resonance imaging (MRI) were both introduced in the 1970s. The former uses x-rays and a computerized algorithm; the latter measures the signal strengths of the various radio frequencies emitted by the proton nuclei of atoms in brain tissue when the protons are placed in a strong magnetic field. Because MRI has superior spatial resolution, it has largely supplanted CT scanning.

The second category of neuroimaging techniques, "functional" neuroimaging, involves the construction of computerized images that measure the brain's activity (and sometimes also its structure). The oldest of these techniques, electroencephalography (EEG), dates back to 1929; it involves the use of electrodes on the scalp to measure the electrical activity of the area of the brain below. A related technique, magnetoencephalography (MEG), measures the magnetic fields produced by brain activity. Positron emission tomography (PET) and single-photon emission computed tomography (SPECT), developed more recently, involve the injection of small amounts of radioactive compounds into the bloodstream; these enter the brain after approximately thirty seconds and, as they begin to decay, can be detected and observed. This allows for the production of images that map the distribution of the tracer compounds throughout the brain—an indication of the localization of the brain's metabolic activity. Researchers infer that the areas with the highest observed metabolic activity are the regions of greatest brain activation at a given time.

Both PET and SPECT have in recent years been eclipsed by functional magnetic resonance imaging (fMRI), a technique that requires no radioactive injections but instead uses MRI technology to detect the concentration of oxygenated blood in particular parts of the brain. Researchers interpret the increase in blood flow to a particular brain region (indicated by an increase in magnetic resonance signal strength) as an increase in cellular activity in that particular region.

Especially with the aid of these neuroimaging techniques, the focus of cognitive neuroscience has expanded from an inquiry into basic sensorimotor and cognitive processes to the exploration of more highly complex behaviors. Over the past decade, scientists have increasingly turned their attention to the neurobiological correlates of behavior and to the links between their science and vexed matters of public policy. Their efforts are motivated largely by the view that, in the words of prominent University of California, San Diego "neurophilosophy" professor Patricia Smith Churchland, "as we

understand more about the details of the regulatory systems in the brain and how decisions emerge in neural networks, it is increasingly evident that moral standards, practices, and policies reside in our neurobiology."

Cognitive neuroscientists have thus brought their tools to bear on contested moral and ethical—and, by extension, legal and political—questions, including the moral status of the human embryo, the brain function of patients diagnosed as minimally conscious or persistently vegetative, and the definition of "brain death." Furthermore, a number of peer-reviewed articles have addressed the cognitive neuroscience of personality traits such as introversion, extroversion and neuroticism, pessimism, empathy, disposition towards cooperation or competition, novelty seeking, harm avoidance, and reward dependence.

Scientists are also investigating the neural mechanisms of emotion, including aversion to unpleasant scenes and social cues (e.g., facial expressions). There is a growing body of neuroimaging research on social attitudes and preferences, including racial attitudes (as well as the cerebral processes involved in making judgments about "similar and dissimilar others"), sexual attraction (as well as the neural mechanisms employed in suppressing such attraction), and even political predilections. Neuroscientists have also explored the neurological dimensions of moral decision-making and religious experiences.

Thus, it is hardly surprising that recent neuroimaging studies have touched on matters, including the detection of deception and the roots of both impulsive and premeditated criminal violence, that could have forensic applications in criminal justice. Several of these recent developments approach what might be thought of—and indeed, are sometimes portrayed in the popular press—as "mind reading." Using neuroimaging techniques developed to study high-level vision, scientists have been able to reliably discern the type of image being viewed or imagined by a research subject, based solely on the pattern of activity in the brain. One recent neuroimaging study purported to identify the "covert goals" that a subject intended to perform (specifically, either the addition or subtraction of two given numbers).

There are numerous practical, technical, epistemic, and interpretive complexities associated with neuroimaging specifically and cognitive neuroscience in general, and while they cannot all be addressed here, it is nevertheless useful to acknowledge and review the most commonly raised objections. One concern is that, given the profound obstacles to isolating, controlling, and studying cognitive processes, it is quite difficult to show a conclusive relationship of necessity between a particular brain region's function and any associated cognitive process. Relatedly, there is the usual

experimental difficulty of distinguishing causation from mere correlation. Concerns have also been raised about the notion of functional specialization of brain regions, a premise that is central to neuroimaging studies. That is, certain brain regions may serve multiple cognitive functions or, alternatively, multiple cognitive functions may activate the same region of the brain. This increases the risk of error in drawing inferences from neuroimaging data about the brain, the mind, and behavior. This difficulty is compounded because the most common forms of neuroimaging depend on proxies (such as blood flow) to serve as indirect measures of regional brain activity. Thus, the measurements derived from these techniques are necessarily attenuated from the ultimate object of interest—namely, cognitive function.

These concerns, in turn, have led to concerns about the interpretation of the data generated by neuroimaging. As neuroscientist Martha J. Farah and social psychiatrist Paul Root Wolpe observed in the *Hastings Center Report*, "Although brainwaves do not lie, neither do they tell the truth; they are simply measures of brain activity." Furthermore, the array of expertise required to produce a single neuroimage (i.e., neuroscience, computational theory, physics, computer science, statistical analysis, and nuclear medicine) presents numerous opportunities for technical error. There is also a lack of standardization among the machines and laboratory procedures used in the field. The wide variability in brain physiology among experimental subjects and the concomitant difficulties in defining normalcy also make it difficult to draw meaningful comparisons. Thus, while it is among the most powerful new tools available, fMRI is rarely used for diagnostic applications and has not yet become part of standard practice for clinicians.

An additional concern is that the use of cognitive neuroimaging data to diagnose psychological conditions relies entirely on the soundness of the diagnostic criteria—which, given the absence of specific biological markers for any psychiatric disorder, can be hotly contested. Moreover, as some scholars have noted, this research raises complicated cultural and anthropological questions. Concepts which are integral to the interpretation of cognitive neuroimaging, such as "personhood," "self," and "consciousness," can vary widely from culture to culture. This only adds further complexity to the analysis of data acquired in such studies. Finally, there is a worry that people will ignore the foregoing technical and interpretive complexities in a rush towards practical application, especially given the current enthusiasm about the potential of neuroimaging to provide seemingly objective and transparent insight into morally and socially relevant behaviors.

Still, assuming for the sake of argument that this new discipline someday acquires the technical capacities its proponents hope for and

expect, the potential social implications are striking. In the words of Judy Illes, director of Stanford University's Program in Neuroethics, this research may ultimately yield the possibility of using neuroimaging when "assess[ing] the truthfulness of statements and memory in law, profiling prospective employees for professional and interpersonal skills, evaluating students for learning potential[,] . . . selecting investment managers[,] . . . and even choosing lifetime partners based on compatible brain profiles for personality, interests, and desires."

