

THE FUTURE OF THE CONSTITUTION

December 28, 2010



Reuters/Ho New

Cognitive Neuroscience and the Future of Punishment

O. Carter Snead



INTRODUCTION



The sentencing phase of the trial, by contrast, had proven far more difficult to bear. The prosecutor had described in excruciating detail the murders themselves in an effort to show that they were especially "heinous, atrocious, and cruel, manifesting exceptional depravity." The prosecutor and counsel for the defense each recounted the details of the defendant's life and character. His broken childhood, marked by unspeakable abuse and neglect. His years of drug and alcohol use. His spotty and unstable employment history. His history of using violence to impose his will and pursue his interests. They even discussed the structure and function of his brain — with reference to an array of colorful poster-board sized images showing diminished activity in the prefrontal cortex (the seat of reasoning, self-restraint, long term planning) and above-average activity in his limbic system (the more primitive part of his brain, associated with fear and aggression). Relying on a raft of neuroimaging studies, the prosecutor argued that the pattern of activation and structural abnormalities in the defendant's brain were consistent with "low arousal, poor fear conditioning, lack of conscience, and decision-making deficits that have been found to characterize antisocial, psychopathic behavior." He further argued that this was not a temporary condition — it was permanent and unlikely to be correctable by any known therapeutic intervention. The prosecutor argued that, taken together, this was the profile of an incorrigible criminal who would certainly kill again if given the chance. The defense argued, to the contrary, that the evidence did not point to any tangible future risk of violence.

The judge then explained to the jurors that they must decide unanimously what punishment was fitting for the crime of conviction: life without parole or a sentence of death. Among other things, the judge explained that "before the death penalty can be considered, the state must prove at least one statutorily-defined aggravating circumstance beyond a reasonable doubt" and that the aggravating factors outweigh all of the mitigating factors. These he described as "any fact or circumstance, relating





O. Carter Snead is an associate professor of law at the University of Notre Dame.

to the crime or to the defendant's state of mind or condition at the time of the crime, or to his character, background or record, that tends to suggest that a sentence other than death should be imposed."

The judge looked up from his jury instructions and turned towards the jury box. "Ladies and gentlemen, let me add a word of caution regarding your judgment about mitigating factors. Some of you may be tempted to ask yourselves 'Was it really the defendant that did this? Or was it his background? Or his brain?' You might be tempted to ask yourselves 'What does this defendant deserve in light of his character, biology, and circumstances?' Some of you might even be tempted to argue to your fellow jurors that 'this man does not *deserve* the ultimate punishment in light of his diminished (though non-excusing) capacity to act responsibly borne from a bad past and a bad brain; capital punishment in this case is *disproportionate* to the defendant's moral culpability." The judge's eyes narrowed and he leaned even farther forward. "But, Ladies and gentlemen of the jury, you must not ask such questions or entertain such ideas. The sole question before you, as a matter of law, is much narrower. The *only* question you are to answer is this: is this defendant likely to present a future danger to others or society? You should treat every fact that suggests that he does present such a danger as an aggravating factor; every fact suggesting the contrary is a mitigating factor. Matters of 'desert,' 'retributive justice,' or proportionality in light of moral culpability are immaterial to your decision. Ladies and gentlemen, this is the year 2040. Cognitive neuroscientists have long ago shown that 'moral responsibility,' 'blameworthiness,' and the like are unintelligible concepts that depend on an intuitive, libertarian notion of free will that is undermined by science. Such notions are, in the words of two of the most influential early proponents of this new approach to punishment, 'illusions generated by our cognitive architecture.' We have integrated this insight into our criminal law. Punishment is not for meting out 'just deserts' based on the fiction of moral responsibility. It is simply an instrument for promoting future social welfare. We impose punishment solely to prevent future crime. And this change has been for the better. As another pioneer of the revolution in punishment — himself an eminent cognitive neuroscientist wisely wrote at the beginning of the twenty-first century: 'Although it may seem dehumanizing to medicalize people into being broken cars, it can still be vastly more humane than moralizing them into being sinners.' So, please ladies and gentlemen of the jury. Keep your eye on the ball, and do not indulge any of the old and discredited notions about retributive justice." With that, the judge adjourned and dismissed the jury so that it could begin its deliberations.



