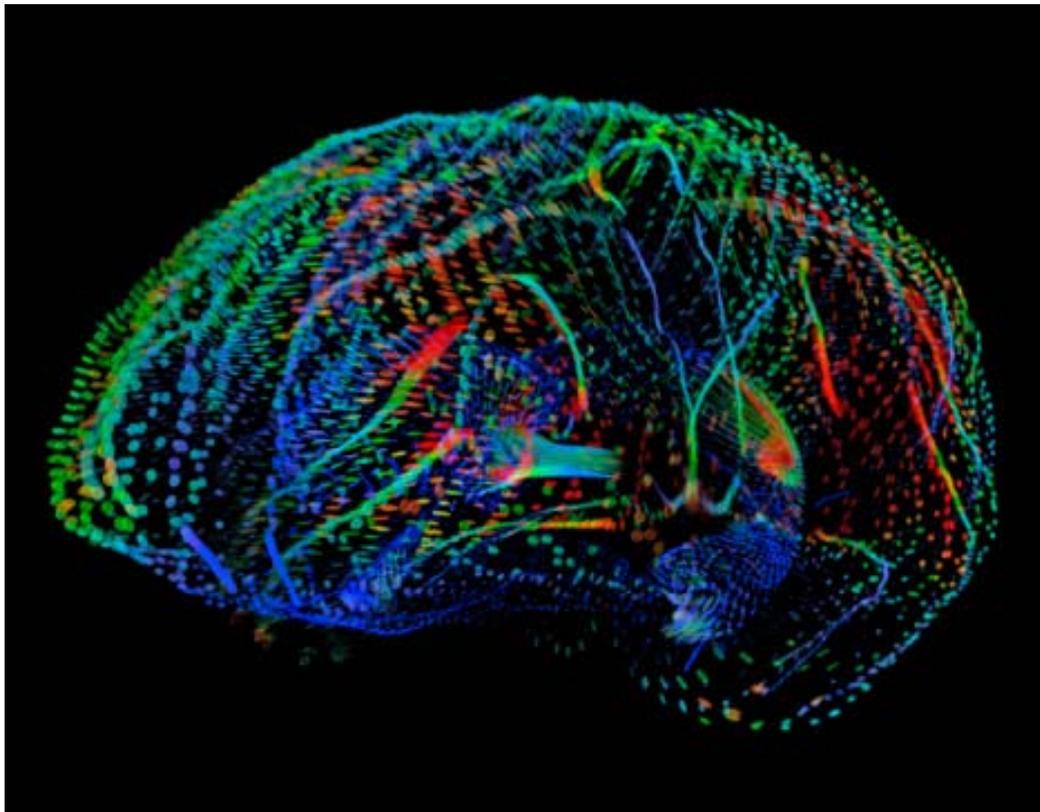




THE FUTURE OF THE CONSTITUTION

February 3, 2011



Arthur Toga/UCLA

Neuroscience and the Future of Personhood and Responsibility

Stephen J. Morse

Collera, a 28 year old man with a life-long history of aggressive behavior, including assaultive conduct and abusive verbal behavior, is driving his large SUV behind a slow moving vehicle on a narrow road with no room to pass. He honks and honks, but the driver in front neither speeds up nor pulls off the road to let Collera pass. Collera starts to curse vehemently and to pull dangerously close to the slower vehicle. Collera's passenger warns that he is taking a very serious risk. Collera finally announces in a fury that he's going to kill the [expletive deleted] in front. He allows his vehicle to drop back a bit, and then he floors the SUV's gas pedal, crashing into the slower vehicle at great speed. Neither he nor his passenger is hurt, but the driver of the slower vehicle is killed.

