LAW AND NEUROSCIENCE

SPEAKER SERIES 2018

FORDHAM UNIVERSITY
THE SCHOOL OF LAW
Heather A. Berlin, Ph.D., MPH
Assistant Professor, Department of Psychiatry
Icahn School of Medicine at Mount Sinai
Weill Cornell Medicine/NY Presbyterian Hospital
Dept. of Neurological Surgery
Sackler Brain and Spine Institute

THE NEUROSCIENCE OF FREE WILL

Reading Materials

1. Biographical information on Heather A. Berlin, Ph.D., MPH
2. Behavior and Responsibility: Views from Law and Neuroscience, in LAW AND NEUROSCIENCE
Dr. Heather Berlin is a cognitive neuroscientist and Assistant Professor of Psychiatry at the Icahn School of Medicine at Mount Sinai where she also completed her NIH post-doctoral fellowship. She practices clinical neuropsychology at Weill Cornell Medicine/New York Presbyterian Hospital in the Department of Neurological Surgery, and is a Visiting Scholar at the New York Psychoanalytic Society and Institute. She explores the neural basis of impulsive and compulsive psychiatric and neurological disorders with the aim of developing novel treatments. She is also interested in the brain basis of consciousness, dynamic unconscious processes, and creativity.

Dr. Berlin was a Visiting Professor at Vassar College, the Swiss Federal Institute of Technology/University of Zurich, and The Hebrew University of Jerusalem. She is the recipient of numerous honors including the Young Investigator Award from the American Neuropsychiatric Association, the International Neuropsychological Society Phillip M. Rennick Award, the Young Investigator Award from the National Education Alliance for Borderline Personality Disorder, and the Clifford Yorke Prize from the International Neuropsychoanalysis Society. She was honored as one of Stony Brook University’s “40 Under Forty”, and won the 2015 BBC University Challenge as part of the Magdalen College, Oxford team. She has published over 40 peer-reviewed articles and chapters including in high-impact journals like *American Journal of Psychiatry* and *Brain*.

Passionate about science communication and promoting women in STEM, Berlin is a committee member of the National Academy of Sciences’ *Science and Entertainment Exchange*, and The New York Times series *TimesTalks*. She hosts *Startalk All-Stars* with Neil DeGrasse Tyson, and hosted the PBS series *Science Goes to the Movies*, and the Discovery Channel series *Superhuman Showdown*. Dr. Berlin also co-wrote and stars in the critically acclaimed off-Broadway and Edinburgh Fringe Festival show, *Off the Top*, about the neuroscience of improvisation. She has made numerous media appearances including on the BBC, History Channel, Netflix, National Geographic, and TEDx, and was featured in the documentary film *Bill Nye: Science Guy*. She received her Ph.D. from the University of Oxford and Master of Public Health from Harvard University.
ASpen Casebook Series

LAW AND NEUROSCIENCE

Owen D. Jones
New York Alumni Chancellor's Professor of Law &
Professor of Biological Sciences
Vanderbilt University
Director, MacArthur Foundation Research
Network on Law and Neuroscience

Jeffrey D. Schall
E. Bronson Ingram Professor of Neuroscience
Vanderbilt University

Francis X. Shen
McKnight Land-Grant Professor & Associate Professor
University of Minnesota Law School
CHAPTER 5

Behavior and Responsibility: Views from Law and Neuroscience

Genetics may yet threaten privacy, kill autonomy, make society homogeneous, and gut the concept of human nature. But neuroscience could do all of these things first.

—The Economist

I'm not a fatalist. But even if I were, what could I do about it?

—Emo Phillips

CHAPTER SUMMARY

This chapter:

- Surveys invocations of cognitive neuroscience in philosophical and legal debates about determinism, free will, and the relationship between mind and brain.
- Highlights the extent to which doctrines of legal responsibility rely on "folk psychology" notions about the causes of human action.
- Explores views—from neuroscientists and from legal thinkers—on whether the implications of cognitive neuroscience on notions of human agency should prompt reconsideration and reform of current approaches to legal responsibility.

INTRODUCTION

Chapter 4 discussed the relationships between law, science, and behavior. In this chapter, we begin to focus on how and where law and neuroscience overlap—and, equally important, where they don't. To begin opening up this complex interaction, consider the man described in this next reading.

‡ Emo Phillips: Biography, Dead-Frog.com.
Patricia Churchland*

The Big Questions: Do We Have Free Will?
New Scientist, Nov. 18-Nov. 24, 2006, at 42

In 2003, the Archives of Neurology carried a startling clinical report.** A middle-aged Virginian man with no history of any misdemeanor began to stash child pornography and sexually molest his 8-year-old stepdaughter. Placed in the court system, his sexual behavior became increasingly compulsive. Eventually, after repeatedly complaining of headaches and vertigo, he was sent for a brain scan. It showed a large but benign tumor in the frontal area of his brain, invading the septum and hypothalamus-regions known to regulate sexual behavior.

After removal of the tumor, his sexual interests returned to normal. Months later, his sexual focus on young girls rekindled, and a new scan revealed that bits of tissue missed in the surgery had grown into a sizable tumor. Surgery once again restored his behavioral profile to “normal.”

This case renders concrete the issue of free will. Did the man have free will? Was he responsible for his behavior? Can a tumor usurp one’s free will?

---

Dr. Churchland’s passage raises many important questions. For example:

1. How did you (and should you) interpret such evidence?
2. What do you think is the probability that the tumor caused the behavior? Why, specifically?
3. What do you mean by “caused,” precisely, in this context?
4. If you think the tumor probably did play some causal role in the behavior, how do you think the legal system should respond?
5. Is this man different in kind, or just in degree, from every other offender?
6. Do you see the potential causal role for the tumor as relevant during guilt determination? Sentencing? Both? Neither? Why?
7. What about potential civil liability? Should a victim’s recovery be affected by the presence of the perpetrator’s tumor?

These questions are not easily answered. And, for a start, all require some consideration of law’s present starting point. So in Section A we consider law’s view of the person. In Section B we examine the implications of neuroscientific “determinism” for free will and legal responsibility. In Section C we explore whether reconceptualizing humans as bundles of neurons should prompt reevaluation of various legal notions, such as retributivism. Throughout this chapter, we use the criminal law to illustrate legal responsibility. But as later chapters in this book will show, debates about brains, behavior, and responsibility are important to many non-criminal legal contexts as well.

---

* [Patricia Churchland is a professor of philosophy at the University of California, San Diego. —Ens.]

** [Right Orbitofrontal Tumor with Pejoraria Sypmptoms and Constructional Apraxia Sign, 60 Archives Neurology 437, 438-40 (Jeffrey M. Burns & Russell H. Sverdlow eds., 2003). —Ens.]
A. LAW'S VIEW OF THE PERSON

Stephen Morse

Neuroscience and the Future of Personhood and Responsibility
Constitution 3.0: Freedom and Technological Change
(Jeffrey Rosen & Benjamin Wittes eds., 2011)

... The law's concern with justifying and protecting liberty and autonomy is deeply rooted in the conception of rational personhood. Human beings are part of the physical universe and subject to the laws of that universe, but, as far as we know, we are the only creatures on earth capable of acting fully for reasons and self-consciously. Only human beings are genuinely reason-responsive and live in societies that are in part governed by behavior-guiding norms. Only human beings have projects that are essential to living a good life. Only human beings have expectations of each other and require justification for interference in each other's lives that will prevent the pursuit of projects and seeking the good. We are the only creatures to whom the questions Why did you do that? and How should we behave are properly addressed, and only human beings hurt and kill each other in response to the answers to such questions. As a consequence of this view of ourselves, human beings typically have developed rich sets of interpersonal, social attitudes, practices, and institutions, including those that deal with the risk we present to each other. Among these are the practice of holding others morally and legally responsible, which depends on our attitudes and expectations about deserved praise and blame, and our practices and institutions that express those attitudes, such as reward and punishment.

There is little evidence at present that neuroscience, especially functional imaging, and genetic evidence are being introduced routinely in criminal cases outside of capital sentencing proceedings. It may well happen in the near future, however, especially as the technology becomes more broadly available and less expensive. So it's worth considering in detail neuroscience's radical challenge to responsibility, which treats people as "victims of neuronal circumstances" or the like. If this view of personhood is correct, it would indeed undermine all ordinary conceptions of responsibility and even the coherence of law itself.

Current Criminal Justice: Persons, Reasons and Responsibility
Criminal law presupposes a "folk psychological" view of the person and behavior. This psychological theory explains behavior in part by mental states such as desires, beliefs, intentions, willings, and plans. Biological, other psychological and sociological variables also play a causal role, but folk psychology considers mental states fundamental to a full causal explanation and understanding of human action. Lawyers, philosophers and scientists argue about the definitions of mental states and theories of action, but that does not undermine the general claim that mental states are fundamental. Indeed, the arguments and evidence disputants use to convince others presuppose the folk psychological view of the person. Brains don't convince each other; people do. Folk psychology presupposes only that human action will at least be rationalizable by mental-state explanations or that it will be responsive to reasons, including incentives, under the right conditions. For example, the folk psychological explanation for why you are reading this
chapter is, roughly, that you desire to understand the relation of neuroscience to criminal responsibility, you believe that reading the chapter will help fulfill that desire, and thus you formed the intention to read it.