The above hypothetical is obviously fanciful. But it borrows concepts and arguments directly from a current debate that has been unfolding alongside the advent of extraordinary advances in cognitive neuroscience (particularly as augmented by revolutionary imaging technology that affords novel ways to examine the structure and function of the brain). Such advances have breathed new life into very old arguments about human agency, moral responsibility, and the proper ends of criminal punishment. A prominent group of cognitive neuroscientists, joined by sympathetic philosophers, lawyers, and social scientists, have drawn upon the tools of their discipline in an effort to embarrass, discredit, and ultimately overthrow retribution as a distributive justification for punishment. The architects of this cognitive neuroscience project 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 criminal defendants, which is meant to be both more humane and more compassionate.

To be sure, not all cognitive neuroscientists subscribe to this program. Indeed, there are many thoughtful voices who raise opposition to this project on various grounds — some prudential and some principled. Whatever one thinks about the cognitive neuroscience project for criminal punishment, however, it deserves to be taken seriously and its arguments should be followed to their ultimate conclusions. This is my aim in the present chapter. In it, I will discuss the contours of the project and explore the radical conceptual challenge that it poses for criminal punishment in America. I will also offer a critique of the project, arguing that jettisoning the notion of retributive justice in criminal punishment will not lead to a more humane legal regime as supporters of the project hope. Rather, by unterthering punishment from moral culpability and focusing entirely on the prediction and prevention of socially harmful behavior, the cognitive neuroscience project eliminates the last refuge of defendants who are legally and factually guilty, but who have diminished culpability owing to some aspect of their character, background, or biology. Indeed, viewed through the lens urged by the cognitive neuroscience project, the only relevance of a non-excusing disposition to criminal behavior is as a justification for incapacitation. The logic of the cognitive neuroscience project could even lead to the embrace of more aggressive use of preventive detention as a solution for categories of criminals that inspire special fears in the polity – including sexual predators and terrorists.

The techniques of cognitive neuroscience are not yet sufficiently developed to support its aspirations. They may never be. But it is always wise to examine the consequences of a nascent moral-technological program *before* it is upon you and in widespread use. My purpose in this chapter is to take seriously the claims of the cognitive neuroscience project so that we may be clear-eyed about its consequences before we consider embracing it.

Cognitive Neuroscience: Premises and Methods

Cognitive neuroscience 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. According to Francis Crick, "the overwhelming question in neurobiology" 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.

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, with the aid of neuroimaging, scientists have increasingly turned their attention to the neurobiological correlates of behavior and to the links between their science and vexing matters of public policy. Their efforts are motivated largely by the view, shared by many in the field, that "[a]s 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."¹ Neuroscientists have explored matters relevant to the criminal justice context, including the detection of deception and the roots of both impulsive and premeditated criminal violence.

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."² Thus, cognitive neuroscience seeks to provide "comprehensive explanations of human behavior in purely material terms."³ Like other disciplines within the modern life and physical sciences, cognitive neuroscience is committed to the premise of *physicalism*, which philosopher of science Alex Rosenberg has noted, is "the assumption that there is only one kind of stuff, substance, or thing in the universe, from matter, material substance, and physical objects all the way down to quarks."⁴ Also like other disciplines within modern science, cognitive neuroscience operates by way of *reduction*. That is, its chief explanatory aspiration is to explicate complex matters in the most simple and elemental terms. It seeks to explain the "macrophenomena" of thought and action solely in terms of the "microphenomena" of the physical brain.⁵



¹ Patricia Smith Churchland, *Moral Decision-making and the Brain, in* NEUROETHICS: DEFINING THE ISSUES IN THEORY, PRACTICE AND POLICY 3, 3 (Judy Illes ed., 2006).

² Joshua Greene & Jonathan Cohen, For the Law, Neuroscience Changes Nothing and Everything, 359 PHIL. TRANSACTIONS ROYAL SOC'Y: BIOLOGICAL SCI. 1775, 1775 (2004) (emphasis added).

³ Martha J. Farah, *Neuroethics: The Practical and the Philosophical*, 9 TRENDS COGNITIVE SCI. 34, 34 (2005). ⁴ ALEX ROSENBERG, DARWINIAN REDUCTIONISM OR HOW TO STOP WORRYING AND LOVE MOLECULAR BIOLOGY, p.2 (2006).

⁵ PATRICIA SMITH CHURCHLAND, BRAIN-WISE: STUDIES IN NEUROPHILOSOPHY 20–21 (2002) ("[A] reduction has been achieved when the causal powers of the macrophenomenon are explained as a function of the physical structure and causal powers of the microphenomenon.").