The Mind as Brain

The foundational premise of cognitive neuroscience is that all aspects of the mind are ultimately reducible to the structure and function of the brain. As Joshua Greene and Jonathan Cohen have described it, cognitive neuroscience is the "understanding of the mind *as* brain"; it seeks, in the words of Martha J. Farah, to provide "comprehensive explanations of human behavior in purely material terms." Put simply, the fundamental premise of cognitive neuroscience is "reductive materialism": it is "reductive" in that it seeks to explain the "macrophenomena" of thought and action solely in terms of the "microphenomena" of the physical brain, and it is "materialist" in that it postulates that human thought and behavior are caused solely by physical processes taking place inside the brain—a three-pound bodily organ of staggering complexity, but a bodily organ nonetheless. In this way, cognitive neuroscience follows the dominant approach of modern science, which seeks to understand and explain all observable phenomena as functions of their component parts. Under this methodology, questions of biology are thought to be reducible to matters of chemistry, which are, by extension, reducible to problems of physics. In principle, this approach will ultimately lead to the analysis of all phenomena in terms of the relationships of motion and rest among their most elemental particles.

In defense of reductive materialism in neuroscience, proponents cite evidence connecting changes in the brain to changes in the mind. The most well-known example of this principle is the nineteenth-century case of Phineas Gage, a law-abiding railway worker who was radically changed into a callous, unreliable troublemaker after an accident in which an iron tamping rod was accidentally driven through his brain. As Harvard experimental psychologist Steven Pinker put it in his 1997 book *How the Mind Works*:

> Another problem [with arguments against materialism] is the overwhelming evidence that the mind is the activity of the brain. The supposedly immaterial soul, we now know, can be bisected with a

> knife, altered by chemicals, started or stopped by electricity, and extinguished by a sharp blow or by insufficient oxygen. Under a microscope, the brain has a breathtaking complexity of physical structure fully commensurate with the richness of the mind.

Reductive materialism is a widely accepted approach among neuroscientists. Michael Gazzaniga told a reporter in 2006 that, in his estimation, "98 or 99 percent" of cognitive neuroscientists share a commitment to reductive materialism in seeking to explain mental phenomena. This near-universal commitment to using material causation to explain the mind and human behavior carries with it profound implications for perennial concepts such as the existence of the soul, free will, selfhood, and consciousness. As Francis Crick put it in his 1994 book *The Astonishing Hypothesis*, "your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules."

To be sure, cognitive neuroscientists (and the philosophers who invoke their research) do not universally agree that materialist accounts of human behavior should wholly alter or displace traditional concepts. One could fill many volumes in an effort to give a responsible account of the debates as they have unfolded. On the issue of free will, some have adopted the posture ("hard determinism") that the reduction of all mental processes to physical events renders the notion of uncaused choice unintelligible, while others (sometimes called "compatibilists") adhere to the view that reductive materialism still leaves a limited amount of room for free choice. But apart from these disagreements, those in the field are of the shared opinion that the findings of cognitive neuroscience compel a deep reevaluation of the philosophical concepts lying at the root of our most weighty moral, ethical, and political decisions. Given the obvious link between free will and personal responsibility, the necessity of reevaluating free will looms large over the aspirations of the cognitive neuroscience project for capital sentencing.

Neuroscience, the Law, and Criminal Violence

Developments in neuroimaging have affected the law both directly and indirectly. The indirect developments are visible in the great deal of discussion that has occurred about speculative applications of nascent technological innovations. The direct impact has occurred where neuroimaging evidence has been introduced in courtrooms and has led to the creation of a body of decisional law that has shaped the legal landscape in this domain.

In an essay in the 2004 collection *Neuroscience and the Law*, Stanford law professor Henry T. Greely provides an excellent account of the *speculative* uses of neuroimaging in the legal context. He suggests that such technology might eventually be used in the courtroom to detect lies or to compel truth, to determine bias (on the part of jurors, witnesses, or parties), to elicit or evaluate memory, to determine competency (e.g., to stand trial, to be executed, or to make medical decisions), to prove the presence of intractable pain, to prove addiction (or susceptibility thereto), to show a disposition to sexual deviance or predatory impulses (for purposes of involuntary civil commitment), or to show future dangerousness.

As for *actual* applications, neuroimaging evidence has been proffered and admitted in a variety of jurisdictions, in both civil and criminal cases, and for a variety of purposes. However, it is difficult to analyze the use of neuroimaging in actual litigation. Many cases in which neuroimaging evidence is introduced may be unreported. Still others may be resolved through informal means, such as settlement or plea agreement. To the extent that such cases can be identified, it is often impossible to reliably discern the role that neuroimaging evidence played in the outcome.

In the civil context, neuroimaging has been proffered and admitted to prove actual harm (and, to a lesser extent, causation) in personal injury cases involving toxic exposure, claims under the National Vaccine Act, head injuries, and medical malpractice. In a recent suit by a video-game-industry trade association to enjoin the enforcement of Illinois laws that restricted the sale of violent and sexually-explicit video games, a federal district court admitted fMRI evidence to show a relationship between playing violent video games and aggressive behavior in children; fMRI evidence was tendered in support of the government's argument that it had a compelling state interest in regulating violent games. In contract disputes, neuroimaging has been admitted—and has been found persuasive by fact finders—to show that one of the parties lacked sufficient cognitive capacity to form a valid contract.

In the criminal context, defendants have proffered neuroimaging evidence at various stages of the process for a variety of purposes. For instance, courts have admitted neuroimaging evidence (or have held that a defendant was entitled to undergo neuroimaging tests) in connection with claims of mental incompetence. Defendants have had mixed success in seeking to admit neuroimaging evidence to show diminished capacity (or an inability to formulate requisite *mens rea*—criminal intent) at the guilt phase of criminal trials or as an adjunct to their insanity defenses. The most famous example of neuroimaging being used in an insanity defense

is the case of John Hinckley, Jr., who attempted to assassinate President Ronald Reagan in 1981. There, the court admitted a CT scan to show that Hinckley's brain had atrophied, which the defense argued—over the vigorous objection of the government's expert—was evidence of organic brain disease.

Defendants have enjoyed the greatest success with neuroimaging evidence at the sentencing phase of capital trials in connection with mitigation claims. In support of these claims, experts have cited evidence from a massive (and growing) body of scientific literature on both the neuroanatomical and neurochemical bases for various types of violence. In 1998 and 1999, an interdisciplinary group of experts was convened under the auspices of the Aspen Neurobehavioral Conference to create a consensus statement on the relationship between the mind, the brain, and violence. To this end, they conducted an exhaustive literature survey of the role of the brain in violent behavior and issued a statement in 2001 noting that the limbic system (structures deep inside the brain) and the frontal lobes (the front-most sections of the cerebrum) "are thought to play preeminent roles in [violent] behavior." The statement asserted that:

> Aggressive behavior has been thought to arise from the operations of the limbic system under certain circumstances, and the amygdala is the structure most often implicated.... [P]refrontal functions may... provide an individual with the capacity to exercise judgment in the setting of complex social situations in which actions have significant consequences. In many cases, this capacity for judgment may serve the important function of inhibiting limbic impulses, which, if acted on, could be socially inappropriate or destructive.... Therefore, there exists a balance between the potential for impulsive aggression mediated by temporolimbic structures and the control of this drive by the influence of the orbitofrontal regions.