Brief reflection should indicate that the law’s psychology must be a folk psychological theory, a view of the person as a conscious (and potentially self-conscious) creature who forms and acts on intentions that are the product of the person’s other mental states. We are the sort of creatures that can act for and respond to reasons. The law treats persons generally as intentional creatures and not simply as mechanistic forces of nature.

Law is primarily action-guiding and could not guide people directly and indirectly unless people could use rules as premises in their reasoning about how they should behave. Otherwise, law as an action-guiding system of rules would be useless, and perhaps incoherent. Legal rules are action-guiding primarily because they provide an agent with good moral or prudential reasons for forbearance or action. Human behavior can be modified by means other than influencing deliberation and human beings do not always deliberate before they act. Nonetheless, the law presupposes folk psychology, even when we most habitually follow the legal rules.

The legal view of the person does not hold that people must always reason or consistently behave rationally according to some pre-ordained, normative notion of rationality. Rather the law’s view is that people are capable of acting for reasons and are capable of minimal rationality according to predominantly conventional, socially constructed standards. The type of rationality the law requires is the ordinary person’s commonsense view of rationality, not the technical notion that might be acceptable within the disciplines of economics, philosophy, psychology, computer science, and the like.

Virtually everything for which agents deserve to be praised, blamed, rewarded, or punished is the product of mental causation and, in principle, responsive to reason, including incentives. Machines may cause harm, but they cannot do wrong and they cannot violate expectations about how people ought to live together. Machines do not deserve praise, blame, reward, punishment, concern or respect because they exist or because of the results they cause. Only people, intentional agents with the potential to act, can violate expectations of what they owe each other and only people can do wrong.

Many scientists and some philosophers of mind and action consider folk psychology to be a primitive or pre-scientific view of human behavior. For the foreseeable future, however, the law will be based on the folk psychological model of the person and behavior described. Until and unless scientific discoveries convince us that our view of ourselves is radically wrong, the basic explanatory apparatus of folk-psychology will remain central. It is vital that we not lose sight of this model lest we fall into confusion when various claims based on neuroscience or genetics are made. If any science is to have appropriate influence on current law and legal decision-making, it must be relevant to and translated into the law’s folk psychological framework.

All of the law’s doctrinal criteria for criminal responsibility are folk psychological. Begin with the definitional criteria, the “elements” of crime. The “voluntary” act requirement is defined, roughly, as an intentional bodily movement (or omission in cases in which the person has a duty to act) done in a
reasonably integrated state of consciousness. Other than crimes of strict liability, all crimes also require a culpable further mental state, such as purpose, knowledge or recklessness. All affirmative defenses of justification and excuse involve an inquiry into the person's mental state, such as the belief that self-defensive force was necessary or the lack of knowledge of right from wrong.

Our concepts of criminal responsibility follow logically from the nature of law itself and its folk psychological concept of the person and action. The general capacity for rationality is the primary condition for responsibility and the lack of that capacity is the primary condition for excusing a person. If human beings were not rational creatures who could understand the good reasons for action and were not capable of conforming to legal requirements through intentional action or forbearance, the law could not adequately guide action. Legally responsible agents are therefore people who have the general capacity to grasp and be guided by good reason in particular legal contexts.

In cases of excuse, the agent who has done something wrong acts for a reason, but is either not capable of rationality generally or is incapable on the specific occasion in question. This explains, for example, why young children and some people with mental disorders are not held responsible. How much lack of capacity is necessary to find the agent not responsible is a moral, social, political, and ultimately legal issue. It is not a scientific, medical, psychological, or psychiatric issue.

Compulsion or coercion is also an excusing condition. Literal compulsion exists when the person's bodily movement is a pure mechanism that is not rationalizable by the agent's desires, beliefs and intentions. These cases defeat the requirement of a "voluntary" act. For example, a tremor or spasm produced by a neurological disorder that causes harm is not an action because it is not intentional and it therefore defeats the ascription of a voluntary act. Metaphorical compulsion exists when the agent acts intentionally, but in response to some hard choice imposed on the agent through no fault of his or her own. For example, if a miscreant holds a gun to an agent's head and threatens to kill her unless she kills another innocent person, it would be wrong to kill under these circumstances, . . . [but] the law may decide as a normative matter to excuse the act because the agent was motivated by a threat so great that it would be supremely difficult for most citizens to resist. Cases involving internal compulsive states are more difficult to conceptualize because it is difficult to define "loss of control." The cases that most fit this category are "disorders of desire," such as addictions and sexual disorders. The question is why these acting agents lack control but other people with strong desires do not? In any case, if the person frequently yields to his or her apparently very strong desires at great social, occupational, or legal cost to himself, the agent will often say that she could not help herself, that she was not in control, and that an excuse or mitigation was therefore warranted.

The criminal law's criteria for responsibility and excuse rest on acts and mental states. In contrast, the criteria of neuroscience are mechanistic: neural structure and function. Conceptually, the apparent chasm between those two types of discourse should be bridgeable, albeit with difficulty. The brain enables the mind. If your brain is dead, you are dead, you have no mind, and you do not behave at all. Therefore, facts we learn about brains in general or about a specific brain in principle could provide useful information about mental states and human
capacities, both in general and in specific cases. While some people doubt this premise, for present purposes, let us assume that what we learn about the brain and nervous system can be potentially helpful in resolving questions of criminal responsibility.

The question is when the new neuroscience is legally relevant because it makes some given proposition about criminal responsibility more or less likely to be true. Any legal criterion must be established independently, and biological evidence must be translated into the criminal law’s folk psychological criteria. That is, the expert must be able to explain precisely how the neuroevidence bears on whether the agent acted, formed a required mens rea, or met the criteria for an excusing condition. If the evidence is not directly relevant, the expert should be able to explain the chain of inference from the indirect evidence to the law’s criteria.

At present, we lack the neuroscientific sophistication necessary to be genuinely legally relevant. The neuroscience of cognition and interpersonal behavior is largely in its infancy and what we know now is quite coarse grained and correlational, rather than causal. We lack the ability neurally to identify the content of a person’s legally relevant mental states, such as whether the defendant acted intentionally or knowingly, but we are increasingly learning about the relationship between brain structure and function and behavioral capacities, such as executive functioning. And these are relevant to broader judgments about responsibility. Over time, these problems may ease, as imaging and other techniques become less expensive and more accurate, and as the sophistication of the science increases.

Dangerous Distractions

It is important quickly to dispose of two dangerous distractions that neuroscience is thought to pose to ascriptions of criminal responsibility. The first is the threat of determinism. Many people think that neuroscience will prove once and for all that determinism (or something like it) is true and that we therefore lack free will and cannot be responsible. In this respect, however, neuroscience provides no new challenge to criminal responsibility. It cannot prove that determinism is true and it is simply the determinism du jour, grabbing the attention previously given to psychological or genetic determinism. This challenge is not a problem for criminal law because free will plays no doctrinal role in criminal law and it is not genuinely foundational for criminal responsibility. Nor is determinism inconsistent with the folk psychological view of the person. Moreover, there is a traditional, respectable philosophical reconciliation of responsibility and the truth of determinism called “compatibilism.”

Related confusions are the view that causes are per se excusing, whether they are biological, psychological or sociological, or that causation is the equivalent of compulsion. If causation were per se an excusing condition or the equivalent of compulsion, then no one or everyone would be responsible because we live in a causal universe, which includes human action. Various causes can produce genuine excusing condition, such as lack of rational or control capacity, but then it is the excusing condition, not causation, that is doing the legal work.

In contrast, the new neuroscientific challenge to personhood, exemplified by treating [a person] as a victim of neuronal circumstances, is not saved by
compatibilism or by the recognition that causation as an excuse cannot explain our practices, which hold most people responsible but excuse some. The radical challenge brain science poses threatens to undermine the very notions of agency that are presupposed by compatibilism and that are genuinely foundational for responsibility and for the coherence of law itself.

The Disappearing Person

At present, the law’s official position—that conscious, intentional, rational, and uncompelled agents may properly be held responsible—is justified. But what if neuroscience or some other discipline demonstrates convincingly that humans are not the type of creatures we think we are? Asking a creature or a mechanistic force that does not act to answer to charges does not make sense. If humans are not intentional creatures who act for reasons and whose mental states play a causal role in our behavior, then the foundational facts for responsibility ascriptions are mistaken. If it is true that we are all automatons, then no one is an agent and no one can be responsible. If the concept of mental causation that underlies folk psychology and current conceptions of responsibility is false, our responsibility practices, and many others, would appear unjustifiable.

This claim is not a strawperson, as neuroscientists Joshua Greene and Jonathan Cohen illustrate . . . [in an excerpt below].

Greene and Cohen are not alone among thoughtful people in making such claims. The seriousness of science’s potential challenge to the traditional foundations of law and morality is best summed up in the title of an eminent psychologist’s recent book, *The Illusion of Conscious Will*. If Greene and Cohen are right, cases that involve alleged abnormalities are really indistinguishable from any other case and thus represent just the tip of the iceberg that will sink our current criminal justice system. In this view, we are all “merely victims of neuronal circumstances.”