Combining the axioms of physicalism and reductionism in this way compels a methodological commitment to *mechanism*, namely, the evaluation and explanation of natural phenomena in terms of the structure, actions and interactions of their most basic physical parts. Thus, cognitive neuroscience follows the dominant approach of modern science, namely *reductive mechanism*. It proceeds according to the postulate 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. It seeks explanations within this framework.⁶

Neuroimaging: The Indispensible Tool of Cognitive Neuroscience.

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.

Henry Greely has provided 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 (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.⁷

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. 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 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.

⁶ For a discussion of such limitations and concerns, see Carter Snead, *Neuroimaging and the* "*Complexity*" of Capital Punishment, NYU LAW REVIEW (2007).

⁷ Henry T. Greely, *Prediction, Litigation, Privacy and Property: Some Possible Legal and Social Implications of Advances in Neuroscience, in NEUROSCIENCE AND THE LAW, 127–48.*

Defendants have had mixed success in seeking to admit neuroimaging evidence to show diminished capacity (or an inability to formulate requisite mens rea) 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 argument 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 invoked cutting-edge neuroimaging research on the biological correlates of criminal violence.

The Cognitive Neuroscience of Criminal Violence

The foundation for many uses of neuroimaging evidence in criminal trials lies in a massive and growing body of scientific literature on both the neuroanatomical and neurochemical bases for the various types of violence. In 1998 and 1999, an interdisciplinary group of experts were 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 and the frontal lobes "are thought to play preeminent roles in [violent] behavior."⁸ 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.⁹

This theory of violence was informed, and has been reinforced by, neuroimaging studies. In addition to the iconic case of Phineas Gage, the nineteenth century railroad worker whose personality was altered after an accident drove an iron rod into his head, there are striking modern examples of the

⁸ Id. at 5.

⁹ Id.

relationship between frontal lobe injuries (or dysfunction) and a disposition to criminal violence. For example, following a concussive injury to his prefrontal cortex, Louis Culpepper found himself no longer able to restrain his impulses to molest his five-year-old stepdaughter. In a similar case, following the development of an egg-sized tumor in his prefrontal lobe, 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. Once the tumor was removed, his inhibitions and capacity for selfrestraint were restored. A more recent example is Andrew Laing, who lost all sexual inhibitions and sense of propriety following a concussive injury to his prefrontal lobe in a skiing accident.

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.

The Cognitive Neuroscience Project for Punishment

Over the past decade, certain eminent cognitive neuroscients (along with sympathetic philosophers, social scientists, and lawyers) have argued for a radical conceptual revision of criminal punishment. More specifically, by using the premises and tools of neuroscience—and neuroimaging in particular— they aim 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, a Harvard philosopher, and Jonathan Cohen, a Princeton neuroscientist, and the other by Robert Sapolsky, a Stanford neuroscientist. As Green and Cohen write:

Our penal system is highly counterproductive from a consequentialist perspective . . . and yet it remains in place because retributivist principles have a powerful moral and political appeal. It is possible, however, that neuroscience will change these moral intuitions by undermining the intuitive, libertarian conceptions of free will upon which retributivism depends.

• • • •

... At this time, the law deals firmly but mercifully with individuals whose behavior is obviously the product of forces beyond their control. Some day, the law may treat all convicted criminals this way. That is, humanely.¹⁰

Greene and Cohen argue that advances in cognitive neuroscience—enabled by neuroimaging—will ultimately demonstrate that "ordinary conceptions of human action and responsibility" are false. "[A]s 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.¹¹ 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."¹²

In defense of this thesis, Greene and Cohen first reprise the familiar dichotomy of consequentialism ("which emerges from the classical utilitarian tradition")¹³ and retribution as both the general and distributive justifications for criminal punishment. They define consequentialism as a doctrine that regards punishment as "merely an instrument for promoting future social welfare"¹⁴ and that seeks to prevent "future crime through the deterrent effect of the law and the containment of dangerous individuals."¹⁵ By contrast, they define retribution as advocating the principle that "in the absence of mitigating circumstances, people who engage in criminal behaviour *deserve* to be punished."¹⁶

Greene and Cohen then turn to the ancient 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, as characterized by Greene and Cohen, 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. In contrast, compatibilism holds that material determinism and free will are reconcilable, though compatibilism's conception of free will is more metaphysically modest than that of libertarianism.