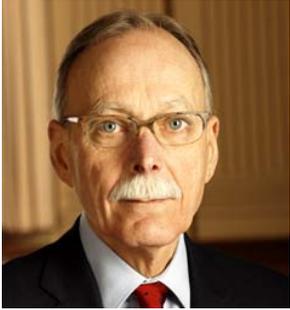
An evaluation of Collera after the killing discloses the following. A functional brain image that measures brain activation discloses that Collera has a type of neurophysiological activity in his right frontal cortex that is associated with poor behavioral self-regulation.¹ Collera's life history includes a history of severe abuse. It is known that such abuse is strongly associated with later antisocial conduct if the person also has a genetic profile that affects particular neurotransmitter levels.² Collera indeed has the genetic profile and the associated neurotransmitter levels.

How should the law respond to people like Collera? Do we treat him, as we now do, as an acting agent who is properly subject to moral assessment and potential liability to just punishment? If so, how does the evaluation bear on his responsibility and future dangerousness? It appears from the limited facts that he has no specific doctrinal defense to murder. In deciding what the just punishment might be, however, how should the information from the evaluation be used? In the alternative, suppose Collera is simply a "victim of neuronal circumstances," as some would claim. Or suppose that although we still think of him as an agent, our prediction and control technology has immeasurably advanced. What should be the proper response?

Imagine that this takes place in the future, when we will have much better information about the biologically causal variables, especially neuroscientific and genetic factors, that produce all dangerous behavior and not just seemingly extreme cases like Collera'. The description of Collera's evaluation results makes no mention of disease or disorder. It simply reports a number of neuroscientific,

¹ Tiffany W. Chow & Jeffrey L. Cummings, *Frontal-Subcortical Circuits*, in Bruce L. Miller & Jeffrey L. Cummings (eds.), *THE HUMAN FRONTAL LOBES: FUNCTIONS AND DISORDERS* (2D ED.) 25, 27-31 (2007). Damage to this region is also associated with antisocial behavior. Steven W. Anderson et al, *Impairment of social and moral behavior related to early damage in human prefrontal cortex*. 2 NAT. NEUROSCI. 1032 (1999); R. James Blair & Lisa Cipolotti, *Impaired social response reversal: a case of acquired sociopathy*, 123 BRAIN 1122 (2000); Jeffrey L. Saver & Antonio R. Damasio, *Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage*, 29 NEUROPSYCHOLOGIA 1241 (1991). Let us assume, however, that Collera is not obviously damaged.

² Avshalom Caspi et al, *Role of genotype in the cycle of violence in maltreated children*, 297 SCIENCE 851 (2002).



Stephen J. Morse is the Ferdinand Wakeman Hubbell Professor of Law and Professor of Psychology and Law in Psychiatry at the University of Pennsylvania Law School and School of Medicine.

genetic and gene-by-environment interaction variables that played an apparently causal role in producing Collera’s behavior and that might have helped us predict it. Will jurisprudence that respects agency, which enhances the dignity, liberty and autonomy of all citizens, survive in a future in which neuroscience and genetics dominate our thinking about personhood and responsibility. Will we abandon the concepts of criminal, crime, responsibility, blame, and punishment, and replace them by concepts such as “dangerous behavior” and “preventive control”? Will people in this brave new world be treated simply as biological mechanisms and will harmdoing be characterized simply as one mechanistic output of the system? As *The Economist* has warned: “Genetics may yet threaten privacy, kill autonomy, make society homogeneous, and gut the concept of human nature. But neuroscience could do all those things first.”³

The law in our liberal democracy responds to the need to restrain dangerous people like Collera by what I have termed “desert-disease” jurisprudence.⁴ As a consequence of taking people seriously as people, as potential moral agents, we believe that it is crucial to cabin the potentially broad power of the state to deprive people of liberty. With rare exceptions, the state may only restrain a citizen if that citizen has been fairly convicted of crime and deserves the punishment imposed. If a citizen has not committed a crime but appears dangerous and not responsible for his or her dangerousness—usually as a result of mental disorder or other diseases that impair rationality—the citizen may be civilly committed. People who are simply dangerous but who have committed no crime and who are responsible agents cannot be restrained. The normative basis of desert-disease jurisprudence is that it enhances liberty and autonomy by leaving people free to pursue their projects unless an agent responsibly commits a crime or unless through no fault of his own the agent is non-responsibly dangerous. In the latter case, the agent’s rationality is impaired and the usual presumption in favor of liberty and autonomy yields to the need for societal protection and preventive detention and involuntary treatment may be warranted.