But are we? Is the rich explanatory apparatus of intentionality simply a post hoc rationalization the brains of hapless homo sapiens construct to explain what their brains have already done? Will the criminal justice system as we know it wither away as an outmoded relic of a prescientific and cruel age? If so, not only criminal law is in peril. What will be the fate of contracts, for example, when a biological machine that was formerly called a person claims that it should not be bound because it did not make a contract? The contract is also simply the outcome of various neuronal circumstances.

This picture of human activity exerts a strong pull on the popular, educated imagination too. In an ingenious recent study, investigators were able to predict accurately based on which part of the brain was physiologically active whether a shopper-subject would or would not make a purchase. This study was reported in the *Science* Times section of the *New York Times*. The story’s spin began with its title: “Findings: The Voices in My Head Say ‘Buy It!’ Why Argue?” It reflects once again the mechanistic view of human activity. What people do is simply a product of brain regions and neurotransmitters. The person disappears. There is no shopper. There is only a brain in a mall.

The law’s fundamental presuppositions about personhood and action are indeed open to profound objection. Action and consciousness are scientific and conceptual mysteries. We do not know how the brain enables the mind, and we do
not know how action is possible. At most we have hypotheses or a priori arguments. Moreover, causation by mental states seems to depend on now largely discredited mind-brain dualism that treats minds and brains as separate entities that are somehow in communication with one another. How can such tenuously understood concepts be justifiable premises for legal practices such as blaming and punishing? And if our picture of ourselves is wrong, as many neuroscientists claim, then our responsibility practices are morally unjustified according to any moral theory we currently embrace.

Given how little we know about the brain-mind and brain-action connections, to claim based on neuroscience that we should radically change our picture of ourselves and our practices is a form of neuroarrogance. Although I predict that we will see far more numerous attempts to introduce neuroevidence in the future, the dystopia that Greene and Cohen predict is not likely to come to pass. There is little reason at present to believe that we are not agents.

Most scientists and philosophers of science are physicalists and monists; they believe, as I do, that all material and non-material elements begin with matter subject to the universe's physical laws and that we do not have minds or souls independent of our bodies. But theorists such as Greene and Cohen go a step further. They appear to assume the validity of a complete reduction of mental states to brain states at the level of (apparently) neural networks. Indeed, the complete post–Enlightenment project of reducing all phenomena to the most basic physical building blocks is controversial even among physicalist monists and most probably is a chimera. Almost certainly, a complete explanation of human behavior will have to use multiple fields and multiple levels within each field. The complete reductionists have to explain how molecules, which have no intentionality or temporal sense, produce intentional creatures with a sense of past, present and future that guides our lives.

It is also possible that if we do ever discover how the brain enables the mind, this discovery will so profoundly alter our understanding of ourselves as biological creatures that all moral and political notions will change. Nevertheless, this argument is different from claiming that we are not agents, that our mental states do no explanatory work, . . .

NOTES AND QUESTIONS

1. Do you find Morse's description and analysis of the criminal law implications of neuroscience accurate and persuasive? Why or why not?
2. Would you extend these arguments beyond criminal responsibility, to apply to responsibility in tort as well?
3. What predictions does this view make about the way law evolves over time?
4. What do you see as the likely result when neuroscience meets folk psychology?
5. Although Morse critiques neuroarrogance, common ground is found in the belief that "we do not have minds or souls independent of our bodies." This belief is a departure from traditional religious views, still shared by a majority of Americans. Does embrace of neurolaw, even in a more modest form as Morse advocates for, necessarily require rejection of the soul?
B. DETERMINISM

Morse considers determinism to be the first of two “dangerous distractions.” Consider the following view:

“Free will has long been a fraught concept among philosophers and theologians. Now neuroscience is entering the fray. For centuries, the idea that we are the authors of our own actions, beliefs, and desires has remained central to our sense of self. We choose whom to love, what thoughts to think, which impulses to resist. Or do we? … Neuroscience suggests something else…. What’s at stake? Just about everything: morality, law, religion, our understanding of accountability and personal accomplishment, even what it means to be human.”

The Chronicle Review, Mar. 23, 2012, at B6. As you will see in the excerpts below, some believe that neuroscientific findings put traditional concepts of free will, moral responsibility, and legal responsibility in jeopardy. As you read the views excerpted below, it will be useful to keep in mind three distinct reactions to this challenge.

One reaction, known as incompatibilism, holds that that because the universe is deterministic it allows no room at all for free will. Adherents to its opposite reaction, known as compatibilism, reject the determinism/free-will dichotomy as false, believing that meaningful quanta of free will exist, even within a universe that operates according to materialistic and deterministic principles. A third reaction, known as metaphysical libertarianism, rejects that the universe is determined, thereby affording expansive room for free will.

Jerry A. Coyne
You Don’t Have Free Will

The term “free will” has so many diverse connotations that I’m obliged to define it before I explain why we don’t have it. I construe free will the way I think most people do: At the moment when you have to decide among alternatives, you have free will if you could have chosen otherwise. To put it more technically, if you could rerun the tape of your life up to the moment you make a choice, with every aspect of the universe configured identically, free will means that your choice could have been different.

Although we can’t really rerun that tape, this sort of free will is ruled out, simply and decisively, by the laws of physics. Your brain and body, the vehicles that make “choices,” are composed of molecules, and the arrangement of those molecules is entirely determined by your genes and your environment. Your decisions result from molecular-based electrical impulses and chemical substances transmitted from one brain cell to another. These molecules must obey the laws of physics, so the outputs of our brain — our “choices” — are dictated by those laws. … And deliberating about your choices in advance doesn’t help matters, or that deliberation also reflects brain activity that must obey physical laws.

* [Jerry A. Coyne is a professor in the department of ecology and evolution at the University of Chicago. — Eds.]
To assert that we can freely choose among alternatives is to claim, then, that we can somehow step outside the physical structure of our brain and change its workings. That is impossible. Like the output of a programmed computer, only one choice is ever physically possible: the one you made. As such, the burden of proof rests on those who argue that we can make alternative choices, for that's a claim that our brains, unique among all forms of matter, are exempt from the laws of physics by a spooky, nonphysical "will" that can redirect our own molecules.

My claim that free will as defined above is an illusion leads to a prediction: Our sense of controlling our actions might sometimes be decoupled from those actions themselves. Recent experiments in cognitive science show that some deliberate acts occur before they reach our consciousness (typing or driving, for example), while in other cases, brain scans can predict our choices several seconds before we're conscious of having made them. Additionally, stimulation of the brain, or clever psychological experiments, can significantly increase or decrease our sense of control over our choices.

So what are the consequences of realizing that physical determinism negates our ability to choose freely? Well, nihilism is not an option: We humans are so constituted, through evolution or otherwise, to believe that we can choose. What is seriously affected is our idea of moral responsibility, which should be discarded along with the idea of free will. If whether we act well or badly is predetermined rather than a real choice, then there is no moral responsibility—only actions that hurt or help others. That realization shouldn't seriously change the way we punish or reward people, because we still need to protect society from criminals, and observing punishment or reward can alter the brains of others, acting as a deterrent or stimulus. What we should discard is the idea of punishment as retribution, which rests on the false notion that people can choose to do wrong.

Although science strongly suggests that free will of the sort I defined doesn't exist, this view is unpopular because it contradicts our powerful feeling that we make real choices. In response, some philosophers—most of them determinists who agree with me that our decisions are preordained—have redefined free will in ways that allow us to have it. I see most of these definitions as face-saving devices designed to prop up our feeling of autonomy. To eliminate the confusion produced by multiple and contradictory concepts of free will, I propose that we reject the term entirely and adopt the suggestion of the cognitive scientist Marvin Minsky: Instead of saying my decision arises from free will, we might say, "My decision was determined by internal forces I do not understand."

Hilary Bok*

Want to Understand Free Will? Don't Look to Neuroscience

As a philosopher, I often find speculation about the implications of neuroscience for free will perplexing. While some neuroscientists describe free will in ways that I recognize, others, including some distinguished and thoughtful scientists, do not.

* [Hilary Bok is an associate professor of philosophy at the Johns Hopkins University. —EBS]
Thus Benjamin Libet: If “our consciously willed acts are fully determined by natural laws that govern the activities of nerve cells in the brain,” then free will is “illusory.”

Most philosophers disagree.

Among philosophers the main division is between compatibilists, who believe that free will is compatible with causal determinism, and incompatibilists, who believe that it is not. Almost all compatibilists think that we are free. Most are not determinists, but they believe that we would be free even if our actions are fully determined.

With the exception of those who work within a religious tradition, philosophers tend to be naturalists who see individual mental events as identical with events in our brains. When we say that a person’s choice caused her action, we do not mean that she swooped in from outside nature and altered her destiny; we mean that an event in her brain caused her to act. On this view, the claim that a person chose her action does not conflict with the claim that some neural processes or states caused it; it simply redescribes it.