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



¹⁰ Greene & Cohen, *supra* note 2, at 1783–84.

¹¹ Id. at 1775.

¹² Id. at 1776.

¹³ Id.

¹⁴ *Id*. at 1775.

¹⁵ *Id.* at 1776. ¹⁶ *Id.*

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 at increasingly odds with modern science. They contend that ultimately neuroimaging will entirely undermine the antimaterialist foundations of the libertarian position on free will, thus removing the grounding necessary for just deserts. Moreover, it is evident that retributive justice is conceptually irreconcilable with hard determinism: if all actions are sufficiently determined by material causes to be beyond anyone's control, the notions of culpability and just deserts upon on which retribution depends are unintelligible.

Greene and Cohen additionally assert that compatibilism's modest account of free will is not sufficiently robust to support the exacting demands of retribution, either as a general aim or as a distributive principle. They describe a compatibilist vision of free will, one that defines free will as the minimal capacity for rational action—namely, the ability to produce "behaviour that serves [one's] desires in light of [his] beliefs."¹⁷ As they describe, Stephen Morse has eloquently argued that the law is constructed with this minimalist conception of free will in mind. Law, Morse maintains, is compatibilist and thus is not threatened by any proof of determinism that neuroimaging may eventually offer. In support of his claim, Morse points to the criminal law, which refuses to excuse from guilt those defendants who are laboring under a defect of mind, so long as they satisfy a minimal cognitive and volitional threshold. Greene and Cohen respond that while the law may formally focus on the question of minimal rationality, what people in society really care about is whether the defendant is responsible in a richer sense one rooted in libertarian conceptions of free will. 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 the materialist determinism of Morse's compatabilism 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 (as Morse contends) formally compatabilist, 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. This forms the basis for their belief that Morse underestimates the transformative power of neuroimaging on the law and especially on criminal punishment.

Cognitive Neuroscience and the Future of Punishment

¹⁷ Id. at 1778.

It is far beyond the scope of this chapter to try to resolve this rich and vexed dispute about the future impact of neuroimaging on the law as a whole. Nevertheless, it is necessary to briefly reflect on how this debate plays out in the context of capital sentencing. Morse is certainly correct that the legal standard for diminished capacity for the purpose of determining legal guilt is modest; defendants rightly can be characterized as needing only "minimal rationality" in order to be held fully accountable for their actions. This is borne out by the small percentage of cases in which defendants raise the defense of legal insanity and the even smaller portion of cases in which such a defense succeeds. But in the context of capital sentencing, which is animated by a particularly rich and textured conception of moral responsibility, Greene and Cohen's analysis is especially accurate. The Supreme Court has construed the Constitution to require the consideration of all mitigating factors relevant to a criminal defendant's culpability in meting out capital punishment. The very doctrine of mitigation is driven by questions like those that Greene and Cohen argue society "really" cares about, such as "[w]as it him," or was it "his brain," "his upbringing," or his "circumstances?" Morse is right that these questions do not currently bear on legal guilt, but they do 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 libertarian incompatibilism can provide adequate support to the principle of retributive justice. And they predict—indeed, they 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.¹⁸

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. In other words, once society internalizes the lessons of cognitive neuroscience as they bear on moral (and thus criminal) responsibility,

¹⁸ Id. at 1781.

Cognitive Neuroscience and the Future of Punishment

retribution—relying as it does on a false understanding of human agency—will be eliminated as a legitimate general or distributive justification for punishment.

Greene and Cohen consider this salutary and desirable. They assert that retributivism is largely responsible for the "counter-productive" state of the American penal system and 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 can promote beneficial effects for society by deterring future harms and incapacitating only those who would visit such harms upon the polity. Greene and Cohen's aspirational framework preserves 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.

They are not alone. 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.¹⁹ Sapolsky echoes (in a fashion) Greene and Cohen: "To understand is not to forgive or to do nothing; 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."²⁰ 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.

The long-term goal of many practitioners of neuroimaging 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 promoting human happiness. As George Fletcher observed, under their utilitarian approach, "No form of punishment could be justified unless it was the cheapest available means for serving these social ends."