This theory of violence was informed, and has been reinforced, by neuroimaging studies. The first such study was published in 1994 by University of Southern California neuroscientist Adrian Raine, who used PET to illustrate diminished activity of the prefrontal cortex of individuals accused of murder. In many of Raine's subsequent works, the model of violence sketched out in the consensus statement figures prominently.

Other articles surveying the neuroimaging literature similarly affirm the widespread association of prefrontal dysfunction and violence. Further literature reviews and articles written by prominent neuroscientists have reached similar conclusions. And in addition to the iconic case of Phineas

Gage, there are striking modern examples of the relationship between frontal lobe injuries (or dysfunction) and a disposition to criminal violence. For example, neurologist Jonathan Pincus, in his 2001 book *Base Instincts*, tells of a Georgia man named Louis Culpepper who, following a concussive injury to his prefrontal cortex, found himself no longer able to restrain his impulses to molest his five-year-old stepdaughter. In a similar case, reported by Nicholas Thompson in a 2006 *Legal Affairs* article, a school teacher with no criminal record and a stable marriage found himself unable to restrain his impulses to view child pornography, solicit sex, and make sexual overtures to his stepdaughter. Physicians discovered an egg-sized tumor growing in his prefrontal lobe; once it was removed, his inhibitions and capacity for self-restraint were restored. Another recent example is Andrew Laing, who, as his mother described to the London *Daily Mail* in 2006, lost all sexual inhibitions and sense of propriety following a concussive injury to his prefrontal lobe in a skiing accident.

A significant and growing area of research concerns the neurobiological correlates of psychopathy and antisocial personality disorder (APD). Psychopathy, as succinctly defined by University of Virginia law professor John Monahan, is "a cluster of personality traits including manipulativeness, lack of empathy, and impulsivity." APD is a related diagnostic construct listed in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders; it is based on behavioral characteristics such as "a pervasive pattern of disregard for, and violation of, the rights of others that begins in childhood or early adolescence and continues into adulthood."

Neuroimaging studies relating to APD and psychopathy reveal a striking connection between abnormal brain activity and psychopathy—although, as the authors of a recent survey of the past decade's worth of papers on the subject concluded, more study is needed "before conclusions can be drawn" about the connection. Adrian Raine's research, meanwhile, has led him to tentatively conclude that a structural deficit in subjects with APD "may underlie the low arousal, poor fear conditioning, lack of conscience, and decision-making deficits that have been found to characterize antisocial, psychopathic behavior." He has also found abnormally high white matter volume in the corpus callosa (the band of tissue connecting the left and right hemispheres of the brain) of "psychopathic antisocial individuals." Raine has speculated that this abnormality might impair interhemispheric communication in a way that bears on the affective deficits typical of psychopaths.

Many other prominent neuroscientists likewise have undertaken inquiries using neuroimaging tools to explore the potential connection between

brain abnormalities and violence. By linking brain abnormalities to specific behaviors—and, specifically, to violent behavior—these studies provide a foundation for the use of neuroimaging evidence in criminal trials.

Aiding Capital Defendants

The short-term aim of those seeking to apply cognitive neuroscience to capital sentencing is straightforward: to bolster defendants' mitigation claims with neuroimaging research that demonstrates a biological disposition to criminal violence. To fully appreciate this goal, it is necessary first to understand the procedural context in which it is pursued. While the precise procedures vary from state to state, virtually all capital sentencing regimes direct the jury to evaluate mitigating and aggravating factors in considering whether to impose the death penalty. The consideration of mitigating evidence is central to the constitutional requirement of individualized sentencing. The Supreme Court has held since *Lockett v. Ohio* (1978) that defendants enjoy wide latitude in their presentation of mitigating evidence bearing on "any aspect of [the] defendant's character or record and any of the circumstances of the offense that the defendant proffers as a basis for a sentence less than death." It is typical for defense experts to testify about the mitigating effects of mental illness or brain damage in an attempt to persuade jurors that a defendant is less than fully culpable and should receive a sentence of life imprisonment rather than death. It is within this procedural framework that cognitive neuroscientists have sought to wield their tools on behalf of defendants facing the death penalty.

Reported cases and public commentary demonstrate that cognitive neuroscientists are increasingly contributing to the mitigation efforts of capital defendants. A significant group of neuroimaging experts has aided capital defendants in constructing their mitigation cases, testifying and using SPECT and PET scans on the behalf of defendants at trial. Some experts have also conducted studies that dovetail with the needs of capital defendants. For example, Raine and Mount Sinai School of Medicine professor Monte Buchsbaum were coauthors of the first PET-scan-based study of the brain function of murderers. Raine has eloquently defended the use of his research as a mechanism for persuading juries (and society more broadly) that capital defendants should receive life sentences rather than the death penalty. Numerous other neuroimaging practitioners work on behalf of defendants at the sentencing phase of capital trials, using EEGs and other methods to supplement their own testimony. Other cognitive neuroscientists who have not personally testified in capital trials have expressed

support for the use of neuroscience testimony in this and related contexts.

When testifying on behalf of a capital defendant, neuroscientists (and other practitioners of neuroimaging) generally argue that, although it does not provide an excuse for purposes of legal guilt, dysfunction in the violence-inhibitory mechanisms of the defendant's brain sufficiently diminishes his moral responsibility such that he deserves a sentence of life imprisonment rather than death. In support of these claims, such experts invoke cutting-edge neuroimaging research on the biological correlates of criminal violence.

Perhaps the most high-profile example of neuroimaging in the capital context is *Roper v. Simmons* (2005), in which the Supreme Court entertained a challenge—under the Eighth Amendment's injunction against cruel and unusual punishment—to a state law permitting the execution of juveniles who were under the age of eighteen at the time they committed a capital offense. Among the numerous amicus briefs submitted, two in particular—one led by the American Psychological Association (APA) and the other led by the American Medical Association (AMA)—captured the public's imagination. Both made novel use of neuroimaging-based evidence to anchor their arguments that adolescents were categorically less morally blameworthy than adults and, as a result, not deserving of the ultimate criminal sanction of death.

According to the briefs, neuroimaging research suggests that adolescents' behavioral immaturity is due, in large measure, to the "anatomical immaturity of their brains." The briefs cite structural and functional neuroimaging studies showing that the neocortical regions of the brain, which are believed to be responsible for risk assessment, impulse control, and high-level cognition, are not yet fully developed in adolescents. Conversely, those subcortical areas of the brain from which impulsivity and violence are thought to arise are fully developed in adolescents and, indeed, are *more* active in teenagers than in adults. Specifically, the AMA brief points to research showing that the limbic system—part of the brain associated with "primitive impulses of aggression, anger and fear"—is overactive in the brains of adolescents. At the same time, the frontal lobes—"the regions of the brain associated with impulse control, risk assessment, and moral reasoning"—are still developing in adolescents and are insufficiently mature to mediate and check the influence of the limbic system.

Both the AMA and APA briefs cite neuroimaging studies showing that the adolescent prefrontal cortex has not yet completed two important processes necessary to its full function: myelination and pruning. "Myelination" is the process by which the axons ("neural fibers that use electrical impulses

to carry information across long distances") are insulated, strengthening and reinforcing their connections and "thereby greatly speeding up the communication between cells, allowing the brain to process information more efficiently and reliably." "Pruning" is the process by which the volume of the brain's gray matter (composed of neurons) is thinned, thus strengthening the connections among the neurons that remain and improving their function. Studies suggest that late in childhood there is a new proliferation of gray matter in the prefrontal cortex, which is then gradually pruned in a process that does not conclude until after adolescence.

The APA brief includes an extensive argument that the exclusion of adolescents from capital punishment should be categorical because, given adolescents' unfixed personal characteristics due in large part to their still-developing brains, the mechanisms of capital sentencing are not sufficient to assess their individual culpability. Capital jurors are thus incapable of accurately weighing the relevant aggravating and mitigating factors in the balancing process required by virtually every jurisdiction that retains the death penalty.

The APA and AMA briefs appeared to have real influence on the Court's consideration of *Roper*. At oral argument, sixteen of the twenty questions asked of the lawyer representing Simmons (one of the killers, then on Death Row) concerned the scientific evidence presented in the two briefs. Moreover, in the opinion itself—which affirmed the Missouri Supreme Court's conclusion that applying the death penalty to juveniles runs afoul of the Eighth Amendment—the Court's reasoning seemed animated by arguments raised in the briefs. Writing for the majority, Justice Kennedy agreed with the briefs' arguments that, because adolescents have a temporarily diminished capacity for sound decision-making and personal restraint, sentencing them to death violates the basic principle of retributive justice on which capital punishment is grounded. Justice Kennedy reasoned that juveniles are less blameworthy principally because their disposition to criminal violence is due to "transient immaturity" rather than "irreparable corruption." The risk that jurors might mistake the former (a mitigating circumstance) for the latter (an aggravating circumstance) makes the individualized sentencing required by the Eighth Amendment impossible when the offender is a juvenile.

The neurobiological theory of violence set forth by the *Roper* amici—with its focus on frontal lobe impairment—fairly represents the capital mitigation arguments used by cognitive neuroscientists generally. As Michael Gazzaniga has put it, those who represent criminal defendants "are looking for that one pixel in their client's brain scan that shows…a malfunction in

the normal inhibitory networks," which would allow them to demand leniency on the grounds that their client could not control his actions.

How has the short-term aim of cognitive neuroscientists fared to date? Capital defense attorneys, encouraged by some successes, now present evidence of frontal lobe dysfunction as mitigation evidence during the sentencing phase; a growing body of scholarly literature encourages the use of such evidence. In fact, some courts have even held that the failure to allow neuroimaging evidence to be introduced at the sentencing phase of a trial constitutes reversible error. At least one court has granted a defendant funds to conduct neuroimaging during a capital trial. This is not surprising; courts can be very permissive when it comes to admitting evidence for purposes of capital mitigation. To be sure, the presentation of neuroimaging evidence doesn't always bring the intended result—juries have often been presented with neuroimaging evidence and nevertheless imposed or recommended a sentence of death—but there is every reason to believe the practice will only grow more routine.

Overthrowing Retribution

Beyond the short-term goal of affecting capital sentencing, the cognitive neuroscientists who are active in this area also have a longer-term, more fundamental aspiration for criminal justice: They aim to work a radical conceptual revision of criminal punishment itself. More specifically, they seek to use the premises and tools of neuroscience—and neuroimaging in particular—to embarrass, undermine, and ultimately overthrow retributive justice as a principle of punishment. Once retribution is discredited, they contend, criminal law will be animated solely by its proper end: namely, the purely forward-looking, consequentialist goal of avoiding socially harmful behavior. This new approach, it is hoped, will usher in a regime of "therapeutic justice," wherein criminal defendants will be treated more humanely.

The most comprehensive articulation and defense of this long-term aspiration for criminal punishment reform was advanced in two papers published in 2004—one by coauthors Joshua Greene and Jonathan Cohen,[1] the other by Robert Sapolsky.[2] Greene and Cohen (now professors in Harvard's and Princeton's psychology departments, respectively) argue that advances in cognitive neuroscience—enabled by neuroimaging—will

1 Joshua Greene and Jonathan Cohen, "For the Law, Neuroscience Changes Nothing and Everything," *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, no. 1451 (November 29, 2004) 1775-1785.

2 Robert M. Sapolsky, "The Frontal Cortex and the Criminal Justice System," *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, no. 1451 (November 29, 2004) 1787-1796.

ultimately demonstrate that "ordinary conceptions of human action and responsibility" are false. "As a result, the legal principles we have devised to reflect these conceptions may be flawed" and must be radically overhauled and replaced with principles that are grounded in a neuroscientific view of the truth about free will and human agency. The primary focus of their critique is the principle of retributive justice—which, they assert, "depends on an intuitive, libertarian notion of free will that is undermined by science."

In defense of this thesis, Greene and Cohen first rehearse the familiar dichotomy of justifications for criminal punishment. *Consequentialism* regards punishment as "merely an instrument for promoting future social welfare" and seeks to prevent "future crime through the deterrent effect of the law and the containment of dangerous individuals"; it is a doctrine that "emerges from the classical utilitarian tradition." By contrast, *retribution* is the principle that "in the absence of mitigating circumstances, people who engage in criminal behavior *deserve* to be punished."

Greene and Cohen then turn to the ancient (yet ongoing) debate over the nature and intelligibility of free will. They articulate a tripartite typology of positions on the issue: hard determinism, libertarianism, and compatibilism. Hard determinism, as the name implies, rejects the concept of free will. It holds that free will is fundamentally incompatible with the premise that all human action can be sufficiently explained by material causes that are necessarily bound by the laws of physics and previous events. Libertarianism (not to be confused with political libertarianism) accepts the claim that free will and determinism are incompatible but nevertheless concludes that the world is not, in fact, completely determined by the laws governing the motion and rest of matter. Compatibilism, meanwhile, holds that material determinism and free will are reconcilable.

Greene and Cohen argue that insofar as advances in neuroscience have begun to reveal the purely material causes of human thought and choice, they have also begun to undermine the fundamental tenets of libertarianism and thus retributive punishment. Libertarianism supplies the strong conception of free will (and thus moral responsibility) on which the doctrine of retribution relies. Greene and Cohen argue, however, that the strength of the concept of free will posited by libertarianism arises from its claim to operate through a nonmaterial mechanism—a proposition increasingly at odds with modern science. They contend that ultimately neuroimaging will entirely undermine the anti-materialist foundations of the libertarian position on free will, thus removing the grounding necessary for just deserts. Moreover, they argue that retributive justice is conceptually irreconcilable with hard determinism: if all actions are sufficiently determined by

material causes beyond anyone's control, the notions of culpability and just deserts upon which retribution depends are unintelligible.