For compatibilists, therefore, the problem of free will is not that neuroscience reveals our choices as superfluous. It does not. Nor do compatibilists deny that our choices cause us to do things. The problem of free will for compatibilists is not to preserve a role for deliberation and choice in the face of explanations that threaten them with elimination; it is to explain how, once our minds and our choices have been thoroughly naturalized, we can provide an adequate account of human agency and freedom.

How can we reconcile the idea that our choices have scientific explanations with the idea that we are free? Determinism does not relieve us of the need to make decisions. And when we make decisions, we need some conception of the alternatives available to us. If we define an alternative as an action that is physically possible, then determinism implies that we never have more than one alternative. But since we cannot know in advance what we will choose, if we define “alternative” this way, we will never know what our alternatives are. For the purposes of deciding what to do, we need to define our alternatives more broadly: as those actions that we would perform if we chose them.

A person whose actions depend on her choices has alternatives; if she is, in addition, capable of stepping back from her existing motivations and habits and making a reasoned decision among them, then, according to compatibilists, she is free.

Whether this view provides an adequate account of free will is not a problem neuroscience can solve. Neuroscience can explain what happens in our brains: how we perceive and think, how we weigh conflicting considerations and make choices, and so forth. But the question of whether freedom and moral responsibility are compatible with free will is not a scientific one, and we should not expect scientists to answer it.

Whatever their views on the compatibility of freedom and determinism, most philosophers agree that someone can be free only if she can make a reasoned choice among various alternatives, and act on her decision; in short, only if she has the capacity for self-government.

Neuroscience can help us to understand what this capacity is and how it can be strengthened. What, for instance, determines when we engage in conscious
self-regulation, and how might we ensure that we do so when we need to? If the exercise of self-government can deplete our capacity for further self-government in the short run, what exactly is depleted, and how might we compensate for its loss? Does self-government deplete our resources in the short run while strengthening them over time, like physical exercise, or does it simply weaken our ability to govern ourselves without any compensating benefit?

Neuroscience can answer those questions, and it can provide causal explanations of human action, but it can’t resolve the question of whether or not such explanations are compatible with free will.

Owen D. Jones

The End of (Discussing) Free Will

The problem with free will is that we keep dwelling on it. Really, this has to stop. Free will is to human behavior what a perfect vacuum is to terrestrial physics—a largely abstract endpoint from which to begin thinking, before immediately moving on to consider and confront the practical frictions of daily existence.

I do get it. People don’t like to be caused. It conflicts with their preference to be fully self-actualized. So it is understandable that, at base, free-will discussions tend to center on whether people have the ability to make choices uncaused by anything other than themselves. But there’s a clear answer: They don’t. Will is as free as lunch. (If you doubt, just try willing yourself out of love, lust, anger, or jealousy.)

All animals are choice machines for two simple reasons. First, no organism can behave in all physically possible ways simultaneously. Second, alternative courses are not all equal. At any given moment, there are far more ways to behave disastrously than successfully (just as there are more ways to break a machine than to fix it). So persistence of existence consistently depends on one’s ability to choose nondisastrous courses of action.

Yet (indeed, fortunately) that choosing is channeled. Choices are initially constrained by the obvious—the time one has to decide, and the volume of brain tissue one can deploy to the task. Choices are also constrained by things we have long suspected but which science now increasingly clarifies.

For example, human brains are not general-purpose processors, idly awaiting culture’s activating infusion of consciousness. Evolutionary processes pre-equip brains in all species with some information-processing predispositions. Generally speaking, these increase the probabilities that some combinations of environmental circumstances—immediate physical and social factors, contexts, and the like—will yield one subset of possible (and generally nondisastrous) behaviors rather than others.

Also, we now know that brains, though remarkable and often malleable, are functionally specialized. That is, different brain regions have evolved to do different things—even though they generally do more than one thing. As a
consequence, impairments to specific areas of the brain—through injury or disease, for example—can impede normal human decision-making. And those impediments can, in turn, relax inhibitions, increase impulsive and addictive behaviors, alter the ability to make moral judgments, or otherwise leave a person situated dissimilarly from the rest of the population.

Which brings us to law. How will insights from the brain sciences affect the ways we assess a person’s responsibility for bad behavior? Answer: only somewhat, but sometimes significantly. Many people assume that legal responsibility requires free will, such that an absence of free will necessarily implies an absence of responsibility. Not true, as many scholars have amply demonstrated. Full, complete, utterly unconstrained freedom to choose among available actions might be nice to have, but it is not in fact necessary for a fair and functioning legal system.

This is not to say that *degrees* of freedom are irrelevant to law. Science hasn’t killed free will. But it has clarified various factors—social, economic, cultural, and biological in nature—that constrain it.

The existence of constraints very rarely excuses behavior, as when a person in an epileptic fit hits someone. But evidence of brain-based constraints—which can vary from small to large—can be, and indeed have been, relevant in determining the severity of punishment. For example, some jurors in a recent Florida case reported that evidence of abnormal brain functioning warranted a murderer spending his life in prison, instead of being executed.

All behaviors have causes, and all choices are constrained. We need to accept this and adapt. Brain sciences are revealing complex and interconnected pathways by which the information-processing activities of multiple brain regions coalesce to influence human decision-making. But this poses an advantage—neither a threat nor a revolutionary transition—to the legal system. In the near term, these complexities are more likely to inform than to utterly transform law’s justice-driven efforts to treat people fairly and effectively.

**Michael S. Gazzaniga**

*Free Will Is an Illusion, but You’re Still Responsible for Your Actions*


Neuroscience reveals that the concept of free will is without meaning... neuroscience, with its ever-increasing mechanistic understanding of how the brain enables mind, suggests that there is no one thing in us pulling the levers and in charge... But brain determinism has no relevance to the concept of personal responsibility.

The exquisite machine that generates our mental life also lives in a social world and develops rules for living within a social network. For the social network to function, each person assigns each other person responsibility for his or her actions. There are rules for traffic that exist and are only understood and adopted when cars interact. It is the same for human interactions. Just as we would not try to understand traffic by studying the mechanics of cars, we should not try to

---

* [Michael S. Gazzaniga is director of the SAGE Center for the Study of the Mind at the University of California at Santa Barbara. — Eds.]
understand brains to understand the idea of responsibility. Responsibility exists at a different level of organization: the social level, not in our determined brains.

Paul Bloom*

*Free Will Does Not Exist. So What?*

...Most of all, the deterministic nature of the universe is fully compatible with the existence of conscious deliberation and rational thought. These (physical and determined) processes can influence our actions and our thoughts, in the same way that the (physical and determined) workings of a computer can influence its output. It is wrong, then, to think that one can escape from the world of physical causation—but it is not wrong to think that one can think, that we can mull over arguments, weigh the options, and sometimes come to a conclusion. After all, what are you doing now?

Patricia Churchland**

**The Big Questions: Do We Have Free Will?**
*New Scientist, Nov. 18-Nov. 24, 2006, at 42*

...I suggest that free will, as traditionally understood, needs modification. Because of its importance in society, any description of free will updated to fit what we know about the nervous system must also reflect our social need for a working concept of responsibility.

Think about what we mean by "free will." As with all concepts, we learn the meaning of this from examples. We learn what to count as fair, or mean-spirited, or voluntary by being given sterling examples of people doing things that are fair, or mean-spirited or voluntary.... Our understanding is balanced by contrasting cases-actions that are obviously not freely chosen: a dreaming man who strangles his wife, the toddler who wets his pants, a startle response to a thunderclap, or a coerced confession. From such prototypes, brains manage to extract a common enough meaning so that we can talk about free will tolerably well.

As well as prototypical cases, there are outlying cases beset with ambiguity, daunting complexity and background cultural differences. Here, the status of an action—freely chosen or not—has no clear answer, and such cases often come before the courts. Andrea Yates, the Texas mother who drowned her five children in a bathtub, was unquestionably psychotic, though her actions were methodical and purposeful, unlike the erratic movements of someone suffering an epileptic seizure. She understood that her actions were against the law, and telephoned the police to say so. Outside of our usual ken, this sort of case divides opinion. The way we currently think about free will, there may be no right answer as to whether she exercised it.

---

* [Paul Bloom is a professor of psychology and cognitive science at Yale University.—Eds.]

** [Patricia Churchland is a professor of philosophy at the University of California, San Diego.—Eds.]
A rigid philosophical tradition claims that no choice is free unless it is uncaused: that is, unless the "will" is exercised independently of all causal influences—in a causal vacuum. In some unexplained fashion, the will—a thing that allegedly stands aloof from brain-based causality—makes an unconstrained choice. The problem is that choices are made by brains, and brains operate causally; that is, they go from one state to the next as a function of antecedent conditions. Moreover, though brains make decisions, there is no discrete brain structure or neural network which qualifies as "the will" let alone a neural structure operating in a causal vacuum. The unavoidable conclusion is that a philosophy dedicated to uncaused choice is as unrealistic as a philosophy dedicated to a flat Earth.

To begin to update our ideas of free will, I suggest we first shift the debate away from the puzzling metaphysics of causal vacuums to the neurobiology of self-control. The nature of self-control and the ways it can be compromised may be a more fruitful avenue to understand cases such as the Virginian man* and Andrea Yates than trying to force the issue of "freely chosen or not."