The long-term aim 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 forwardlooking 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."²¹ This view led Wootton to argue for a complete abandonment of mens rea as an



¹⁹ Robert Sapolsky, *The Frontal Cortex and the Criminal Justice System*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y: BIOLOGICAL SCI., 1787, 1794 (2004).

²⁰ Id.

²¹ H.L.A. Hart, Book Review, 74 Yale L.J. 1325, 1328 (1965) (reviewing Barbara Wootton, Crime and the Criminal Law (1963)).

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. 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.

It is worth noting that while Greene and Cohen fundamentally share with Wootton the same view of the aims of criminal law, they have opposite views on whether the reasons for antisocial choices can be known. Wootton regards such reasons as unknowable, whereas Greene and Cohen are confident that someday they will become discernible through neuroimaging.

The Consequences: The End of Mitigation and Punishment as Prevention

The first consequence of the cognitive neuroscience project is the elimination of the doctrine of mitigation, traditionally understood. 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. At this stage of the criminal process, therefore, "it is impossible to offer an 'excuse' for the defendant's acts. The jury already knows that no justifiable excuse exists for what the defendant did."²² Mitigation evidence is presented in order to "inspire[] compassion . . . offer[ing] neither justification, nor excuse for the capital crime."²³ A mitigation claim is thus a plea for leniency in spite of a prior finding of legal guilt.

There are many ways in which capital defendants and the experts working on their behalf seek to move jurors towards leniency. One of the most frequent strategies is to introduce evidence that the defendant was laboring under a mental disturbance or incapacity that, while not an excuse for purposes of guilt, should nevertheless reduce his culpability.



²² Peter T. Hansen, *Mitigation: An Outline of Law, Method and Strategy*, CAP. DEF. DIG., Apr. 1992, at 29, 32.

²³ John M. Fabian, Death Penalty Mitigation and the Role of the Forensic Psychologist, 27 LAW & PSYCHOL. REV. 73, 78 (2003). (quoting Russell Stetler, *Mental Disabilities and Mitigation*, CHAMPION, Apr. 1999, at 49, 50).

This strategy, however, is squarely rooted in a distributive theory of punishment that proponents of the cognitive neuroscience project explicitly repudiate as a principal source of the irrationality and brutality that plague the current system. Paul Robinson 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."²⁴ The Supreme Court's death penalty jurisprudence confirms that the concept of mitigation grows directly out of the requirements of retributive justice:

Underlying [the Supreme Court precedents bearing on the doctrine of mitigation] is the principle that punishment should be directly related to the personal culpability of the criminal defendant. If the sentencer is to make an individualized assessment of the appropriateness of the death penalty, "evidence about the defendant's background and character is relevant because of the belief, long held by this society, that defendants who commit criminal acts that are attributable to a disadvantaged background, or to emotional and mental problems, may be less culpable than defendants who have no such excuse."²⁵

For purposes of capital mitigation, defendants increasingly invoke cognitive neuroscience evidence to advance the claim that the defendant "is a human being with redeeming value, and that he or she suffered . . . neurological . . . damage . . . that make[s] him or her less than 100 percent morally culpable for his or her behavior."²⁶ 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. This claim indulges precisely the same principle of punishment that the cognitive neuroscience project rules out of bounds.

The cognitive neuroscience project 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. As John Monahan has observed, "Assessing the likelihood of future crime . . . is a central task of sentencing under the forward-looking principle of crime control."²⁷ By contrast, such considerations

²⁶ Michael N. Burt, Forensics as Mitigation,

²⁴ Paul H. Robinson, The A.L.I.'s Proposed Distributive Principle of "Limiting Retributivism": Does It Mean in Practice Anything Other than Pure Desert?, 7 BUFF. CRIM. L. REV. 3, 5 (2003).

 ²⁵ Penry v. Lynaugh, 492 U.S. 302, 319 (1989) (quoting California v. Brown, 479 U.S. 538, 545 (1987) (O'Connor, J., concurring)), *overruled in part by* Atkins v. Virginia, 536 U.S. 304 (2002).

http://www.goextranet.net/Seminars/Dallas/BurtForensics.htm (last visited July 28, 2007). ²⁷ John Monahan, A Jurisprudence of Risk Assessment: Forecasting Harm Among Pris- oners, Predators, and Patients, 92 VA. L. REV. 391, 396 (2006).

are "jurisprudentially irrelevant to sentencing under the backward-looking principle of punishment [according to] just deserts."²⁸ Nothing in capital sentencing embodies the purely consequentialist spirit of the long-term cognitive neuroscience project as much as the aggravating factor of future dangerousness. To fully appreciate the impact that the long-term plan would like to have on capital sentencing, it is thus necessary to explore briefly the nature and contours of this element of death penalty jurisprudence.