Meanwhile, compatibilism's modest account of free will, Greene and Cohen argue, is not sufficiently robust to support the exacting demands of retribution. While some have argued that the law is constructed with compatibilism in mind—accepting at least the minimal capacity for rational action—what society really cares about, Greene and Cohen contend, is whether the defendant is responsible in a richer sense. That is, even if the defendant is shown to be minimally rational in a legal sense, citizens will still ask whether it was "really him" who committed the crime, or whether it was "his upbringing," "his genes," "his circumstances," or "his brain" that were truly responsible. These questions, Greene and Cohen argue, arise from a libertarian vision of free will that does not accept compatibilism but rather is animated by a *dualist* premise that the brain and the mind are distinct (though interacting) entities. Thus, while the law as written may be formally compatiblist, it is actually driven by the "libertarian moral intuitions" of the citizens who implement it.

Greene and Cohen characterize this tension between the law's formal requirements and society's richer conception of free will as an unstable "marriage of convenience." They predict that neuroimaging will force a crisis in this union: cognitive neuroscience (aided by neuroimaging) will ultimately show that there is no difference between "him" and "his brain"—thus proving that the foundations of the libertarian dualist intuitions about human agency are untenable.

Greene and Cohen's analysis applies especially well to the context of capital sentencing, wherein the Supreme Court (as described above) has construed the Constitution to require the consideration of all mitigating factors relevant to a criminal defendant's culpability. The very doctrine of mitigation is driven by questions like those that Greene and Cohen argue society "really" cares about, such as "was it him," or was it "his brain," "his upbringing," or his "circumstances?" While these questions have little or no bearing on earlier trial stages—they are not necessary for establishing legal guilt—they bear significantly on the kind of punishment imposed on the legally guilty. So it would seem that capital sentencing is largely driven by a metaphysically ambitious conception of human agency—one that is at odds with the conception that animates our determinations of guilt and innocence.

According to Greene and Cohen, only the libertarian understanding of free will can provide adequate support to the principle of retributive justice. But they predict—indeed, hope—that cognitive neuroscience will shatter

this foundation. They note that while philosophical arguments against free will have not proven persuasive to the general population, science supported by neuroimaging will succeed where philosophy has failed:

> Arguments are nice, but physical demonstrations are far more compelling. What neuroscience does, and will continue to do at an accelerated pace, is elucidate the 'when,' 'where,' and 'how' of the mechanical processes that cause behavior. It is one thing to deny that human decision-making is purely mechanical when your opponent offers only a general, philosophical argument. It is quite another to hold your ground when your opponent can make detailed predictions about how these mechanical processes work, complete with images of the brain structures involved and equations that describe their function.

Greene and Cohen argue that when and if the notion of human agency is shown to be illusory, societal attitudes may well change. Eventually the law of punishment will have to follow suit and reflect the newly revealed truths about free will. Once society internalizes the lessons of cognitive neuroscience as they bear on moral (and thus criminal) responsibility, the principle of retribution—relying as it does on a false understanding of human agency—will be eliminated as a legitimate general or distributive justification for punishment.

And this, they think, is all to the good. Greene and Cohen assert that retributivism is largely responsible for the "counter-productive" state of the American penal system. They advance consequentialism as the sole legitimate justification for punishment. Without free will—and hence, without retribution—punishment can be fashioned solely with the future benefits to society in mind. Criminal offenders can still be held "responsible" for their actions, but without the moral stigma and judgment that retributive justice implies. Sentencing promotes beneficial effects for society by deterring future harms and incapacitating only those who would visit such harms upon the polity. Greene and Cohen would preserve excuse defenses (such as insanity and duress) for those cases where it can be shown that the deterrence of such offenders would not be effective. But retribution would be laid to rest forever as a pernicious fiction based on the "illusion" of free will.

Joining Greene and Cohen in their criticisms of retributive justice, Stanford neuroscientist Robert Sapolsky notes that "at a logical extreme, a neurobiological framework may indeed eliminate blame," but adds that the institution of criminal punishment is still necessary for the purpose of protecting society from future harms. "To understand is not to forgive

or to do nothing," Sapolsky writes. "Whereas you do not ponder whether to forgive a car that, because of problems with its brakes, has injured someone, you nevertheless protect society from it." Human beings, in this depiction, are essentially analogous to mechanical devices. Sapolsky shares Greene and Cohen's desire to shed a framework that implicitly regards criminal defendants as morally blameworthy, preferring a consequentialist system even though it adopts an arguably diminished understanding of human personhood. "It may seem dehumanizing to medicalize people into being broken cars," he admits, "but it can still be vastly more humane than moralizing them into being sinners." Sapolsky, Greene, Cohen and others engaged in this effort see no irony in the pursuit of such humane dehumanization.

Practitioners of neuroimaging whose work already contributes directly or indirectly to the short-term project of aiding capital defendants with mitigation claims have also embraced these long-term aspirations. For example, Vickie Luttrell (a psychologist with neuroscience training) and Jana Bufkin (a criminologist), both of Drury University, reached conclusions similar to those of Cohen, Greene, and Sapolsky following their 2005 review of seventeen neuroimaging studies of criminal violence. Retributive justice, they write in an article in the journal *Trauma, Violence, and Abuse*, accounts for factors "with no inherent explanatory worth" that "are summoned to justify less-than-stellar community-level interventions and unproductive institutionalization." They believe that the new findings of cognitive neuroscience should steer society away from retribution and towards a regime of "therapeutic justice" in which offenders will be held to "scientifically rational and legally appropriate degree[s] of accountability." They also believe that neuroimaging research will ultimately lead to the refinement and improvement of the instruments used for the classification and prediction of violent criminal behavior.

The long-term goal of overthrowing retributive justice is very much in the spirit of late-eighteenth-century thinkers such as Jeremy Bentham and Cesare, Marquis of Beccaria, who regarded punishment of the guilty as justified only insofar as it was instrumental to the protection of society and the promotion of human happiness. It also mirrors, in many respects, the work of Barbara Wootton, Baroness of Abinger. Lady Wootton, a twentieth-century criminologist, rejected the notion of criminal "punishment" altogether, arguing instead that the only intelligible goal for the criminal law is to be a "system of purely forward-looking social hygiene in which our only concern when we have an offender to deal with is with the future and the rational aim of prevention of further crime." This view led

Wootton to argue for a complete abandonment of *mens rea* as an element of guilt in favor of a system of strict criminal liability. She believed that a person's intentions at the time of a crime are not knowable and, indeed, not relevant to the question of guilt. (It is worth noting that Greene and Cohen disagree with Wootton here—they are confident that someday the reasons for antisocial choices will become discernible through the techniques of neuroscience—even while they fundamentally share her view of the aims of criminal law.) A defendant's mental state, to Wootton, would only be relevant as a predictive instrument to be used in preventing the same defendant from offending in the future. Under her approach, the state would take custody of an offender upon his conviction for a criminal act and give him medical treatment or incarcerate him. Wootton's approach blurs the distinction between prisons and hospitals: Both are "places of safety" where "offenders will receive the treatment which experience suggest [sic] is most likely to evoke the desired response" of preventing future crime. Wootton's framework thus explicitly and intentionally conflates punishment with therapy.

Mitigation and Just Deserts

On the surface, the long- and short-term aims of the cognitive neuroscience project for capital punishment share clearly humanitarian ambitions: namely, success in helping convicted capital defendants persuade jurors and judges not to impose a sentence of death, and the ultimate creation of a more compassionate and humane legal regime for such defendants. Unfortunately, it seems likely that the criminal regime desired by cognitive neuroscientists would, tragically and ironically, prove far harsher and less humane for capital defendants than the current system.

Why? Simply put, the project, taken as a whole, is utterly at war with itself. Its short-term aim relies on a particular theory of mitigation that is firmly grounded in retribution—a principle whose foundations are explicitly rejected by the architects of the cognitive neuroscience project against capital punishment. Conversely, the project's long-term aim is devoted to dismantling the doctrinal foundation upon which the short-term aspiration depends. Thus, the success of its long-term goal would necessarily defeat the project's short-term goal. Worse still, the extant mechanisms that the long-term project would explicitly leave in place—those features of the capital sentencing framework animated solely by the consequentialist goal of avoiding societal harms—constitute arguably the gravest threat to a capital defendant's life. If the capital sentencing regime were remade according

to the aspirations of the long-term plan, this threat would be dramatically amplified precisely *because of* the research of cognitive neuroscientists.

Consider the context, discussed above, in which cognitive neuroscientists seek to implement their short-term aim: the mitigation phase of capital sentencing. Mitigation involves the presentation of evidence regarding the character, background, or other pertinent features of an already convicted defendant that might convince the jury that the defendant's degree of culpability merits life imprisonment rather than death. However, defendants who reach the sentencing phase have, by necessity, already satisfied the prerequisite legal thresholds for sanity, competence, and the capacity to formulate the relevant *mens rea*. Mitigation evidence is presented to inspire leniency in spite of a prior finding of legal guilt.

This strategy, however, is squarely rooted in a distributive theory of punishment that proponents of the use of cognitive neuroscience in capital sentencing explicitly repudiate as a principal source of the irrationality and brutality that plague the current system. Paul H. Robinson, law professor at the University of Pennsylvania, has called this theory "punishment according to desert," as it is an approach that distributes punishment "according to the offender's personal blameworthiness for the past offense, which takes account not only of the seriousness of the offense, but also the full range of culpability, capacity, and situational factors that we understand to affect an offender's blameworthiness."

The Supreme Court's death penalty jurisprudence confirms that the concept of mitigation grows directly out of the requirements of retributive justice. So does the AMA's amicus brief in *Roper*—a powerful illustration that the short-term aspiration is driven entirely by an appeal to the culpability-mitigating effects of the defendant's neurological condition. And it is clear that Justice Kennedy's opinion in *Roper* was likewise principally animated by concerns about just deserts: "Retribution is not proportional if the law's most severe penalty is imposed on one whose culpability or blameworthiness is diminished, to a substantial degree, by reason of youth and immaturity." Kennedy also observed that, given their diminished capacity for self-control and risk assessment, it was "unclear" whether the death penalty would have a sufficient deterrent effect on potential juvenile offenders. In this way, Justice Kennedy's doubts about deterrence (a key component of the consequentialist principle justifying the death penalty as punishment) further bolstered his more emphatic arguments that retributive justice categorically requires sparing adolescents from the ultimate punishment.

Justice Kennedy's majority opinion likewise echoed the arguments

made in the AMA and APA briefs about the inadequacy of individualized capital sentencing as a safeguard against error and abuse in the capital context. Indeed, Kennedy went further, holding that juries would be incapable of treating youth as a mitigating factor on a case-by-case basis:

> An unacceptable likelihood exists that the brutality or cold-blooded nature of any particular crime would overpower mitigating arguments based on youth as a matter of course.... In some cases a defendant's youth may even be counted against him. In this very case... the prosecutor argued Simmons' youth was aggravating rather than mitigating.

Cognitive neuroscientists who invoke neuroimaging evidence for purposes of capital mitigation embrace the strategy outlined in the AMA brief and adopted in Justice Kennedy's opinion. But this approach trades on the very dichotomy of "him" versus "his brain" that just deserts invites—one that proponents of the long-term aspiration deplore as unintelligible. Thus, the short-term aspiration depends on precisely the principle of punishment that the long-term approach rules out of bounds.

Prediction and Prevention

Conversely, the long-term aspiration of cognitive neuroscience for capital punishment seeks to undermine and destroy the very distributive principle of retributive justice upon which its short-term counterpart depends. Proponents of the long-term goal regard just deserts as anathema to the only suitable goal of the criminal law—preventing future criminal harms. The long-term aspiration would thus preclude the introduction of mitigation evidence that bears on diminished culpability. It would leave in place only those mechanisms that promote the avoidance of crime. The mechanisms of capital sentencing best suited to this end are those that are calibrated to *predict* the social harms to be contained or avoided.

Nothing in capital sentencing embodies the purely consequentialist spirit of the long-term cognitive neuroscience project as much as the commonly invoked aggravating factor of "future dangerousness." Prosecutors seeking the death penalty bear the burden of persuading jurors beyond a reasonable doubt that at least one aggravating factor exists to make the defendant death-eligible. As Peter T. Hansen, a capital defense expert, put it in a 1992 article, this is the stage of the trial where prosecutors "suggest to the jury that the defendant is a living hazard to civilization and a menacing threat to society." To this end, prosecutors submit the testimony of experts or laypersons regarding a defendant's future dangerousness or

simply argue it themselves based on a variety of evidence, or they try to establish it through cross-examination of the defense's mitigation experts.

There are two principal scientific approaches to assessing future dangerousness. The first is clinical prediction, which relies on the judgment of experts (such as psychologists and psychiatrists) or laypersons (such as police or probation officers) to evaluate the defendant as an individual. The second is actuarial (or statistical) prediction, which evaluates defendants according to "explicit rules specifying which risk factors are to be measured, how those risk factors are to be scored, and how the scores are to be mathematically combined to yield an objective estimate of violence risk." Among experts in the field, actuarial methods are thought to be significantly more reliable than clinical methods, though there are commentators who argue that all predictive efforts are insufficiently reliable to be permitted in capital sentencing.

Because the rules of evidence that govern criminal trials often do not apply to capital sentencing hearings, courts have wide latitude in deciding whether to admit evidence of future dangerousness at such proceedings. In one capital case in Illinois, the court admitted as evidence of the defendant's future dangerousness that he had two tattoos of the Grim Reaper and another of a cross surrounded by skulls, and that his e-mail address was "Cereal Kilr 2000." In a case in Texas, the court allowed the testimony of law enforcement officers "derived from their observations of [the] defendant, about that defendant's character and the likelihood of future violence." In a Virginia case, the court admitted evidence that the defendant had been cruel to animals twenty years earlier. In some cases, clinicians have been permitted to testify even where they have not examined the defendant.

Prosecutors regularly invoke diagnoses of psychopathy or antisocial personality disorder in capital sentencing, likely because both are highly correlated with recidivist violence. Courts have specifically permitted both diagnoses to be introduced as evidence of future dangerousness at the sentencing phase of capital trials. This has proven to be a highly effective strategy for prosecutors given that the diagnostic criteria for each sound to the lay juror essentially like a straightforward description of "irreparable corruption" (to borrow Justice Kennedy's phrase from *Roper*). Either diagnosis can have a devastating effect on the defendant's mitigation claims and can create an expectation in jurors' minds that rehabilitation is impossible. The APD diagnosis, in fact, has been dubbed "the kiss of death."