Self-control can come in many degrees, shades, and styles. We have little direct control over autonomic functions such as blood pressure, heart rate and digestion, but vastly more control over behavior that is organized by the cortex of the brain. Self-control is mediated by pathways in the prefrontal cortex, shaped by structures regulating emotions and drives. . . .

Ulysses famously bound himself to the mast of his ship to avoid seduction by the sirens, and monkeys will deviate from a direct route to avoid a temptation known to be troublesome. This is the prefrontal cortex using cognition for impulse control.

Self-control also allows us to make sense of difficult cases where free will is unhelpful. Self-control may be diminished in persons with brain lesions or tumors. Self-control is also diminished during an epileptic seizure, while intoxicated or under anesthesia. . . .

How do neural networks achieve these effects that we call self-control, and what is different in the brain when self-control functions are impaired? Although little is known so far about the exact nature of the mechanisms, relevant experimental details have begun to pour in from many directions: on the properties of neurons sensitive to reward and punishment, on the generation of fear responses by neurons in the amygdala, and on the response profiles of "decision" neurons in parietal regions of cortex when the animal makes a choice. . . .

These sorts of discoveries promise that eventually we will understand, at least in general terms, the neurobiological profile of a brain that has normal levels of control, and how it differs from a brain that has compromised control.

So is anyone ever responsible for anything? Civil life requires it be so. . . .

NOTES AND QUESTIONS

1. Where and how do the views above overlap? Differ?
2. With which of the perspectives do you most closely identify, and with what implications for whether you are an incompatibilist, compatibilist, or metaphysical libertarian?

* [The reference is to the Virginian pedophile in the chapter-opening excerpt.—Eds.]
3. Consider this passage: "I found myself... looking down on the face of the dead man. I hadn’t consciously thought about doing that— but when it became apparent what my body was up to, I didn’t veto the action either." Robert J. Sawyer, Mindscan 138 (2005). To what extent does this suggest an important distinction between "free will" and "free won’t"? For more on this subject, see Michael S. Gazzaniga, Who’s in Charge?: Free Will and the Science of the Brain (2011).

4. Consider the two examples mentioned thus far of individuals who committed serious crimes against children: the Virginia man who molested his step-daughter and Andrea Yates, the woman who methodically drowned her five children. If you were forced to absolve one, whom would you absolve? Why? Can you reconcile that choice, in a principled way, with neuroscientific perspectives? Does that matter?

5. Stephen J. Morse has argued: "There is no bright line between free and unfree choices. Harder and easier choices are arranged along a continuum of choice: there is no scientifically dictated cutting point where legal and moral responsibility begins or ends." The Twilight of Welfare Criminology, 49 S. Cal. L. Rev. 1247, 1253 (1976). Do you agree or disagree? Why or why not? Must the law define cut points?

C. THE DISAPPEARING PERSON

A second "dangerous distraction"—in Morse’s view—is the case of the "disappearing person," wherein the very notion of personhood is winnowed by the view, incorrect in Morse’s opinion, that to identify a cause is to identify an excuse. Consider this passage (to which the Morse excerpt above refers). As you read, consider: Do you find Greene/Cohen, Sapolsky, or Morse more persuasive on these issues? Or do you disagree with all of them?

Joshua Greene & Jonathan Cohen

For the Law, Neuroscience Changes Nothing and Everything

1. INTRODUCTION

The law takes a long-standing interest in the mind. In most criminal cases, a successful conviction requires the prosecution to establish not only that the defendant engaged in proscribed behavior, but also that the misdeed in question was the product of mens rea, a "guilty mind." Narrowly interpreted, mens rea refers to the intention to commit a criminal act, but the term has a looser interpretation by which it refers to all mental states consistent with moral and/or legal blame. . . . Thus, for centuries, many legal issues have turned on the question: "what was he thinking?" . . .

Given the law’s . . . concern for mental states, along with its preference for "hard" evidence, it is no surprise that interest in the potential legal implications of cognitive neuroscience abounds. But does our emerging understanding of the mind as brain really have any deep implications for the law? . . . Some have argued . . . that new neuroscience contributes nothing more than new details
and that existing legal principles can handle anything that neuroscience will throw our way in the foreseeable future.

... Existing legal principles make virtually no assumptions about the neural bases of criminal behavior, and as a result they can comfortably assimilate new neuroscience without much in the way of conceptual upheaval. ... We maintain, however, that our operative legal principles exist because they more or less adequately capture an intuitive sense of justice. In our view, neuroscience will challenge and ultimately reshape our intuitive sense(s) of justice. New neuroscience will affect the way we view the law, not by furnishing us with new ideas or arguments about the nature of human action, but by breathing new life into old ones. Cognitive neuroscience, by identifying the specific mechanisms responsible for behavior, will vividly illustrate what until now could only be appreciated through esoteric theorizing: that there is something fishy about our ordinary conceptions of human action and responsibility, and that, as a result, the legal principles we have devised to reflect these conceptions may be flawed.

Our argument runs as follows: first, we draw a familiar distinction between the consequentialist justification for state punishment, according to which punishment is merely an instrument for promoting future social welfare, and the retributivist justification for punishment, according to which the principal aim of punishment is to give people what they deserve based on their past actions. We observe that the common-sense approach to moral and legal responsibility has consequentialist elements, but is largely retributivist. Unlike the consequentialist justification for punishment, the retributivist justification relies, either explicitly or implicitly, on a demanding ... conception of free will. We therefore consider the standard responses to the philosophical problem of free will. "Liberarians" (no relation to the political philosophy) and "hard determinists" agree on "incompatibilism," the thesis that free will and determinism are incompatible, but they disagree about whether determinism is true, or near enough true to preclude free will. Libertarians believe that we have free will because determinism is false, and hard determinists believe that we lack free will because determinism is (approximately) true. "Compatibilists," in contrast to libertarians and hard determinists, argue that free will and determinism are perfectly compatible.

We argue that current legal doctrine, although officially compatibilist, is ultimately grounded in intuitions that are incompatibilist and, more specifically, libertarian. In other words, the law says that it presupposes nothing more than a metaphysically modest notion of free will that is perfectly compatible with determinism. However, we argue that the law's intuitive support is ultimately grounded in a metaphysically overambitious, libertarian notion of free will that is threatened by determinism and, more pointedly, by forthcoming cognitive neuroscience. At present, the gap between what the law officially cares about and what people really care about is only revealed occasionally when vivid scientific information about the causes of criminal behavior leads people to doubt certain individuals' capacity for moral and legal responsibility, despite the fact that this information is irrelevant according to the law's stated principles. We argue that new neuroscience will continue to highlight and widen this gap. That is, new neuroscience will undermine people's common sense, libertarian conception of free will and the retributivist thinking that depends on it, both of which have heretofore been shielded by the inaccessibility of sophisticated thinking about the mind and its neural basis.
The net effect of this influx of scientific information will be a rejection of free will as it is ordinarily conceived, with important ramifications for the law. As noted above, our criminal justice system is largely retributivist. We argue that retributivism, despite its unstable marriage to compatibilist philosophy in the letter of the law, ultimately depends on an intuitive, libertarian notion of free will that is undermined by science. Therefore, with the rejection of common-sense conceptions of free will comes the rejection of retributivism and an ensuing shift towards a consequentialist approach to punishment, i.e. one aimed at promoting future welfare rather than meting out just deserts. Because consequentialist approaches to punishment remain viable in the absence of common-sense free will, we need not give up on moral and legal responsibility. . . . We can, however, recognize that free will, as conceptualized by the folk psychology system, is an illusion and structure our society accordingly by rejecting retributivist legal principles that derive their intuitive force from this illusion.

2. TWO THEORIES OF PUNISHMENT: CONSEQUENTIALISM AND RETRIBUTIVISM

There are two standard justifications for legal punishment. According to the forward-looking, consequentialist, . . . punishment is justified by its future beneficial effects. Chief among them are the prevention of future crime through the deterrent effect of the law and the containment of dangerous individuals. Few would deny that the deterrence of future crime and the protection of the public are legitimate justifications for punishment. The controversy surrounding consequentialist theories concerns their serviceability as complete normative theories of punishment. Most theorists find them inadequate in this regard, and many argue that consequentialism fundamentally mischaracterizes the primary justification for punishment, which, these critics argue, is retribution. As a result, they claim, consequentialist theories justify intuitively unfair forms of punishment, if not in practice then in principle. One problem is that of Draconian penalties. It is possible, for example, that imposing the death penalty for parking violations would maximize aggregate welfare by reducing parking violations to near zero. But, retributivists claim, whether or not this is a good idea does not depend on the balance of costs and benefits. It is simply wrong to kill someone for double parking. A related problem is that of punishing the innocent. It is possible that, under certain circumstances, falsely convicting an innocent person would have a salutary deterrent effect, enough to justify that person’s suffering, etc. Critics also note that, so far as deterrence is concerned, it is the threat of punishment that is justified and not the punishment itself. Thus, consequentialism might justify letting murderers and rapists off the hook so long as their punishment could be convincingly faked. . . .