"Future dangerousness" is a commonly invoked aggravating factor in capital sentencing. 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 one capital defense expert puts it, this is the stage of the trial where the government "suggest[s] to the jury that the defendant is a living hazard to civilization and a menacing threat to society."²⁹ To this end, prosecutors often submit the testimony of experts or laypersons regarding a defendant's future dangerousness or simply argue it themselves based on a variety of evidence. In those jurisdictions that prohibit the state from submitting an expert prediction of violence, prosecutors often try to establish future dangerousness through cross-examination of the defense's mitigation experts. Sometimes, prosecutors will try to turn the defendant's own mitigation claims against him, arguing that evidence of a violent disposition borne of abuse or a personality disorder is, in fact, demonstrative of future dangerousness.

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. Thus, actuarial and clinical evidence of future dangerousness has been admitted in jurisdictions following both the *Daubert* and *Frye* standards, and clinicians have been permitted to testify even where they have not examined the defendant. David Faigman, an expert on the use and admissibility of scientific evidence in court, has observed that "[m]ost courts either entirely ignore evidentiary standards for expert testimony concerning future violence, or give it scant attention."³⁰

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*). More importantly, courts do not regard either psychopathy or APD as an excusing condition for guilt or competence; neither is



²⁸ Id.

²⁹ Hansen, *supra* note 22, at 32.

³⁰ DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: SOCIAL AND BEHAVIORAL SCIENCE ISSUES 79-80 (2002) (including detailed discussion of why courts give little weight to evidentiary standards).

thought to sufficiently diminish the defendant's cognitive or volitional capacity for those purposes. Thus, either diagnosis both can have a devastating effect on the defendant's mitigation claims and can create an expectation in jurors' minds "that no rehabilitation is possible and that future criminal violence is inevitable."³¹

The 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 (often without ever having examined the defendant) testified to the effect that "the defendant 'has a severe antisocial personality disorder and is extremely dangerous and will commit future acts of violence.'"³² In the seminal case of *Barefoot v. Estelle*, he testified with "reasonable psychiatric certainty" that Barefoot fell in the "most severe category" of sociopaths and that Barefoot would, with "*one hundred percent and absolute*" certainty, commit future criminal acts, constituting a continuing threat to society. Prosecutors often reprise these arguments directly or raise them in response to mitigation claims of nondangerousness, in those few jurisdictions that prohibit the prediction of dangerousness in the prosecution's aggravation case-in-chief.

Studies have shown that capital juries often regard evidence of future dangerousness as the most important aggravating factor in their sentencing calculus. Indeed, two commentators have noted that "Future dangerousness takes precedence in jury deliberations over any mitigating evidence, such as remorse, mental illness, intelligence, or drug/alcohol addiction, and any concern about the defendant's behavior in prison."³³ 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.

As Paul Robinson has observed, within the context of sentencing, desert and dangerousness inevitably conflict as distributive criteria: "To advance one, the system must sacrifice the other. 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

³¹ Mark D. Cunningham & Thomas J. Reidy, *Antisocial Personality Disorder and Psychopathy: Diagnostic Dilemmas in Classifying Patterns of Antisocial Behavior in Sentencing Evaluations*, 16 BEHAV. SCI. & L. 333, 333 (1998); *see also* ohn H. Blume & David P. Voisin, *Capital Cases*, 24 CHAMPION 69, 69 (2000) (calling APD diagnosis "the kiss of death").

³² Blume & Voisin, *supra* note 31, at 69 (quoting Satterwhite v. Texas, 486 U.S. 249, 253 (1988)). For a provocative profile of Grigson, see Ron Rosenbaum, *Travels with Dr. Death*, VANITY FAIR, May 1990, at 140, *as reprinted in* NINA RIVKIND & STEVEN F. SCHATZ, CASES AND MATERIALS ON THE DEATH PENALTY 513 (2d ed. 2005).