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

³*The Ethics of Brain Science: Open Your Mind*, *ECONOMIST*, May 25, 2002, at 77.

⁴Stephen J. Morse, *Neither Desert Nor Disease*, 5 *LEGAL THEORY* 265, 267-70 (1999).

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. Unless people are capable of understanding and then using legal rules to guide their conduct, law would be powerless to affect human behavior.

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 common sense 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."⁵ 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 herself, 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

⁵ Stephen J. Morse, *Uncontrollable Urges and Irrational People*, 88 VIRGINIA L. REV. 1025 (2002).

principle could provide useful information about mental states and human capacities, both in general and in specific cases. While some people doubt this premise,⁶ 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.⁷ 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.⁸ Nor is determinism inconsistent with the

⁶ M.R. Bennett and P.M.S. Hacker, *Philosophical Foundations of Neuroscience* (2003); Michael Pardo and Dennis Patterson, 'Philosophical Foundations of Neuroscience' (2010) *U.Ill.L. Rev.* 1211

⁷ Gregory A. Miller, *Mistreating Psychology in the Decades of the Brain*, 5 *PERSPECTIVES PSYCHOL. SCI.* 716 (2010).

⁸ Stephen J. Morse, *The non-problem of free will in forensic psychiatry and psychology*, 25 *BEHAV. SCI.& L.* 203 (2007)

folk psychological view of the person. Moreover, there is a traditional, respectable philosophical reconciliation of responsibility and the truth of determinism called “compatibilism.”⁹

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 Collera 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:

[A]s 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 positions....What neuroscience does, and will continue to do at an

⁹ Robert Kane, *A CONTEMPORARY INTRODUCTION TO FREE WILL* (2006), pp.12-22.

accelerated pace, is elucidate the “when,” “where” and “how” of the mechanical processes that cause behavior. It is one thing to deny that human behavior 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...At some further point...[p]eople may grow up completely 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? Will jurors of the future wonder whether the defendant... *could have done otherwise*? Whether he really *deserves* to be punished...? 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.¹⁰

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*.¹¹ 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. We are, after 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

¹⁰ Joshua Greene & Jonathan Cohen, *For the law, the new neuroscience changes nothing and everything*, in Samir Zeki & Oliver Goodenough (eds.) *LAW & THE BRAIN* 207, 217-18(2006).

¹¹ Daniel Wegner, *THE ILLUSION OF CONSCIOUS WILL* (2002).

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?"¹² 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.¹³ 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.¹⁴ 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

¹² John Tierney, *Findings: The Voices in My Head Say "Buy It!" Why Argue?*, N.Y. TIMES, Jan. 16, 2007, at F1.

¹³ Paul R. McHugh & Philip R. Slavney, *THE PERSPECTIVES OF PSYCHIATRY* 11-12 (2d. Ed. 1998)

¹⁴ Carl Craver, *EXPLAINING THE BRAIN; MECHANISMS AND THE MOSSAIC UNITY OF NEUROSCIENCE* (2007); Alva Noe, *OUT OF OUR HEADS: WHY YOU ARE NOT YOUR BRAIN AND OTHER LESSONS FROM THE BIOLOGY OF CONSCIOUSNESS* (2009).

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.

The Evidence for the “Victims of Neuronal Circumstances” Thesis

The real question behind the “victims of neuronal circumstances” thesis (VNC) is whether scientific and clinical investigations have shown that agency is rare or non-existent, that conscious will is largely or entirely an illusion. Four kinds of indirect evidence are often adduced: first, demonstrations that a very large part of our activity is undeniably caused by variables we are not in the slightest aware of; second, studies indicating that more activity than we think takes place when our consciousness is divided or diminished; third, laboratory studies that show that people can be experimentally misled about their causal contribution to their apparent behavior; and, fourth, evidence that particular types of psychological processes are associated with heightened physiological activation in specific regions of the brain. None of these types of indirect evidence offers logical support to the VNC. Although the science behind the claims for VNC is often good, the VNC claim itself is a non-sequitur because these studies do not demonstrate that mental states play no causal role in most behavior we now consider intentional.¹⁵

There is also allegedly direct experimental evidence of VNC from studies done by neuroscientist Benjamin Libet and his followers, which have generated an immense amount of comment. Indeed, many claim that Libet’s work is the first direct neurophysiological evidence of VNC. Libet’s studies demonstrate that measurable electrical brain activity associated with the intentional actions of raising one’s finger or flexing one’s wrist at random occurs in the relevant motor area of the brain about 550 milliseconds before the subject actually acts and about 350-400 milliseconds before the subject is consciously aware of the intention to act. Let us assume, with cautious reservations, the basic scientific methodological validity of these studies. The crucial question is whether the interpretation of these findings as supporting VNC is valid. Do these findings mean that brain events are the entire causal explanation and that mental states played no causal role in explaining the subjects’ finger-raising and wrist-flexings? This claim has been contested by numerous people, including legal philosopher, Michael Moore, and philosopher, Alfred Mele, who have usefully shown that the Libetian conception of the role of brain events in causing behavior does not offer a coherent conceptual or empirical account of the relation between brain states and behavior.¹⁶ My own

¹⁵ Stephen J. Morse, *Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience*, 9 MINN. J. L. SCI. & TECH. 1 (2008).

¹⁶ Michael S. Moore, *Libet’s Challenges to Responsible Human Agency*, in Walter Sinnott-Armstrong and Lynn Nadel (eds), *CONSCIOUS WILL AND RESPONSIBILITY: A TRIBUTE TO BENJAMIN LIBET* (2011), ch 18; .Alfred Mele, *EFFECTIVE INTENTIONS: THE POWER OF THE CONSCIOUS WILL* (2009).

work has contributed a more empirical and common sense critique.¹⁷ Indeed, it is at present an open question whether Libet's paradigm is representative of intentional actions in general because Libet investigated such trivial behavior. At least at present, this body of work does not remotely indicate that mental states play no causal role whatsoever in our intentional actions. I doubt that future science will change this conclusion.

Answers to VNC are rooted in common sense, a plausible theory of mind, our evolutionary history, and practical necessity. Virtually every neurologically intact person consistently has the experience of first person agency, the experience that one's intentions flow from one's desires and beliefs and result in action. Indeed, this folk-psychological experience is so central to human life and so apparently explanatory that it is difficult to imagine giving it up or a good reason to do so, even if it were possible to give it up. Folk psychology has much explanatory power, and we are capable of investigating its claims scientifically. There is compelling psychological evidence that intentions play a causal role in explaining behavior. Indeed, it is hard to imagine the nature of a scientific study that would prove conclusively that mental states do *no* work to creatures that have created that study and will assess it with mental states.

The plausible theory of mind that might support mental state explanations is thoroughly physical, but non-reductive and non-dualist.¹⁸ It hypothesizes that all mental and behavioral activity is the causal product of physical events in the brain, that mental states are real, that they are caused by lower level biological processes in the brain, that they are realized in the brain—the mind-brain—but not at the level of neurons, and that mental states can be causally efficacious. It accepts that a fully causal story about behavior will be multifold and multilevel.

There is a plausible evolutionary story about why folk psychology is causally explanatory and why human beings need rules such as those provided by law. We have evolved to be self-conscious creatures that act for reasons and are responsive to reasons. Acting for reasons is inescapable for creatures like ourselves who inevitably care about the ends we pursue. Because we are social, language-using creatures whose interactions are not governed primarily by innate repertoires, it is inevitable that rules will be necessary to help order our interactions in any minimally complex social group. Our ancestors would have been much less successful, and therefore much less likely to be our ancestors, if they had been unable to understand the intentions of others and predict their behavior accordingly. Psychologists call this having a "theory of mind," and people who do not develop one adequately experience profound difficulties in their interpersonal lives.

None of these considerations is an incontrovertible analytic argument against VNC, but surely the burden of persuasion is on those who argue to the contrary.

¹⁷ Stephen J. Morse, *Lost in Translation? An Essay on Law and Neuroscience*, in Michael Freeman (ed.), *LAW AND NEUROSCIENCE* (2011), p.529.

¹⁸ E.g., John R. Searle, *MIND: A BRIEF INTRODUCTION* (2004), pp.113-114.

At the very least, we remain entitled to presume that conscious intentions are causal until the burden is met.

But let's suppose that we were convinced by the mechanistic view that we are not intentional, rational agents after all. (Of course, the notion of being "convinced" would be an illusion, too. Being convinced means that we are persuaded by evidence or argument, but a mechanism is not persuaded by anything. It is simply neurophysically transformed.) What should we do now? We know that it is an illusion to think that our deliberations and intentions have any causal efficacy in the world. We also know, however, that we experience sensations such as pleasure and pain and that we care about what happens to us and to the world. We cannot just sit quietly and wait for our brains to activate, for determinism to happen. We must, and will of course, deliberate and act.

If we still thought that VNC were correct and that standard notions of genuine moral responsibility and desert were therefore impossible, we might nevertheless continue to believe that the law would not necessarily have to give up the concept of incentives. Indeed, Greene and Cohen concede that we would have to keep punishing people for practical purposes. Such an account would be consistent with "black box" accounts of economic incentives that simply depend on the relation between inputs and outputs without considering the mind as a mediator between the two. For those who believe that a thoroughly naturalized account of human behavior entails complete consequentialism, such a conclusion might not be unwelcome.

On the other hand, this view seems to entail the same internal contradiction just explored. What is the nature of the "agent" that is discovering the laws governing how incentives shape behavior? Could understanding and providing incentives via social norms and legal rules simply be epiphenomenal interpretations of what the brain has already done? How do "we" "decide" which behaviors to reward or punish? What role does "reason" — a property of thoughts and agents, not a property of brains — play in this "decision"?

If the truth of pure mechanism is a premise in deciding what to do, this premise yields no particular moral, legal or political conclusions. It will provide no guide as to how one should live or how one should respond to the truth of VNC. Normativity depends on reason and thus VNC is normatively inert. If reasons do not matter, then we have no reason to adopt any morals or politics, any legal rule, or to do anything at all.

Given what we know and have reason to do, the allegedly disappearing person remains fully visible and necessarily continues to act for good reasons, including the reasons not yet to accept VNC. We are not Pinocchios and our brains are not Giapettos pulling the strings.

In a world which accepts VNC, who knows what would happen to Collera? We would have no good, potentially action-guiding reason even to care. On the other hand, suppose it is true that we are incapable of not deliberating, no matter what we come to believe in theory. Perhaps the system will look not so different

from the system we have today.

Do Liberty and Autonomy Hang By a Technological Thread?

The more plausible scenario is a world in which we continue to accept that agency is genuine, but in which the science of prediction and control has markedly advanced. This might realistically lead to alternative regimes based on prevention.

Our present scientific understanding of the causation of behavior and the ability to predict and control behavior are still quite limited despite the vast increase in knowledge in recent decades. There are still good reasons to limit a purely consequential response to potentially dangerous people; we will make many errors in prediction; we will intervene unnecessarily in many cases; and we often will not be able to fix the problems involved anyway.

If the power of science to predict and control human behavior becomes vastly more advanced than it currently is, however, that could change. Perfectly conscious, intentional and rational behaviors, for which people can be held responsible, may nonetheless prove controllable and predictable. There is no inconsistency between predictability and control and responsibility. Consequently, we might well decide to move to a pure system of prediction and prevention for social control on consequential grounds. The lure of a safe, harmonious society might irresistibly cause us to abandon notions of guilt and innocence and instead simply to employ concepts like dangerousness and preventive intervention. Responsibility would not be rejected on conceptual grounds; it would simply be treated as normatively irrelevant because the attractions of social engineering were so great.