The backward-looking, retributivist account does a better job of capturing these intuitions. Its fundamental principle is simple: in the absence of mitigating circumstances, people who engage in criminal behavior deserve to be punished, and that is why we punish them. Some would explicate this theory in terms of criminals’ forfeiting rights, others in terms of the rights of the victimized, whereas others would appeal to the violation of a hypothetical social contract, and so on. Retributivist theories come in many flavors, but these distinctions need not
concern us here. What is important for our purposes is that retributivism captures the intuitive idea that we legitimately punish to give people what they deserve based on their past actions—in proportion to their “internal wickedness,” to use Kant’s phrase—and not, primarily, to promote social welfare in the future.

The retributivist perspective is widespread, both in the explicit views of legal theorists and implicitly in common sense. There are two primary motivations for questioning retributivist theory. The first, which will not concern us here, comes from a prior commitment to a broader consequentialist moral theory. The second comes from skepticism regarding the notion of desert, grounded in a broader skepticism about the possibility of free will in a deterministic or mechanistic world.

3. FREE WILL AND RETRIBUTIVISM

The problem of free will is old and has many formulations. Here is one, drawing on a more detailed and exacting formulation by Peter Van Inwagen: determinism is true if the world is such that its current state is completely determined by (i) the laws of physics and (ii) past states of the world. Intuitively, the idea is that a deterministic universe starts however it starts and then ticks along like clockwork from there. Given a set of prior conditions in the universe and a set of physical laws that completely govern the way the universe evolves, there is only one way that things can actually proceed.

... There are three standard responses to the problem of free will. The first, known as “hard determinism,” accepts the incompatibility of free will and determinism (“incompatibilism”), and asserts determinism, thus rejecting free will. The second response is libertarianism (again, no relation to the political philosophy), which accepts incompatibilism, but denies that determinism is true. This may seem like a promising approach. After all, has not modern physics shown us that the universe is indeterministic? The problem here is that the sort of indeterminism afforded by modern physics is not the sort the libertarian needs or desires. If it turns out that your ordering soup is completely determined by the laws of physics, the state of the universe 10,000 years ago, and the outcomes of myriad subatomic coin flips, your appetizer is no more freely chosen than before. Indeed, it is randomly chosen, which is no help to the libertarian. What about some other kind of indeterminism? What if, somewhere deep in the brain, there are mysterious events that operate independently of the ordinary laws of physics and that are somehow tied to the will of the brain’s owner? In light of the available evidence, this is highly unlikely. Say what you will about the “hard problem” of consciousness, there is not a shred of scientific evidence to support the existence of causally effective processes in the mind or brain that violate the laws of physics. In our opinion, any scientifically respectable discussion of free will requires the rejection of what Strawson famously called the “panicky metaphysics” of libertarianism.

Finally, we come to the dominant view among philosophers and legal theorists: compatibilism. Compatibilists concede that some notions of free will may require indefensible, panicky metaphysics, but maintain that the kinds of free will “worth wanting,” to use Dennett’s phrase, are perfectly compatible with determinism. Compatibilist theories vary, but all compatibilists agree that free will is a perfectly natural, scientifically respectable phenomenon and part of the ordinary human condition. They also agree that free will can be undermined by various kinds of psychological deficit, e.g. mental illness or “infancy.” Thus, according to this view,
a freely willed action is one that is made using the right sort of psychology—rational, free of delusion, etc.

Compatibilists make some compelling arguments. After all, is it not obvious that we have free will? Could science plausibly deny the obvious fact that I am free to raise my hand at will? For many people, such simple observations make the reality of free will non-negotiable. But at the same time, many such people concede that determinism, or something like it, is a live possibility. And if free will is obviously real, but determinism is debatable, then the reality of free will must not hinge on the rejection of determinism. That is, free will and determinism must be compatible. Many compatibilists skeptically ask what would it mean to give up on free will. Were we to give it up, wouldn’t we have to immediately reinvent it? Does not every decision involve an implicit commitment to the idea of free will? And how else would we distinguish between ordinary rational adults and other individuals, such as young children and the mentally ill, whose will—or whatever you want to call it—is clearly compromised? Free will, compatibilists argue, is here to stay, and the challenge for science is to figure out how exactly it works and not to peddle silly arguments that deny the undeniable.

The forward-looking-consequentialist approach to punishment works with all three responses to the problem of free will, including hard determinism. This is because consequentialists are not concerned with whether anyone is really innocent or guilty in some ultimate sense that might depend on people’s having free will, but only with the likely effects of punishment. (Of course, one might wonder what it means for a hard determinist to justify any sort of choice. We will return to this issue in §8.) The retributivist approach, by contrast, is plausibly regarded as requiring free will and the rejection of hard determinism. Retributivists want to know whether the defendant truly deserves to be punished. Assuming one can deserve to be punished only for actions that are freely willed, hard determinism implies that no one really deserves to be punished. Thus, hard determinism combined with retributivism requires the elimination of all punishment, which does not seem reasonable. This leaves retributivists with two options: compatibilism and libertarianism. Libertarianism, for reasons given above, and despite its intuitive appeal, is scientifically suspect. At the very least, the law should not depend on it. It seems, then, that retributivism requires compatibilism. Accordingly, the standard legal account of punishment is compatibilist.

[The authors discuss, among other things, the views of those who claim that neuroscience will change nothing of significance in the law.]

6. NEUROSCIENCE AND THE TRANSPARENT BOTTLENECK

We have argued that, contrary to legal and philosophical orthodoxy, determinism really does threaten free will and responsibility as we intuitively understand them. It is just that most of us, including most philosophers and legal theorists, have yet to appreciate it. This controversial opinion amounts to an empirical prediction that may or may not hold: as more and more scientific facts come in, providing increasingly vivid illustrations of what the human mind is really like, more and more people will develop moral intuitions that are at odds with our current social practices.

Neuroscience has a special role to play in this process for the following reason. As long as the mind remains a black box, there will always be a donkey on which to
pin dualist and libertarian intuitions. For a long time, philosophical arguments have persuaded some people that human action has purely mechanical causes, but not everyone cares for philosophical arguments. 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.

Thus, neuroscience holds the promise of turning the black box of the mind into a transparent bottleneck. There are many causes that impinge on behavior, but all of them—from the genes you inherited, to the pain in your lower back, to the advice your grandmother gave you when you were six—must exert their influence through the brain. Thus, your brain serves as a bottleneck for all the forces spread throughout the universe of your past that affect who you are and what you do. Moreover, this bottleneck contains the events that are, intuitively, most critical for moral and legal responsibility, and we may soon be able to observe them closely.

At some time in the future we may have extremely high-resolution scanners that can simultaneously track the neural activity and connectivity of every neuron in a human brain, along with computers and software that can analyze and organize these data. Imagine, for example, watching a film of your brain choosing between soup and salad. The analysis software highlights the neurons pushing for soup in red and the neurons pushing for salad in blue. You zoom in and slow down the film, allowing yourself to trace the cause-and-effect relationships between individual neurons—the mind’s clockwork revealed in arbitrary detail. You find the tipping-point moment at which the blue neurons in your prefrontal cortex out-fire the red neurons, seizing control of your pre-motor cortex and causing you to say, “I will have the salad, please.”

At some further point this sort of brainware may be very widespread, with a high-resolution brain scanner in every classroom. People may grow used to the idea that every decision is a thoroughly mechanical process, the outcome of which is completely determined by the results of prior mechanical processes. What will such people think as they sit in their jury boxes? Suppose a man has killed his wife in a jealous rage. Will jurors of the future wonder whether the defendant acted in that moment of his own free will? Will they wonder if it was really him who killed his wife rather than his uncontrollable anger? Will they ask whether he could have done otherwise? Whether he really deserves to be punished, or if he is just a victim of unfortunate circumstances? We submit that these questions, which seem so important today, will lose their grip in an age when the mechanical nature of human decision-making is fully appreciated. The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless.
At least in our more reflective moments. Our intuitive sense of free will runs quite deep, and it is possible that we will never be able to fully talk ourselves out of it.

8. FREE WILL, RESPONSIBILITY AND CONSEQUENTIALISM

Even if there is no intuitively satisfying solution to the problem of free will, it does not follow that there is no correct view of the matter. Ours is as follows: when it comes to the issue of free will itself, hard determinism is mostly correct. Free will, as we ordinarily understand it, is an illusion. However, it does not follow from the fact that free will is an illusion that there is no legitimate place for responsibility. Recall from §2 that there are two general justifications for holding people legally responsible for their actions. The retributive justification, by which the goal of punishment is to give people what they really deserve, does depend on this dubious notion of free will. However, the consequentialist approach does not require a belief in free will at all. As consequentialists, we can hold people responsible for crimes simply because doing so has, on balance, beneficial effects through deterrence, containment, etc. It is sometimes said that if we do not believe in free will then we cannot legitimately punish anyone and that society must dissolve into anarchy. In a less hysterical vein, Daniel Wegner argues that free will, while illusory, is a necessary fiction for the maintenance of our social structure. We disagree. There are perfectly good, forward-looking justifications for punishing criminals that do not depend on metaphysical fictions. (Wegner’s observations may apply best to the personal sphere: see below.)