Diagnoses of APD and psychopathy have played a prominent role as aggravating factors in the capital context. Dr. James Grigson, an iconic and notorious figure in the jurisprudence of future dangerousness, serves

as an extreme but illustrative example of how government experts sometimes wield their power to make these diagnoses. In over 140 cases, Dr. Grigson—nicknamed "Dr. Death"—testified to the effect that the defendant had severe APD and was thus very likely to be violent in the future—often without ever having examined the defendant. In the seminal case of *Barefoot v. Estelle*, he testified with "reasonable psychiatric certainty" that Thomas Barefoot, the convicted murderer, fell in the "most severe category" of sociopaths and that he would, with "*one hundred percent and absolute*" certainty, commit future criminal acts, constituting a continuing threat to society.

Studies have shown that capital juries often regard evidence of future dangerousness as the most important aggravating factor in their sentencing calculus. In fact, it has been observed that even in those jurisdictions that do not explicitly direct the capital jury to consider future dangerousness as an aggravating factor, jurors do so anyway.

The Costs of Repudiating Retribution

In the context of sentencing, desert and dangerousness inevitably conflict. "To advance one, the system must sacrifice the other," as Paul Robinson has observed. "The irreconcilable differences reflect the fact that prevention and desert seek to achieve different goals. Incapacitation concerns itself with the future—avoiding future crimes. Desert concerns itself with the past—allocating punishment for past offenses." This tension is played out in dramatic fashion in capital cases. On the one hand, capital defendants introduce mitigating evidence to diminish their moral culpability, thus seeking a final refuge in the concept of retribution. On the other, the prosecution tenders evidence of future dangerousness, trying to stoke the consequentialist fears of the jury about violent acts that the defendant might commit if he is not permanently incapacitated by execution. In capital sentencing, pure consequentialism is the gravest threat to the defendant's life, while appeals to retributive justice are often his last best hope.

The long-term aspiration of cognitive neuroscience decisively resolves this conflict between desert and crime control in favor of the latter by removing any consideration of diminished culpability. In so doing, the long-term scheme eliminates the last safe haven for a capital defendant whose sanity, capacity for the requisite *mens rea*, competence, and guilt are no longer at issue. Thus, in a final ironic twist, once retribution is replaced with a regime single-mindedly concerned with the prediction of crime and the incapacitation of criminals, the only possible use in capital

sentencing of neuroimaging research on the roots of criminal violence is to demonstrate the aggravating factor of future dangerousness.

Imagine for a moment how a jury concerned solely with avoiding future harms would regard an fMRI or PET image that purported to show the biological causes of a non-excusing disposition to criminal violence. Likely, neuroimaging would radically amplify, in the minds of jurors, the aggravating effect of a diagnosis of APD or psychopathy. In a sentencing system that focused the jury's deliberation solely on the question of identifying and preventing crime, the work of the cognitive neuroscience project's architects would be transformed from a vehicle for seeking mercy into a tool that counsels the imposition of death.

It is only through the lens of just deserts that such evidence could possibly be regarded as mitigating. In the regime contemplated by the long-term aspiration—where claims of diminished culpability are untenable—the only permissible inference that jurors can draw is that the defendant is "'damaged goods' and beyond repair," as one capital defense expert has put it. Arguing for compassion or leniency in such a system would be as nonsensical as (to return to Sapolsky's metaphor) seeking mercy for a dangerously defective car on its way to the junkyard to be crushed into scrap metal. Reconciliation and forgiveness are not useful concepts as applied to soulless machines; they are only intelligible as applied to sinners.

The grave implications of the long-term aspiration for capital sentencing come into even sharper relief when one considers the role that retributive justice has played in modern death penalty jurisprudence. Contrary to the intuitions of the project's architects, retribution has served as a crucial *limiting* principle on capital sentencing. The Supreme Court itself has referred (in the 2002 case *Atkins v. Virginia*) to a "narrowing jurisprudence" of just deserts, which "seeks to ensure that only the most deserving of execution are put to death." In the name of retributive justice, the Court has barred the execution of mentally retarded defendants, defendants who were under the age of eighteen when their offense was committed, rapists, and defendants convicted of felony murder who did not actually kill or attempt to kill the victim. In each instance, the Court ruled that such defendants were not eligible for the death penalty because such punishment would be categorically disproportionate to their personal culpability. These same results could not have been reached if deterrence were the sole animating principle guiding the Court.

One might be tempted to think that without the engine of retribution, the political will to continue a regime of capital punishment will wither away. In fact, the average voter's desire for retribution has been blamed

for the continued existence of capital punishment in the United States. But a review of recent political rhetoric on the death penalty contradicts this conclusion, suggesting that politicians prefer to couch their public arguments in terms of deterrence. For example, in one of the presidential debates of the 2000 election, Vice President Al Gore and Governor George W. Bush agreed that the only reason to support the death penalty was for its deterrent effects:

> **Moderator:** Do both of you believe the death penalty actually deters crime?…
>
> **Bush:** I do. It's the only reason to be for it.…I don't think you should support the death penalty to seek revenge. I don't think that's right. I think the reason to support the death penalty is because it saves other people's lives.
>
> **Gore:** I think it is a deterrent. I know that's a controversial view, but I do believe it's a deterrent.

More recently, in a 2007 *USA Today* article describing the recent efforts on the part of various states to expand the death penalty to a wider array of offenses, *every politician quoted* cited deterrence as the sole motivation for the initiatives. Empirical studies purporting to show the deterrent effect of the death penalty have had a profound impact on the Supreme Court and the lay public by casting the death penalty as a life-saving institution. Finally, and perhaps most powerfully, social science evidence shows that the aggravating factor of "future dangerousness" is the second-most decisive consideration (next to the fact of the underlying crime itself) for jurors contemplating the imposition of a death sentence. Thus, in actual deliberations, it is nearly certain that jurors privilege the question of deterrence above almost all other factors.

In fact, the widely shared intuition that seems to be motivating the long-term aspiration—namely, that retributive justice is the primary source of the brutality and harshness of the modern American criminal justice system—may generally be misguided. As Paul Robinson has argued, "the harshness of the current system may be attributed in largest part to the move to rehabilitation, incapacitation, and deterrence, which *disconnected criminal punishment from the constraint of just desert.*" Many features of the criminal justice system that are frequently criticized as draconian and inhumane are, in fact, motivated by a purely consequentialist crime-control rationale. Such measures include laws that authorize life

sentences for recidivists (such as the "three strikes" laws), laws that reduce the age at which offenders can be tried as adults, laws that punish gang membership, laws that require the registration of sex offenders, laws that dramatically increase sentences by virtue of past history, and, most paradigmatically, laws that provide for the involuntary civil commitment of sexual offenders who show difficulty controlling their behavior. These laws are the progeny of the principle animating the long-term aspiration of those wielding neuroscience in capital cases, and some are worrisome examples of its possible implications.

Robinson points to the possibility that "if incapacitation of the dangerous were the only distributive principle, there would be little reason to wait until an offense were committed to impose criminal liability and sanctions; it would be more effective to screen the general population and 'convict' those found dangerous and in need of incapacitation." Indeed, the short-term project—using cognitive neuroscience to identify the roots of criminal violence—may someday create novel and powerful opportunities to interfere with individual liberty.

Questions of whether a given individual poses a continuing threat to society are central to the criminal justice system. In addition to capital sentencing, fact-finders are charged with making such determinations in the context of non-capital sentencing, civil commitment hearings, parole and probation hearings, pretrial detention, and involuntary civil commitment of sexual offenders. Regardless of neuroimaging's capacity or incapacity to predict criminal behavior reliably, there is already a powerful demand for the use of such techniques in crime control (a possibility occasionally explored in science fiction). Moreover, far more controversial methods for predicting future social harms have already been accepted by the Supreme Court in the capital sentencing context. This problem would be dramatically aggravated by adopting a criminal framework that places an even higher premium on the prediction and prevention of violence than the present one does.

A Fundamental Conflict

The architects of the cognitive neuroscience project against capital punishment might reasonably raise several possible rejoinders in response to these concerns. First, they may defend the use of neuroimaging evidence for mitigation by attempting to ground the concept of mitigation in the consequentialist value of deterrence rather than in just deserts. That is, they could argue that capital punishment cannot deter individuals with certain types of brain abnormalities, so presenting evidence of

these abnormalities merely shows that there is no deterrence rationale for imposing a death sentence on such individuals. In fact, Greene and Cohen take this very approach by arguing in favor of retaining certain excuse defenses (such as diminished capacity, infancy, insanity, and duress) because there is no deterrence value in punishing people in circumstances where such excuses apply.

But this argument fails to provide an alternative, consequentialist justification for the short-term aspiration's theory of mitigation for at least two reasons. First, the argument is a "spectacular *non sequitur*," as the late Oxford legal philosopher H. L. A. Hart noted when analyzing an essentially identical dispute between Jeremy Bentham and William Blackstone. The argument purports to claim that offenders with a brain defect or personality disorder are not deterrable by the *threat* of the death penalty. It is possible that, as Hart pointed out, the *infliction* of the punishment might nevertheless "secure a higher measure of conformity to law on the part of normal persons than is secured by the admission of excusing conditions," thus decreasing the overall amount of crime. In addition, it is obvious that a death sentence will be effective as a *specific deterrent* on the convicted offender.

But an attempt to ground the short-term aspiration's theory of mitigation in this way fails for a more fundamental reason: the law of capital sentencing does not presently accept the proposition that defendants who are afflicted by the kinds of conditions that cognitive neuroscientists invoke in mitigation (such as hypoactivity of the prefrontal lobe, APD, or psychopathy) are undeterrable. Sentencing law does not accept (nor do the mitigation experts argue) that these conditions make it impossible for a defendant to appreciate his actions, conform with the law, or form the requisite *mens rea*. If the law did accept this claim, such defendants would prevail at the guilt stage of their trials and would not face sentencing.

It does not seem possible, under present capital sentencing categories, to characterize the theory of mitigation invoked by cognitive neuroscientists under the short-term aspiration as anything other than an argument for diminished culpability rooted in the overarching distributive principle of just deserts. That being so, the long-term and short-term aspirations, as argued above, remain at loggerheads.

Next, those who would apply cognitive neuroscience in capital cases might retort that the project is aimed at rehabilitation and not mere incapacitation—that is, it aspires to bring about a regime of "therapeutic justice." But this argument, too, fails to rescue the effort, both for principled and prudential reasons. First, it is not clear how a regime devoted solely

to avoiding future harms to society compels the pursuit of rehabilitation for offenders. It cannot be because neurologically-defective defendants *deserve* a punishment less than death due to their diminished culpability, since this would be an appeal to retribution. Moreover, for offenders with the neurological conditions associated with a predisposition to criminal violence, rehabilitation may very well prove impossible. Future prospects of rehabilitation for those laboring under one such condition—psychopathy—seem quite bleak, according to the conventional wisdom of the mental health community. (Indeed, precisely therein lies its persuasive force as an aggravating factor, as Justice Kennedy observed in *Roper*.) As Robinson has observed, once just deserts has been eliminated as a legitimate distributive principle of punishment and therapies for criminal misconduct have proven unavailing, societies often turn to incapacitation as the sole animating value of the criminal law.

Finally, cognitive neuroscientists might argue that the prediction of a draconian, inhumane, and invasive system of criminal law will not come to pass as a result of their efforts because the public would not stand for it; the revulsion and fear that citizens would feel toward such a system would simply not be tolerated. Greene and Cohen make this argument in response to the concern that their vision for the criminal law would produce massive over-punishing. They argue, by way of example, that society would not tolerate the inhumanity of executing parking-law violators.

To be sure. But that example has no bearing on the cognitive neuroscience project for *capital sentencing*. Capital defendants are arguably the most hated and feared members of society; it is unlikely that there would be substantial social discontent with a regime that was geared toward their permanent incapacitation. For better or worse, there is not now widespread opposition to the execution of defendants who are factually and legally guilty, sane, competent, and with the requisite *mens rea*, yet who present neuroimaging evidence suggesting that they have a neurological condition or personality disorder that inclines them (if not irresistibly) towards murderous acts. The continued successful invocation of APD and psychopathy as aggravating factors by prosecutors, and the substantial number of defendants claiming frontal lobe dysfunction who have nevertheless been sentenced to death, illustrates that our society is (again, for better or worse) comfortable executing such individuals.

Besides, even if there were widespread societal discomfort with executing those whose brains make them incorrigibly violent, this would not prevent such executions under Greene and Cohen's framework. Rather, this anxiety could be explained away as a function of the irrational dichotomy between

"him" and "his brain" that the cognitive neuroscience project deplores. As Greene and Cohen argue, accepting an appealing fiction (such as free will) may be useful for decisions relating to small things, just as having faith in Euclid and Newton is safe when negotiating the aisles of the grocery store. But in matters as important as capital sentencing, Greene and Cohen would argue that we must have the courage to abandon such fictions and embrace the truths of cognitive neuroscience—just as we must turn to Lobachevsky and Einstein when we want to launch a rocket into space.

In the end, while the goals of making capital sentencing more rational and humane are laudable, the cognitive neuroscience project to do so is ill-conceived. Because that project's short- and long-term aims are intractably opposed to one another, its impact likely will be the opposite of what it seeks. It is unclear whether it is possible (or desirable) to salvage the project in a way that will preserve its humanitarian ends. It may be that the reductive materialist account of human personhood and human agency posited by cognitive neuroscience—and, indeed, by modern science more generally—is fundamentally incompatible with the account on which the criminal law is premised. This should lead us to question the assumptions underlying cognitive neuroscience no less than those underlying the criminal law. If we examine both through the lens of our humanitarian aspirations, we are likely to discover that the wisdom behind our laws fares a good deal better than we imagined against the assumptions (often masquerading as hard facts) behind the new science of the brain. Not surprisingly, dehumanization turns out not to be humane.