This system would differ vastly from current practices. Actually causing harm would not be necessary; it would simply be evidence of dangerousness, which we might be able to establish by other means. We would not require any showing that dangerousness resulted from lack of rational capacity. A therapeutic model of intervention would not be necessary either. Predicted dangerousness from any cause would be a sufficient justification for the necessary intervention of any kind. Casting the whole scheme in the language of disorder and therapy might make it all seem more justifiable and less of an intrusion on liberty, but this would simply be a sleight of hand. Widespread screening to identify potentially dangerous people and involuntary intervention would be consistent with such a regime.

We might all be safer in such a world, but would we want to live in it? Would there be any limits on the state's power to predict and control for the good of us all? This is impossible to predict at present, but perhaps traditional notions of liberty and autonomy do hang by a technological thread.

There is a less dystopian version of this scenario, in which the concepts of responsibility and desert-disease jurisprudence are maintained, but in which dangerous agents in either the civil or criminal justice systems are fixed, rather than punished. The term of incarceration would be only as long as necessary to fix

the agent so that he or she could become a safe member of free society again. This sounds humane and it would be vastly more protective of liberty and autonomy than the scheme just presented. Such a system would have problems of its own, however. It would undermine the deterrent value of the criminal law, especially if the interventions were effective, speedy, and not terribly painful or otherwise intrusive. The only reason not to offend would be the stigma of conviction and the inconvenience of the intervention, but these costs might be acceptable to many potential criminals. Moreover, what types of interventions would be acceptable and should they be imposed against the will of the criminal? Over a half-century ago, C.S. Lewis warned against the “humanitarian theory of punishment” because it threatened dignity and respect for persons and it could become the instrument of vast oppression.¹⁹ Those concerns are no less valid today.

In our first regime, we must assume that Collera’s volatile aggressiveness would have been identified very early, perhaps even prenatally, and fixed. In the second regime, the question would be whether we could alter him without his consent or how long we could and would incapacitate him for social safety. It is possible, however, that he would have been fixed after he had committed his first assaultive crime and that the homicide would never have occurred.

Conclusion

It is fun, I suppose, to speculate that neuroscience threatens a radical revision in our moral and legal doctrines and practices, and, indeed, in our lives as a whole. Neuroscience poses no such threat at present, however, and it is unlikely ever to pose that threat unless it decisively resolves the mind-body problem. Until that happens, neuroscience might contribute to the reform of doctrines that do not accurately reflect truths about human behavior, to the resolution of individual cases, and to the efficient operation of various legal practices. If the power to predict and prevent dangerous behavior becomes sufficiently advanced, however, traditional notions of responsibility and guilt might simply become irrelevant.

¹⁹ C.S. Lewis, *The Humanitarian Theory of Punishment*, 6 RES JUDICATAE 224 (1953).

Stephen J. Morse is an expert in criminal and mental health law whose work emphasizes individual responsibility in criminal and civil law. Educated in law and psychology at Harvard, Morse has written for law reviews, journals of psychology, psychiatry, and philosophy; and he has edited collections. He was a contributing author (with Larry Alexander and Kimberly Kessler Ferzan) to *Crime and Culpability: A Theory of Criminal Law* (Cambridge University Press, 2009), he co-edited (with Leo Katz and Michael S. Moore) *Foundations of Criminal Law* (Foundation Press, 2000), and he is working on a new book, *Desert and Disease: Responsibility and Social Control*. Morse was Co-Director of the MacArthur Foundation's Law and Neuroscience Project and he co-directed the Project's Research Network on Criminal Responsibility and Prediction. Morse is a Diplomate in Forensic Psychology of the American Board of Professional Psychology; a past president of Division 41 of the American Psychological Association (the American Psychology-Law Society); a recipient of the American Academy of Forensic Psychology's Distinguished Contribution Award; a member of the MacArthur Foundation Research Network on Mental Health and Law (1988-1996); and a trustee of the Bazelon Center for Mental Health Law in Washington, D.C. (1995-present).

Governance Studies

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

Editors

Jeffrey Rosen
Benjamin Wittes

Production & Layout

John S Seo

**E-mail your comments to
gscments@brookings.edu**

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