The vindication of responsibility in the absence of free will means that there is more than a grain of truth in compatibilism. The consequentialist approach to responsibility generates a derivative notion of free will that we can embrace. In the name of producing better consequences, we will want to make several distinctions among various actions and agents. To begin, we will want to distinguish the various classes of people who cannot be deterred by the law from those who can. That is, we will recognize many of the ‘diminished capacity’ excuses that the law currently recognizes such as infancy and insanity. We will also recognize familiar justifications such as those associated with crimes committed under duress (e.g. threat of death). If we like, then, we can say that the actions of rational people operating free from duress, etc. are free actions, and that such people are exercising their free will.

At this point, compatibilists such as Daniel Dennett may claim victory: “what more could one want from free will?” In a word: retributivism. We have argued that commonsense retributivism really does depend on a notion of free will that is scientifically suspect. Intuitively, we want to punish those people who truly deserve it, but whenever the causes of someone’s bad behavior are made sufficiently vivid, we no longer see that person as truly deserving of punishment. This insight is expressed by the old French proverb: “to know all is to forgive all.” It is also expressed in the teachings of religious figures, such as Jesus and Buddha, who preach a message of universal compassion. Neuroscience can make this message more compelling by vividly illustrating the mechanical nature of human action.

Our penal system is highly counter-productive from a consequentialist perspective, especially in the USA, 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 on which retributivism depends.

As advocates of consequentialist legal reform, it behooves us to briefly respond to the three standard criticisms levied against consequentialist theories of punishment. First, it is claimed that consequentialism would justify extreme overpunishing. As noted above, it is possible in principle that the goal of deterrence would justify punishing parking violations with the death penalty or framing innocent people to make examples of them. Here, the standard response is adequate. The idea that such practices could, in the real world, make society happier on balance is absurd. Second, it is claimed that consequentialism justifies extreme underpunishment. In response to some versions of this objection, our response is the same as above. Deceptive practices such as a policy of faking punishment cannot survive in a free society, and a free society is required for the pursuit of most consequentialist ends. In other cases consequentialism may advocate more lenient punishments for people who, intuitively, deserve worse. Here, we maintain that a deeper understanding of human action and human nature will lead people—more of them, at any rate—to abandon these retributivist intuitions. Our response is much the same to the third and most general criticism of consequentialist punishment, which is that even when consequentialism gets the punishment policy right, it does so for the wrong reasons. These supposedly right reasons are reasons that we reject, however intuitive and natural they may feel. They are, we maintain, grounded in a metaphysical view of human action that is scientifically dubious and therefore an unfit basis for public policy in a pluralistic society.

Finally, as defenders of hard determinism and a consequentialist approach to responsibility, we should briefly address some standard concerns about the rejection of free will and conceptions of responsibility that depend on it. First, does not the fact that you can raise your hand “at will” prove that free will is real? Not in the sense that matters. As Daniel Wegner has argued, our first-person sense of ourselves as having free will may be a systematic illusion. And from a third-person perspective, we simply do not assume that anyone who exhibits voluntary control over his body is free in the relevant sense, as in the case of Mr. Puppet.

A more serious challenge is the claim that our commitments to free will and retributivism are simply inescapable for all practical purposes. Regarding free will, one might wonder whether one can do so much as make a decision without implicitly assuming that one is free to choose among one’s apparent options. Regarding responsibility and punishment, one might wonder if it is humanly possible to deny our retributive impulses. This challenge is bolstered by recent work in the behavioral sciences suggesting that an intuitive sense of fairness runs deep in our primate lineage and that an adaptive tendency towards retributive punishment may have been a crucial development in the biological and cultural evolution of human sociality. Recent neuroscientific findings have added further support to this view, suggesting that the impulse to exact punishment may be driven by phylogenetically old mechanisms in the brain. These mechanisms may be an efficient and perhaps essential, device for maintaining social stability. If retributivism runs that deep and is that useful, one might wonder whether we have any serious hope of, or reason for, getting rid of it. Have we any real choice but to see one another as free agents who deserve to be rewarded and punished for our past behaviors?
We offer the following analogy: modern physics tells us that space is curved. Nevertheless, it may be impossible for us to see the world as anything other than flatly Euclidean in our day-to-day lives. And there are, no doubt, deep evolutionary explanations for our Euclidean tendencies. Does it then follow that we are forever bound by our innate Euclidean psychology? The answer depends on the domain of life in question. In navigating the aisles of the grocery store, an intuitive, Euclidean representation of space is not only adequate, but probably inevitable. However, when we are, for example, planning the launch of a spacecraft, we can and should make use of relativistic physical principles that are less intuitive but more accurate. In other words, a Euclidean perspective is not necessary for all practical purposes, and the same may be true for our implicit commitment to free will and retributivism. For most day-to-day purposes it may be pointless or impossible to view ourselves or others in this detached sort of way. But—and this is the crucial point—it may not be pointless or impossible to adopt this perspective when one is deciding what the criminal law should be or whether a given defendant should be put to death for his crimes. These may be special situations, analogous to those routinely encountered by “rocket scientists,” in which the counter-intuitive truth that we legitimately ignore most of the time can and should be acknowledged.

Finally, there is the worry that to reject free will is to render all of life pointless: why would you bother with anything if it has all long since been determined? The answer is that you will bother because you are a human, and that is what humans do. Even if you decide, as part of a little intellectual exercise, that you are going to sit around and do nothing because you have concluded that you have no free will, you are eventually going to get up and make yourself a sandwich. And if you do not, you have got bigger problems than philosophy can fix.

Robert M. Sapolsky

The Frontal Cortex and the Criminal Justice System

1. INTRODUCTION

...[S]ome findings in neuroscience should seem nothing short of flabbergasting to any intelligent person. In some instances, these findings must challenge our sense of self...[N]eurobiology is beginning to provide the first hints of mechanistic explanations for our personalities, propensities and passions.

These insights can be of extraordinary relevance, in that neurobiology often must inform some of our decision-making. Is a loved one, sunk in a depression so severe that she cannot function, a case of a disease whose biochemical basis is as "real" as is the biochemistry of, say, diabetes, or is she merely indulging herself? Is a child doing poorly at school because he is unmotivated and slow, or because there is a neurobiologically based learning disability? Is a friend, edging towards a serious problem with substance abuse, displaying a simple lack of discipline, or suffering from problems with the neurochemistry of reward?...

Arguably, the most important arena in which a greater knowledge of neuroscience is needed is the criminal justice system. In some cases, the criminal justice
system has accommodated well the lessons of neurobiology. If someone with
epilepsy, in the course of a seizure, flails and strikes another person, that epileptic
would never be considered to have criminally assaulted the person who they struck.
But in earlier times, that is exactly what would have been concluded, and epilepsy
was often assumed to be a case of retributive demonic possession. Instead, we are
now a century or two into readily dealing with the alternative view of, “it is not him,
it is his disease.” . . .

[T]he criminal justice system in the USA has been dominated increasingly by a
view that an inability to tell right from wrong is the sole basis of an acceptable
insanity defense. I will now examine how contemporary neuroscience strongly
argues against this trend. Instead, we have come to understand increasingly the
organic basis of impaired impulse control.

[The author surveys literature concerning the relationship between impulse
control problems and damage to a region of the brain—the prefrontal cortex
(or “PFC”)—that is known to be importantly involved in (among other things)
choosing the harder but analytically “more correct” course of action.]

We have come to recognize numerous realms in which a biological abnormal-
ity gives rise to aberrant behavior. And such recognition has often then given rise
to an expectation that people now exert higher-order control over that abnormal-
ity. For example, as noted, we would never consider an epileptic violent who strikes
someone in the process of a seizure: “it is not him; it is his disease.” However, we
expect that epileptic to not drive a car if their seizures are uncontrolled. Or we are
coming to understand the neurochemistry of context-dependence relapse into
drug dependency in organisms. Thus, we have come to expect ex-addicts to
avoid the settings in which they previously abused drugs.

There is a false dichotomy in this manner of thinking. It is as if we artificially
demarcate an area in which biology dominates: yes, there is something organic that
gives rise to this person having uncontrolled and synchronous neuronal discharges
(i.e. a seizure), or who has certain pathways potentiated that project onto
dopamine-releasing ‘pleasure’ pathways (one theory about the neurochemistry
of substance abuse relapse). But it is as if, with that area of organic impairment
identified and given credence, we expect it to be bounded, and for the rest of our
“us-ness,” replete with free will, to now shoulder the responsibility of keeping that
organic impairment within the confines of its boundaries. It cannot possibly work
this way. What the literature about the PFC shows is that there is a reductive, mate-
rialist neurobiology to the containment, resulting in the potential for volitional
control to be impaired just as unambiguously as any other aspect of brain function.
It is possible to know the difference between right and wrong but, for reasons of
organic impairment, to not be able to do the right thing.

The most obvious implication of this concerns how individuals with demon-
strable PFC damage are treated in the criminal justice system. As the simplest con-
clusion, everything about this realm of contemporary neurobiology argues against
the retrenchment back towards a sole reliance on M’Naghten that has gone on in
recent decades. [For discussion of the M’Naghten standard, see Chapter 1.]