³³ Erica Beecher-Monas & Edgar Garcia-Rill, Danger at the Edge of Chaos: Pre- dicting Violent Behavior in a Post-Daubert World, 24 CARDOZO L. REV. 1845, 1897 (2003); see also Brief for the American Psychological Ass'n & the Missouri Psychological Ass'n as Amici Curiae Supporting Respondent at 20, Roper v. Simmons, 543 U.S. 551 (2006) (No. 03-633) at 18–19.

past—allocating punishment for past offenses."³⁴ The thrust-and-parry of this conflict is played out in dramatic fashion in the capital context. 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, and best, hope.

The cognitive neuroscience project 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 project 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 the 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 nonexcusing 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. This conclusion is bolstered by capital defense experts who have observed that "Evidence of neurological impairment can be devastatingly damaging to the case for life. In presenting such evidence to a jury, counsel must be careful to avoid creating the impression that the defendant is 'damaged goods' and beyond repair."³⁵ In the regime contemplated by the cognitive neuroscience project—where claims of diminished culpability are untenable—this is the only permissible inference that jurors can draw. Arguing for compassion or leniency in such a system would be as nonsensical as 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 cars; they are only intelligible as applied to sinners.

The grave implications of the cognitive neuroscience project for capital sentencing come into even sharper relief when one considers the role that

 ³⁴ Paul H. Robinson, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114
HARV. L. REV. 1429, 1441 (2001).
³⁵ Part Lemma 24

³⁵ Burt, *supra* note 26.

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 to a "narrowing jurisprudence" of just deserts, which limits the ultimate punishment to "a narrow category of the most serious crimes" and defendants "whose extreme culpability makes them 'the most deserving of execution.'"³⁶ 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. General deterrence may be a contested issue. However, specific deterrence is always advanced by the execution of a defendant, since execution guarantees that the same defendant will not cause future harm.

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. 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 (e.g., "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, and some are worrisome examples of its possible implications.

Paul Robinson has offered a provocative genealogy for such laws that provides further grounds for caution. He makes a powerful argument that abandoning retributive justice in favor of consequentialist values of rehabilitation laid the groundwork for the draconian measures described above. According to Robinson's account, once "the limited ability of social and medical science to rehabilitate offenders became clear," reformers tried to salvage what was left of the consequentialist project by turning to incapacitation as the principle means of avoiding future crimes. He concludes that "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.*"³⁷ Robinson also points to the possibility that "if incapacitation of the

³⁶ Roper v. Simmons, 543 U.S. 551, 568 (2005).

³⁷ Robinson, *Limiting Retributivism, supra* note 24 at 14.

Cognitive Neuroscience and the Future of Punishment

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."³⁸

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 noncapital 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. 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.

Indeed, as Stephen Morse has observed, the law already has mechanisms for the preventive detention of a restricted class of individuals based on predictions about their disposition to engage in criminal misconduct. At present, these mechanisms for preventive detention of this sort apply only to narrow circumstances, namely, where the individual in question is not morally responsible for his dangerous because he is laboring under a cognitive impairment (like a mental disorder); or where he has committed a criminal act and deserves punishment. the imposition of which also involves incapacitation aimed at preventing future harms during the time of incarceration. Morse terms these constraints, respectively, "disease" and "desert." In the former case, pre-emptive involuntary civil commitment is deemed a fitting response to non-culpable dangerousness. In the latter case, judgments about future harms are integrated into the larger calculus of punishment.

The cognitive neuroscience project threatens to enlarge the "disease" justification above to encompass *all potential criminals*. That is, the cognitive neuroscience project adopts the premise that *no one* is morally responsible for his or her actions, removing the threshold safeguard for involuntary civil commitment. This seemingly opens the door to treating all individuals for whom there is good predictive evidence of a criminal disposition as we currently treat dangerous mentally ill patients. Indefinite incapacitation without stigma appears to be the logical terminus of the cognitive neuroscience project.

We have already seen the seeds of such an approach in the sexual predator context. In the landmark cases of *Kansas v. Crane* and *Kansas v. Hendricks*, the Supreme Court upheld the use of indefinite involuntary civil commitment for convicted sexual predators who had completed their prison terms, provided that he or she "suffers from a mental abnormality [defined as a "congenital or acquired

³⁸ Robinson, *supra* note 34 at 1439-40.

condition affecting the emotional or volitional capacity which predisposes the person to commit sexually violent offenses in a degree constituting such person a menace to the health and safety of others"] or personality disorder which makes the person *likely to engage in* the predatory acts of sexual violence."

It is in some ways not surprising that the Court permitted this intervention, given the fear and loathing that sexual predators inspire. But there is no reason to think that this might be the only class of individuals subject to such deprivations of liberty in the name of diminished responsibility and a disposition to violence. One obvious category of especially grave perceived threats would be individuals deemed likely to engage in terrorism. Even the most committed advocates of civil liberties and human rights, such as David Cole and Doug Cassel have begun to entertain the possibility of a narrowly circumscribed regime of administrative detention for those suspected of terrorism. The moral anthropology of the cognitive neuroscience project, if it is accepted, makes these argument all the more appealing.

Conclusion

If adopted, the cognitive neuroscience project will yield very inhumane consequences for criminal defendants — defeating the chief aspirations of its architects. Thus, for those committed to improving the lot of criminal defendants, especially those facing the death penalty, should be very wary of embracing the project. Is it possible or desirable to salvage the cognitive neuroscience project in a way that will preserve its humanitarian ends? Or is the reductive mechanist account of human personhood and human agency posited by cognitive neuroscience — and, indeed, by modern science more generally — incommensurable with the account on which the criminal law is premised? Perhaps most fundamentally, is the account of human behavior that undergirds the cognitive neuroscience project indeed an empirically demonstrated scientific conclusion, or rather simply the repackaging and extension of an undemonstrable axiom or metaphysical postulate of modern science? Understanding where an argument leads in principle and practice is a necessary precursor to appraising its wisdom— an appraisal that is particularly essential when human lives hang in the balance.

O. Carter Snead is an associate professor of law at the University of Notre Dame. He joined the university in 2005. His principal area of research is Public Bioethics – the governance of science, medicine, and biotechnology in the name of ethical goods. His scholarly works have explored issues relating to neuroethics, enhancement, stem cell research, abortion, and end-of-life decisionmaking. His articles appear in such publications as the New York University Law Review, the Harvard Law Review Forum, the Vanderbilt Law Review, Constitutional Commentary, Quaderni Costituzionali, the Yale Journal of Health Policy, Law and Ethics, the Journal of Medicine and Philosophy, and Political Science Quarterly. Professor Snead teaches Law & Bioethics, Torts, and Constitutional Criminal Procedure.

In addition to his scholarship and teaching, Professor Snead has provided advice on the legal and public policy dimensions of bioethical questions to officials in all three branches of the U.S. government, and in several intergovernmental fora. Prior to joining the law faculty at Notre Dame, Professor Snead served as General Counsel to The President's Council on Bioethics (Chaired by Dr. Leon R. Kass), where he was the primary drafter of the 2004 report, "Reproduction and Responsibility: The Regulation of New Biotechnologies." In 2006, he testified in the U.S. House of Representatives on regulatory questions concerning RU-486 (the abortion pill). From 2004 to 2005, Professor Snead led the U.S. government delegation and served as its chief negotiator for the Universal Declaration on Bioethics and Human Rights (adopted in October 2005). He served (along with Dr. Edmund Pellegrino) as U.S. government's Permanent Observer to the Council of Europe's Steering Committee on Bioethics (CDBI), where he assisted in its efforts to elaborate international instruments and standards for the ethical governance of science and medicine. In conjunction with the American Association for the Advancement of Science (AAAS), he regularly lectures to state and federal judges on the uses of neuroimaging in the courtroom. He was recently appointed by the Director-General of UNESCO to a four-year term on the International Bioethics Committee (IBC), a 36-member body of independent experts that advises member states on bioethics, law, and public policy. The IBC is the only bioethics commission in the world with a global mandate.

Professor Snead received his J.D., magna cum laude, from Georgetown University (where he was elected to the Order of the Coif), and his B.A. from St. John's College (Annapolis, MD). He clerked for the Hon. Paul J. Kelly, Jr. of the U.S. Court of Appeals for the Tenth Circuit.



Governance Studies

The Brookings Institution 1775 Massachusetts Ave., NW Washington, DC 20036 Tel: 202.797.6090 Fax: 202.797.6144 www.brookings.edu/governance.aspx

Editor Jeffrey Rosen

Benjamin Wittes

Production & Layout John S Seo

E-mail your comments to gscomments@brookings.edu

This paper is distributed in the expectation that it may elicit useful comments and is subject to subsequent revision. The views expressed in this piece are those of the author and should not be attributed to the staff, officers or trustees of the Brookings Institution.