Amid the seeming obviousness of this conclusion, there is always a valid
counter-point that can be raised: there are individuals with substantial amounts
of PFC damage who, nonetheless, do not commit crimes. At present, knowing
that someone has sustained PFC damage does not give much power in predicting
whether that person's disinhibition will take the form of serial murder or merely being unable to praise a nearly inedible meal prepared by a host. This seems to weaken the "volition can be organically impaired, just like any other aspect of brain function" argument; in these interstices of unpredictability seem to dwell free will.

However, we can begin to imagine tree diagrams of variables that, with each new layer, add more predictive power. We can already see two layers in the realm of PFC function. The first layer might query, "PFC: normal or damaged?" (while recognizing that this is a false dichotomy). The second might then query, "if damaged: damaged in childhood or later?" This same structure of increasing predictive power was shown in a recent, landmark study concerning clinical depression ... [that] generates an impressive predictive power as to which adults succumb to clinical depression. If free will lurks in those interstices, those crawl spaces are certainly shrinking.

A second way in which findings about the PFC are relevant to the criminal justice system concerns individuals who have committed grotesquely violent, sociopathic crimes, but who have no demonstrable PFC damage. Initially, it seems a fatuous tautology to say that there must be an organic abnormality in such cases—"it is only an organically abnormal brain that produces abnormal behavior"—and that we simply lack sufficiently sensitive techniques for demonstrating it. However, it must be emphasized that most of the neurobiological techniques used to demonstrate PFC abnormalities in humans (predominantly structural and functional brain imaging) did not exist a decade or two ago. It would be the height of hubris to think that we have already learned how to detect the most subtle ways in which PFC damage impairs volitional control. Instead, we probably cannot even imagine yet the ways in which biology can go awry and impair the sorts of volitional control that helps define who we are.

At the most disturbing level, findings about the PFC are relevant to the criminal justice system with respect to those of us with a normal PFC and who have never behaved criminally. It is here that the tendency of science to function in continua comes up against the legal culture of jury decisions. Among sociopaths without overt PFC damage, the smaller the volume of the PFC, the greater the tendency towards aggressive and antisocial behavior. Similarly, as noted, among humans with no neurological impairments or histories of antisocial behavior, the greater the level of metabolic activity in parts of the PFC, the lower the activity of the amygdala. There is little support for the idea that over the range of PFC function, there is a discontinuity, a transition that allows one to dichotomize between a healthy PFC in an individual expected to have a complete capacity to regulate behavior, and a damaged PFC in someone who cannot regulate their behavior. The dichotomy does not exist.

A conclusion like this makes sense to neurobiologists, but may seem alien to legal scholars. The emphasis on continua seems to hold the danger of a world of criminal justice in which there is no blame and only prior causes. Whereas it is true that, at a logical extreme, a neurobiological framework may indeed eliminate blame, it does not eliminate the need for forceful intervention in the face of violence or antisocial behavior. 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.
Legal scholars have objected to this type of thinking for a related reason, as well. In this view, it is desirable for a criminal justice system to operate with a presumption of responsibility because, “to treat persons otherwise is to treat them as less than human.” There is a certain appealing purity to this. But 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.

NOTES AND QUESTIONS

1. Are the views of Greene/Cohen and Sapolsky radical, reasonable, or both? Would it be fair to characterize Greene/Cohen and Sapolsky, or either of those excerpts, as a plea for greater recognition that humans are victims of neuronal circumstances? How would you describe the main implications they draw for law from neuroscientific perspectives? Given that, do you find the views of Greene/Cohen, Sapolsky, or Morse more appealing? For what reasons?

2. How persuasive do you find Sapolsky’s argument for medicalizing criminal justice?

3. What are your reactions to Sapolsky’s and Greene/Cohen’s arguments for abandoning retributivist approaches to criminal justice in favor of consequentialist ones? Why?

4. What evidence would be required for you to reach the conclusion that a defendant did not have the requisite prefrontal resources to be responsible for his crime? How would you determine that a defendant’s moral engine was broken beyond repair? Would it affect your decisions about culpability, mitigation, or both?

5. Is hard determinism completely incompatible with retributivism? To what extent is free will necessary to justify retributive justice?

6. Greene and Cohen argue (in another passage) that “syndromes and other causes do not have excusing force unless they sufficiently diminish rationality in the context in question.” Would prefrontal cortex damage and its resulting impaired volitional control that Sapolsky describes meet that requirement of sufficiently diminished rationality?

7. What is your reaction to the following reasoning?

Imagine this futuristic courtroom scene. The defence barrister stands up, and pointing to his client in the dock, makes this plea: “The case against Mr X must be dismissed. He cannot be held responsible for smashing Mr Y’s face into a pulp. He is not guilty, it was his brain that did it. Blame not Mr X, but his overactive amygdala.” . . . [Yet] those who blame the brain should be challenged as to why they stop at the brain when they seek the causes of bad behaviour. Since the brain is a physical object, it is wired into nature at large. “My brain made me do it!” must mean (ultimately) that “The Big Bang” made me do it . . . And there is a contradiction built into the plea of neuromitigation. The claim “my brain made me do it” suggests that I am not my brain; even that my brain is some kind of alien force, one of the founding notions of neurolaw, however, is that the person is the brain. If I were my brain, then “My brain made me do it” would boil down to “I made me do it” and that would hardly get me off the hook.

8. “Sam Harris [author of a recent book titled *Free Will*] predicts that a declaration by the scientific community that free will is an illusion would ‘set off a culture war far more belligerent than the one that has been waged on the subject of evolution.’” *Is Free Will an Illusion?, The Chronicle Review, March 23, 2012, at B6*. What do you think of that prediction?

9. Philosopher Nicole Vincent argues that there is good reason to turn to neuroimaging for individualized assessment of responsibility. She recognizes ten challenges to this proposal, but argues that none of these worries can be addressed. As you consider this list, do you agree?

... [T]en worries and problems with the suggestion that neuroimaging can help us to individually assess people’s responsibility: (1) that not all people with abnormal brains commit crimes; (2) that our understanding of how the human brain works is still very rudimentary; (3) that brain plasticity might make it difficult to diagnose who has which capacities; (4) that methodological and technological problems with current neuroimaging techniques cast doubt over the usefulness of neuroimaging data; (5) that we can’t go back in time and check what capacities a person had at the time when they committed their crime; (6) that neuroimaging evidence for a person’s incapacity may actually damn them even harder rather than exculpating them; (7) that social factors, and not just neurological impairments, also play some role in determining our behaviour; (8) that responsibility assessments also depend in part on normative assumptions which are at least partially independent of what cognitive neuroscience tells us about the human mind; (9) that people might be responsible for their own incapacity and thus that they might be responsible for what they do on account of that; and (10) that people who know about their own incapacities may be responsible for what they do if they fail to avoid situations in which those incapacities may become a problem.


10. Consider this statement: “There is something more to being a person than biology can tell us.” Steven K. Erickson, *Blaming the Brain*, 11 Minn. J.L. Sci. & Tech. 27 (2010). To what extent do you agree or disagree, and why? To what extent could neuroscience be thought to ever shift crime policy “from a legal framework to one engineered by science”? What would be the hallmarks of such a shift? What would be the appropriate legal response?

11. Neuroscientist Benjamin Libet performed a landmark study showing that when a subject has been instructed to flick his wrist at will, and to note the time at which he decides to do so, neural signals leading to the movement actually precede the subject’s conscious awareness that he is about to do so (summarized in his article *Do We Have Free Will?* 6 J. Consciousness Stud. 47 (1999)). A great deal of commentary has debated whether or not these findings are meaningful, both scientifically and normatively. If neuroscience shows conclusively that brain processes precede (and cause) the experience of “will,” in what sense are human agents free? For further exploration of the controversies
over the implications of Libet’s work, see Conscious Will and Responsibility: A Tribute to Benjamin Libet (Walter Sinnott-Armstrong & Lynn Nadel eds., 2010).

12. Neuroscientist David Eagleman suggests that systems of law should be consistent with contemporary brain science. To measure this congruence, he proposes a “neurocompatibility index,” which states:

Criminal jurisprudence is often driven more by intuition and political need than by scientific understanding. A brain-compatible system prizes fairness and long-term crime prevention over harsh but inconsequential punishment. We have developed an index to measure the degree to which systems of law around the world are aligned with modern brain science. The criteria include:

understanding of mental illness
meaningful methods for rehabilitation
individualized sentencing based on risk assessment
eyewitness identification standards
specialized court systems (mental health, drug, juvenile)
incentive structuring based on behavioral economics
a minimum standard of science education for policy-makers . . .

David M. Eagleman & Sarah Isgur Flores, Defining a Neurocompatibility Index for Criminal Justice Systems: a Framework to Align Social Policy with Modern Brain Science, in Law of the Future 161-72 (Ser. No. 1, 2012). Is an index such as this a useful lens through which to view the justice system?

FURTHER READING

On Free Will and Responsibility:
Michael S. Pardo & Dennis Patterson, Philosophical Foundations of Law and Neuroscience, 2010 U. Ill. L. Rev. 1211 (2010).

On the Folk Psychology of Intentionality:

In Addition: