Minnesota Journal of Law, Science & Technology

Volume 9 | Issue 1 Article 3

2007

Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience

Stephen J. Morse

Articles

Determinism and the Death of Folk Psychology: Two Challenges To Responsibility from Neuroscience

Stephen J. Morse*

I. INTRODUCTION

Free will and human agency are considered foundational for ascriptions of criminal responsibility in Anglo-American jurisprudence. As United States Supreme Court Justice Oliver Wendell Holmes famously observed, "even a dog distinguishes

* Ferdinand Wakeman Hubbell Professor of Law & Professor of Psychology and Law in Psychiatry, University of Pennsylvania.

This paper was first presented as a Deinard Memorial Lecture in Law and Medicine cosponsored by the University of Minnesota's Joint Degree Program in Law, Health & the Life Sciences and the Center for Bioethics with major support from the law firm of Leonard, Street and Deinard, and then presented at a conference on "Neuroethics and Empirical Moral Psychology" at the Department of Philosophy, Classics, History of Art and Ideas of the University of Oslo. I would like to thank Leonard, Street and Deinard and Susan Wolf for their hospitality in Minneapolis, and Jakob Elster, Lene Bomann-Larsen and Øistein Schmidt Galaaen for their hospitality in Oslo. Thanks to Barry Feld, Vidar Halvorsen and Monica Luciano for helpful commentary. The paper was also presented at a symposium on "Neuroscience and Moral and Legal Responsibility" at the University of Cincinnati College of Law and School of Medicine, as a plenary address at the 2007 annual meeting of the Association for Politics and the Life Sciences, and as a lecture at the Sage Center for the Study of the Mind at the University of California, Santa Barbara. I thank all those at each of these venues who made helpful comments, especially John Bickle, Mike Gazzaniga and Douglas Mossman. I also thank Ed Greenlee for his invaluable assistance. As always, I thank my personal attorney, Jean Avnet Morse, for her sound, sober counsel and moral support.

MORSE S. DETERMINISM AND THE DEATH OF FOLK PSYCHOLOGY: TWO CHALLENGES TO RESPONSIBILITY FROM NEUROSCIENCE. MINN. J.L. SCI. & TECH. 2008;9(1):1-36.

between being stumbled over and being kicked."[1] And, as Justice Jackson wrote in Morissette v. U.S., concisely noting both conditions:

> The contention that an injury can amount to a crime only when inflicted by intention is no provincial or transient notion. It is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil. A relation between some mental element and punishment for a harmful act is almost as instinctive as the child's familiar exculpatory "But I didn't mean to". . . .[2]

Now, however, the discoveries of the new neuroscience challenge both foundations for responsibility.[3] The new neuroscience seems poised to demonstrate that our behavior is determined by physical events in the brain and that we therefore cannot be responsible. Neuroscientific discoveries also are alleged to demonstrate that mental states do not causally explain our behavior. If this is true, it provides another, independent ground for the claim that responsibility is impossible.

I argue that neither challenge succeeds. The challenge to free will from neurophysical determinism is familiar to similar challenges in the past, but it fails for three reasons. First, free will is not a criterion for the application of any legal rule. Second, free will is not foundational for criminal responsibility. Third, there is a philosophically plausible response to those who claim that determinism—whether based on the theories and findings of neuroscience or any other discipline—and responsibility are incompatible. Thus, I conclude that, for the moment, the positive doctrines of legal and moral responsibility are normatively safe from the newest metaphysical assault. The neuroscientific attack on agency is more troubling because it claims that the presumptions of morality and the law about human agency are inconsistent with our new understanding of the link between brain and behavior. Roughly speaking, the law implicitly adopts the folk-psychological model of the person,

1. OLIVER WENDELL HOLMES, JR., THE COMMON LAW 7 (Transaction ed. 2005) (1963).

2. Morissette v. United States, 342 U.S. 246, 250–51 (1952).

3. I recognize that legal and moral responsibility need not coincide. Because Anglo-American punishment theory holds that desert is a necessary condition for punishment—at least at the core of criminal law—I shall assume that moral and legal responsibility do coincide in the core.

MORSE S. DETERMINISM AND THE DEATH OF FOLK PSYCHOLOGY: TWO CHALLENGES TO RESPONSIBILITY FROM NEUROSCIENCE. MINN. J.L. SCI. & TECH. 2008;9(1):1-36.

which explains behavior in terms of desires, beliefs and intentions. If practical reason plays no role in explaining our behavior, as some neuroscientists and others claim, current responsibility doctrines and practices would have to be radically altered or jettisoned altogether. I suggest, however, that the conceptual and scientific support for this argument is thin at present and that there is good ground to believe that our conception of persons as agents is unlikely to disappear. Consequently, legal and moral doctrines that depend on agentic personhood are secure—at least for now.

The paper proceeds as follows: part II briefly describes the law's concept of the responsible person and demonstrates that free will is not a criterion of legal responsibility in general and criminal responsibility in particular. Thus, even if determinism or universal causation is a true account of the metaphysics of the universe, this truth cannot cast doubt generally on specific legal criteria. Part III considers the metaphysical free will problem and shows that free will is also not foundational for responsibility. Neurophysical determinism is akin to genetic, psychological, or sociological determinism and subject to the same compatibilist responses that have deflected earlier deterministic challenges. In sum, determinism is irrelevant to positive law, even if it is true. Part IV addresses the allegedly disappearing person. It reconsiders the folk-psychological concept of the person embedded in law and morality and identifies philosophical arguments and types of evidence that would justify abandoning the commonsense conception. It then canvasses the evidence for the current assault on agency from neuroscience and allied disciplines and concludes that folk psychology is alive and well. The person is most definitely visible. Part V is a brief conclusion.

II. THE NON-PROBLEM OF FREE WILL IN LAW

The genuine metaphysical problem about free will concerns, roughly, whether human beings possess the ability or power to act uncaused by anything other than themselves. Such ability is often termed libertarian freedom, contra-causal freedom, agent origination and the like. But there is no such problem in law. The importance of having this power or ability results from the controversial belief that libertarian freedom underwrites the possibility of holding people genuinely

responsible.[4] Solving the free will problem would have profound implications for responsibility doctrines and practices, such as blame and punishment, but, at present, having or lacking libertarian freedom is not a criterion of any civil or criminal law doctrine. Law addresses problems genuinely related to responsibility, including consciousness, the formation of mental states such as intention and knowledge, the capacity for rationality, and compulsion, but it never addresses the presence or absence of free will.[5] People sometimes use "free will" loosely to refer to genuine responsibility doctrines, but this distracts from the real issues and perpetuates confusion. The only practical free will problem in law is the confusion among lawyers, scientists and others who think that free will is a legal criterion or who speak and write as if it is.

A. THE LEGAL VIEW OF THE PERSON

Consciousness and action are central to the law's view of the person. The capacity for intentional activity or stillness—the capacity for agency—is a central aspect of personhood and is integral to what it means to be a responsible person. We act because we intend. Responsibility judgments depend on the mental states that produce and accompany bodily movement and stillness. This is how we think about ourselves, and this is the concept of the person that morality and law both reflect.

The law's view of the person is thus the so-called "folk-psychological" model: a view of the person as a conscious (and potentially self-conscious) creature capable of practical reason,

---

4. *See* ROBERT KANE, A CONTEMPORARY INTRODUCTION TO FREE WILL 1–5 (2005); GEORGE SHER, IN PRAISE OF BLAME, at viii-ix (2006); Editorial, *Free to Choose? Modern Neuroscience Is Eroding the Idea of Free Will*, ECONOMIST, Dec. 23, 2006, at 16.

5. It is virtually impossible to prove a negative, but perusal of any American criminal code or judicial opinions will confirm the absence of libertarian free will as a genuine criterion. On rare occasions, a statute might include the phrase. *See, e.g.*, CAL. PENAL CODE § 261.6 (2007) (Consent to sexual activity must be "pursuant to an exercise of free will."). It is clear, however, that free will in such instances simply is a proxy for more familiar, less metaphysical criteria, such as the absence of compulsion. Judges, too, sometimes write as if freedom of the will were a foundation for responsibility. *See* Morissette v. United States, 342 U.S. 246, 250–51 (1952). Such locutions are either unanalyzed boilerplate or once again are proxies for more familiar responsibility criteria. Again, Part III *infra* explains why libertarian free will is not required for criminal responsibility.

an agent who forms and acts on intentions that are the product of the person's desires and beliefs. We are the sort of creatures that can act for and respond to reasons. The law properly treats persons generally as intentional creatures and not as mechanical forces of nature. Law and morality are action-guiding[6] and could not guide people ex ante and ex post unless people could use rules as premises in their practical reasoning. Otherwise, law and morality as action-guiding normative systems of rules would be useless, and perhaps incoherent. Law is a system of rules that, at the least, is meant to guide or influence behavior and thus to operate as a potential cause of behavior. As John Searle wrote,

> Once we have the possibility of explaining particular forms of human behavior as following rules, we have a very rich explanatory apparatus that differs dramatically from the explanatory apparatus of the natural sciences. When we say we are following rules, we are accepting the notion of mental causation and the attendant notions of rationality and existence of norms. . . .
>
> . . . The content of the rule does not just describe what is happening, but plays a part in *making it happen.*[7]

Legal and moral rules are not simply mechanistic causes that produce "reflex" compliance. They operate within the domain of practical reason. Agents are meant to and can only use these rules as potential reasons for action as they deliberate about what they should do.[8] Moral and legal rules are thus action-guiding primarily because they provide an agent with good moral or prudential reasons for forbearance or action. Unless people are capable of understanding and then using legal rules as premises in deliberation, law would be powerless to affect human behavior.[9] People use legal rules as

---

6. SHER, *supra* note 4, at 123 (stating that although philosophers disagree about the requirements and justifications of what morality requires, there is widespread agreement that "the primary task of morality is to guide action"); John R. Searle, *End of the Revolution*, 49 N.Y. REV. OF BOOKS 33, 35 (2002).

7. Searle, *supra* note 6, at 35.

8. Scott J. Shapiro, *Law, Morality, and the Guidance of Conduct*, 6 LEGAL THEORY 127 (2000).

9. *Id*. at 131. This view assumes that law is sufficiently knowable to guide conduct, but a contrary assumption is largely incoherent. As Shapiro writes:

> Legal skepticism is an absurd doctrine. It is absurd because the law cannot be the sort of thing that is unknowable. If a system of norms were unknowable, then that system would not be a legal system. One important reason why the law must be knowable is that its function is to guide conduct. *Id*.

premises in the practical syllogisms that guide much human action. No instinct governs how fast a person drives on the open highway. However, among the various explanatory variables, the posted speed limit and the belief in the probability of suffering the consequences for exceeding it surely play a large role in the driver's choice of speed. Human behavior can be modified by means other than influencing deliberation and human beings do not always deliberate before they act. Nonetheless, law operates through practical reason, even when we most habitually follow the legal rules. Law can directly and indirectly affect the world we inhabit only by its influence on practical reason.[10]

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[11] and, in principle, responsive to reason. 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.

I do not assume that legal rules are always clear and thus capable of precise action guidance. If most rules in a legal system were not sufficiently clear most of the time, however, the system could not function. Further, the principle of legality dictates that criminal law rules should be especially clear.

10. *Id*. at 131–32.

11. I do not mean to imply dualism here. I am simply accepting the folk-psychological view that mental states—which are fully produced by and realizable in the brain—play a genuinely causal role in explaining human behavior. *See infra* Part IV.

B. THE GENERAL LEGAL CRITERIA FOR RESPONSIBILITY AND EXCUSE

The law's concept of responsibility follows logically from the nature of law itself and its concept of the person. As a system of rules that guides and governs human interaction, law tells citizens what they may and may not do, what they must or must not do, what abilities are required competently to perform certain tasks, and what consequences will follow from their conduct. If human beings were not rational creatures who could understand the good reasons for action, including the relevant facts and rules, and were not capable of conforming to legal requirements through intentional action or forbearance, the law would be powerless to affect human action.[12] Legally responsible agents are therefore people who have the general capacity to grasp and be guided by good reason in particular legal contexts.[13] For example, they must be generally capable of properly using the rules as premises in practical reasoning. The usual legal presumption is that most adults are so capable.

Note that the law requires possession of a general capacity at the time in question rather than an exercise of that capacity. Failure to exercise a capacity does not necessarily mean that one lacks that capacity. Indeed, acting irrationally and foolishly is common even among people with the greatest capacity for rational conduct. Under the law, if a person is capable of exercising the capacity for rationality if there is good reason to do so—as there always is when important interests are at stake—then that person may be held responsible even if she failed to exercise that capacity.

The general capacity for rationality is not self-defining. It must be understood according to some contingent, normative notion both of rationality and of how much capability is required. For example, legal responsibility might require the capability of understanding the reason for an applicable rule, as well as the rule's command and the consequence for failure to comply. These are matters of moral, political and, ultimately, legal judgment, about which reasonable people differ. There is no uncontroversial definition of rationality or of the type and amount required for responsibility in various legal contexts. These are normative issues and, whatever the

12. R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 51–63 (1994).

13. *Id.* at 74–83.

outcome might be within a polity and its legal system, the debate is about human action—intentional behavior that is responsive to reasons.

Now let us turn to the law's excusing conditions, those situations in which the agent will not be held responsible and may be treated specially. If the general capacity for rationality is the primary responsibility condition, then lack of that capacity is the primary excusing condition. It explains, for example, why young children, some people with dementia, and some people with mental disorder are not held responsible, at least in some contexts. Again, how much lack of capacity is necessary to find the agent not responsible is a normative moral, social, political, and ultimately legal issue. It is not a scientific, medical, psychological or psychiatric issue.[14]

Compulsion or coercion is also an excusing condition, but it is less well-understood than lack of rational capacity.[15] It is useful first to distinguish cases of literal and metaphorical compulsion. Literal compulsion exists when the person's bodily movement is a pure mechanism, the product of mechanistic causes and is not rationalizable by the agent's desires, beliefs and intentions.[16] For example, if a much stronger person pulls an agent's arm and literally forces it against the head of a third person despite the agent's best efforts not to move his arm, the movement of the arm is not the agent's act at all. It is not a product of his or her intention. For another example, a tremor produced by a neurological disorder is not an action because it is not intentional. Again, the movement of the arm is pure mechanism.

In contrast, 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.[17] In cases of metaphorical compulsion, it is useful to distinguish two party and one party cases. The former occurs when another person threatens an agent by placing him or her in a "do-it-or-

14. Scientific or clinical evidence may help resolve legal questions, but the ultimate issue is always legal and not scientific or clinical.

15. *See* Stephen J. Morse, *Uncontrollable Urges and Irrational People*, 88 VA. L. REV. 1025, 1066 (2002) (addressing the difficulty with compulsion excuses and suggesting that most cases seeming to require such an excuse are better understood in terms of irrationality).

16. *Id.* at 1055–63.

17. *Id.*

else" situation. 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 the innocent other. If the agent does kill, however, the killing is fully intentional and a perfectly rational response to the desire to live and to the belief that she would otherwise be killed. Although it would be wrong to kill under these circumstances, the law may decide as a normative matter that the act should be excused because the agent was motivated by a threat so great that it would be supremely difficult for most citizens to resist.

One party or internal compulsion cases are more difficult to understand. Recognize, first, that if the internally compelled agent is irrational, there is no need for an independent compulsion excuse. Only in cases in which the agent is seemingly rational do we need to consider compulsion. The cases that most fit this category are "disorders of desire," such as addictions, paraphilias, compulsive gambling, pyromania, and the like. Note again, however, that when the addict seeks and uses substances, when the pedophile molests a child, when the compulsive gambler places a bet, and when the pyromaniac sets a fire, the agent is acting. She acts intentionally to satisfy her craving for the purpose of achieving relief, of obtaining pleasure, or of both. In these cases, 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. Consequently, although the agent undeniably acted in response to her desires, morality or the law may again hold that metaphorically and normatively she was compelled and should be excused.

Note that none of the law's general criteria for responsibility or excuse refers to free will or its absence. Lack of action, lack of rationality, and compulsion all excuse, but none of these conditions has anything to do with libertarian freedom of the will. There may be problems conceptualizing and evaluating the lack of rational capacity or compulsion. These are real problems for law, but they are not free will problems. Lawyers, scientists and forensic practitioners often speak and write as if these are "free will" problems, as if lack of free will were a synonym for lack of action, lack of rational capacity, or compulsion. Nevertheless, free will is doing no work whatsoever, independent of these genuine excusing conditions, and it thus threatens to confuse the issues.

To explore further the error of believing that there is a genuine and independent free will problem in positive law, let us turn specifically to criminal responsibility, which is the legal context in which talk of free will is probably most common (and most distracting).

C. CRIMINAL RESPONSIBILITY

Current common law holds an agent prima facie criminally responsible if the agent acts intentionally and with the appropriate mental state, the mens rea, required by the definition of the offense, such as purpose, knowledge, recklessness, or negligence. Criminal law typically defines an act as an intentional bodily movement performed by an agent whose consciousness is reasonably intact. Mental states have their ordinary language, common sense meanings. No degree of commitment or rationality is included in the definitions of mens rea. An act committed ambivalently for irrational reasons is considered intentional if it was done on purpose. Even if the agent is fully prima facie responsible, however, the agent ultimately may still not be criminally responsible if an excusing condition, an affirmative defense, such as legal insanity (essentially a rationality defect) or duress (a compelling "hard choice" situation, such as a "do-it-or-else" threat at gunpoint) was present when the agent committed the offense. Although one might quibble about details, and there is substantial variation across jurisdictions, this account accurately reflects Anglo-American law's current, core conception of criminal responsibility.[18]

The logic of the foregoing account is that a defendant who wants to avoid imputation of criminal responsibility must create reasonable doubt about whether he or she acted intentionally, consciously, and with the required mens rea, or the defendant must establish an affirmative defense.[19] If the

18. Any good treatise will bear out this claim. *See, e.g.*, JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW (4th ed. 2006).

19. I use the vague locution, "establish an affirmative defense," because the Supreme Court has made clear that the burden of persuasion for affirmative defenses may be placed on the defendant. *See, e.g.*, Leland v. Oregon, 343 U.S. 790, 799 (1952) (permitting shifting the persuasion burden for legal insanity to the defendant); Dixon v. United States, 126 S. Ct. 2437, 2442 (2006) (permitting shifting the persuasion burden for duress to the defendant). The defendant will effectively retain the production burden,

agent does not act at all because the bodily movement was not intentional or the agent's consciousness was substantially compromised, the agent is not prima facie responsible. For example, a reflex or behavior in an altered state of consciousness, such as sleepwalking, will not be considered the defendant's action, even if the defendant's bodily movements caused a harm.[20] An agent who does not act is acquitted outright. Similarly, if the agent lacks a requisite mental state, the agent is also not prima facie criminally responsible and must be acquitted outright of the crime requiring that mental state. For example, suppose a defendant shoots a creature he actually believes to be a space alien impersonating a police officer, but the victim turns out to be a genuine police officer. The shooting was an intentional action, but the defendant cannot be convicted of intentional homicide of a person knowing that the victim was a police office because he did not intend to kill a human being and did not know the victim was a police officer.[21] If he or she was sufficiently careless, however, the defendant might be convicted of negligent homicide, defined as killing in a situation in which a reasonable person should have recognized that his or her conduct created a substantial and unjustifiable risk of death to a human being.[22]

Like the definitions of crimes, affirmative defenses also have specific criteria. Consider first the insanity defense, using the Model Penal Code test as an example. To be found legally insane, at the time of the crime the defendant must have been suffering from a mental disorder and, as a result, lacked substantial capacity to appreciate the criminality of his action or to conform his action to the requirements of the law.[23] Thus, if the defendant is not sufficiently disordered to meet the law's definition of mental disorder, or did not lack the requisite substantial capacity, a legal insanity defense will fail. Now, consider the Model Penal Code standard for the affirmative defense of duress. Duress is established if the defendant is threatened with unlawful force—usually death or grievous bodily harm—against his person unless he harms another, and

---

however, whether or not the state shifts the persuasion burden.

20. DRESSLER, *supra* note 18, at 95.

21. *See* Clark v. Arizona, 126 S. Ct. 2709, 2717–18 (2006) (stating that Defendant Clark claimed that he suffered from paranoid schizophrenia and held these beliefs).

22. DRESSLER, *supra* note 18, at 583–84.

23. MODEL PENAL CODE § 4.01(1) (Proposed Official Draft 1962).

a person of reasonable firmness would have yielded in this situation.[24] The defense will fail if the threat was of lesser force or a threat to destroy only the defendant's valued property, or if a person of reasonable firmness would not have yielded, say, killing five people to save one's own life.

To establish prima facie guilt or to defeat an affirmative defense, the prosecution need not prove that the defendant had free will. To defeat the prosecution's prima facie case, the defendant must simply cast reasonable doubt on the elements of conscious, intentional action and mens rea. To establish an affirmative defense, the defense must introduce sufficient evidence of the criteria for the defense. To avoid criminal responsibility either by negating the prosecution's prima facie case or by establishing an affirmative defense, the defendant need not demonstrate that he or she lacked free will. People often say that a defendant who acted under duress or who was legally insane lacked free will. In such cases, however, free will is simply a confusing and conclusory way to say that the genuine legal criteria for excuse were met. Lack of free will, independent of the behavioral legal criteria for excuse, does not explain why such a defendant is excused.

For a final confirmation of the thesis that free will plays no role in the positive criteria for criminal responsibility, consider the United States Supreme Court's recent decision, *Clark v. Arizona*, in which the Supreme Court had a rare opportunity to clarify the relation between mens rea and insanity.[25] The questions presented were whether Arizona's unusually narrow insanity defense test, which asked only if the defendant could distinguish between right and wrong, violated substantive due process rights and whether an Arizona rule that excluded virtually all expert evidence concerning mental disorder offered for the purpose of negating mens rea violated procedural due process.[26] Legal insanity and the presence of mens rea are probably the criminal law issues to which free will is allegedly most relevant, and there was extensive discussion of the history of legal insanity and of the role of mens rea.[27] Nevertheless, the Court mentions the term "free will" only

24. *Id*. § 2.09(1).
25. Clark v. Arizona, 126 S. Ct. 2709 (2006).
26. *Id*. at 2716.
27. *Id*. at 2720–22, 2724–26.

once, and only then because it is included in a quotation from the American Psychiatric Association's Insanity Defense Work Group, which claimed that psychiatrists who give ultimate opinions about legal or moral constructs such as free will exceed their expertise and are likely to confuse juries.[28] Libertarian free will as a criterion of or foundation for criminal responsibility was not discussed. There were many problems with the Court's analysis,[29] but failure to discuss free will was not among them.

In short, free will or lack of it is not a criterion for criminal responsibility or non-responsibility. Once again, it is irrelevant to the actual practice of criminal law. It is true that people, including judges, practicing lawyers and a few law professors, talk as if free will were important in criminal law, but this is clearly wrong as a matter of positive law. They sometimes mean, however, that free will is a necessary foundational justification for responsibility, even if it is not a discrete criterion for any legal doctrine. The next part of this article, which discusses the metaphysical free will problem, demonstrates that having libertarian free will is not necessary to justify responsibility doctrines and practices according to an entirely plausible and practical resolution of the metaphysical free will problem.

III. THE GENUINE FREE WILL PROBLEM AND RESPONSIBILITY

Many people believe that libertarianism is a foundational assumption for law. They believe that responsibility is possible only if we genuinely possess contra-causal freedom. Thus, if we do not have this extraordinary capacity, they fear that many legal doctrines and practices, especially those relating to responsibility, may be entirely incoherent. As we shall see, however, metaphysical libertarianism is not a necessary support for current responsibility doctrines and practices.

Only a small number of philosophers and scientists believe that human beings possess libertarian freedom of action and will, which has been termed a "panicky" metaphysics[30] because

28. *Id.* at 2736.

29. Stephen J. Morse & Morris B. Hoffman, *The Uneasy Entente Between Insanity and Mens Rea: Beyond* Clark v. Arizona, J. CRIM. L. & CRIMINOLOGY (forthcoming 2008).

30. P.F. Strawson, *Freedom and Resentment*, *in* FREE WILL 59, 80 (G.

it is so implausible.[31] Most philosophers and scientists believe that the universe is deterministic or universally caused, or nearly so, especially above the sub-atomic level. There is no uncontroversial definition of determinism and we will never be able to confirm that it is true or not. As a working definition, however, let us assume, roughly, that all events have causes that operate according to the physical laws of the universe and that were themselves caused by those same laws operating on prior states of the universe in a continuous thread of causation going back to the first state.

Even if this assumption is too strong, the universe seems so sufficiently regular and lawful that rationality demands that we must adopt the hypothesis that universal causation is approximately correct.[32] If determinism is true, the people we are and the actions we perform have been caused by a chain of causation over which we mostly had no rational control and for which we could not possibly be responsible. People do not have contra-causal freedom. How can responsibility be possible for action or for anything else in such a universe? How can it be rational and fair for civil and criminal law to hold anyone accountable for anything, including blaming and punishing people because they allegedly deserve it?

It is important to understand that, for the determinist, biological causes, including those arising from the brain, pose no new or more powerful general metaphysical challenge to responsibility than non-biological or social causes. As a conceptual and empirical matter, humans do not necessarily have more control over psychological or social causal variables than over biological causal variables. More important, in a world of universal causation or determinism, biological causation creates no greater threat to a person's life hopes than psychological or social causation. For purposes of the metaphysical free will debate, a cause is just a cause, whether it is neurological, genetic, psychological, sociological, or astrological. Neuroscience is simply the newest "bogey" in a dispute about the general possibility of responsibility that has

---

Watson ed., 1982).

31. HILARY BOK, FREEDOM AND RESPONSIBILITY 42–51 (1998).

32. Galen Strawson, *Consciousness, Free Will, and the Unimportance of Determinism*, 32 INQUIRY 3, 12 (1989) (terming this hypothesis the "realism constraint").

been ongoing for millennia. Though it is more scientifically respectable than earlier bogeys, such as astrology and psychoanalysis, and appears to produce very compelling graphic representations of the brain,[33] neuroscience evidence for causation does no more work in the general free will/responsibility debate than other kinds of causal evidence.

No analysis of the determinism/responsibility problem could conceivably persuade everyone. There are no decisive, analytically incontrovertible arguments to resolve the metaphysical question. Moreover, the question is metaphysical, not scientific. Indeed, the debate is so fraught that even theorists who adopt the same general approach to the metaphysical challenge substantially disagree. Nevertheless, as I shall argue, compatibilism is a plausible metaphysical

33. These brain graphics are almost always misleading to those who do not understand how they are constructed. We have long become accustomed to seeing in the media and elsewhere what appear to be pictures of the brain with various superimposed shaded areas (which can be in any color the investigator wishes), indicating which region was activated by the stimulus presented. These are not pictures of the brain, however, and they do not necessarily reflect the activity in any individual brain. The underlying brain image is a "standardized" brain structure, whereas we know that there is enormous variation in the structure of individual brains, much as there is enormous variation in all the biological structures of the body. The superimposed shaded areas indicate that in that region there was a change in activity from base rate activity, but the activity being measured is not neural activity. Rather, the shading discloses (with a short time lag) changes in blood oxygenation, which is thought to be a good proxy for neural activation in that area. Finally, the shaded areas do not show actual activity. Instead, they are computer generated statistical postdictions of the likelihood that the change was produced by the stimuli rather than by a random fluctuation in activity level. In a sense, these images are pictorial representations of a confidence interval, and they represent an average that may not be the true value for any individual subject in the study. *See* CHRIS FRITH, MAKING UP THE MIND: HOW THE BRAIN CREATES OUR MENTAL WORLD 116–17 fig.CP2 (2007).

It is also the case that when a claim is allegedly backed by brain imaging evidence, non-experts rate the scientific reasoning supporting the claim to be superior to such reasoning based on equally persuasive evidence of other sorts. David P. McCabe & Alan D. Castel, Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning (2007) (unpublished manuscript, on file with Cognition). Moreover, non-experts are more likely to believe the claim even if the neuroscience is actually irrelevant to the logic of the explanation. Deena Skolnick Weisberg et al., *The Seductive Allure of Neuroscience Explanations*, 20 J. COGNITIVE NEUROSCI. 1, 6 (2008). Much of this misunderstanding has been attributed to misleading reports of neuroscience in the media. Eric Racine et al., *fMRI in the Public Eye*, 6 NATURE REVIEWS NEUROSCI. 159, 159–60 (2005). *See infra* text accompanying notes 56–61 for a more recent example.

contender in the debate and thus we are warranted in thinking that determinism is not inconsistent with facts about human behavior we have reason to believe and with legal and moral practices we adopt in light of those facts.

A. HARD DETERMINISM

Hard determinism holds that free will and determinism are inconsistent, that free will is necessary for moral responsibility, that determinism is true, and therefore that no one can be morally responsible for one's actions.[34] This theory does not try either to explain or to justify our responsibility concepts and practices. It simply assumes that genuine responsibility is metaphysically unjustified. For example, a central hard determinist argument is that people can be responsible only if they could have acted otherwise than they did, but they could not have acted differently if determinism is true.[35] Consequently, the hard determinist claims that even if an internally coherent account of responsibility and related practices can be given, it will be a superficial basis for responsibility, which is only an illusion.[36] There is no real or ultimate responsibility. Hard determinists properly concede that western systems of law and morality hold some people accountable and excuse some people, but they argue there is no genuinely justifiable basis for distinguishing responsible from non-responsible people. Hard determinism thus provides an external critique of responsibility. If determinism is true and is genuinely inconsistent with responsibility, then no one can ever be "really" or "ultimately" responsible for anything and responsibility attributions cannot properly justify further action. The question, then, is whether as rational agents we must swallow our pride, accept hard determinism because it is so self-evidently true and somehow transform the legal system and our moral practices accordingly.

---

34. SAUL SMILANSKY, FREE WILL AND ILLUSION 4 (2000).

35. This is sometimes called the "principle of alternate possibilities." It has generated endless disputes between incompatibilists, who believe it is flatly inconsistent with responsibility, and compatibilists, who believe that it is not inconsistent with responsibility. *See* WALLACE, *supra* note 12, at 115–17, 251–65.

36. *See* SMILANSKY, *supra* note 34, at 40–73, 145–219 (arguing that free will is an illusion, but an illusion that is indispensable).

B. COMPATIBILISM

Compatibilists, who agree with incompatibilist hard determinists that determinism is true, have three basic answers to the incompatibilist challenge. First, they claim that responsibility attributions and related practices are human activities constructed by us for good reason and that they need not conform to any ultimate metaphysical facts about genuine or "ultimate" responsibility.[37] Indeed, some compatibilists deny that conforming to ultimate metaphysical facts is even a coherent goal in this context.[38] Second, compatibilism holds that positive doctrines of responsibility are fully consistent with determinism.[39] Third, compatibilists believe that responsibility doctrines and practices are normatively desirable and consistent with moral, legal, and political theories that we firmly embrace.[40] The first claim is theoretical; the third is primarily normative. There are very powerful arguments for the first and third claims. For the present purpose, however, which is addressed to whether free will is foundational for law, the second claim is the most important.

Let us begin with the most general responsibility and excusing conditions. Recall that the capacity for rationality is the primary responsibility criterion and its lack is the primary excusing condition. Human beings have different capacities for rationality in general and in specific contexts. For example, young children in general have less developed rational capacity than adults. It is also true that rationality differences affect agents' capacity to grasp and to be guided by good reason. Differences in rational capacity and its effects are real even if determinism is true.[41] Compulsion is also an excusing condition, but it is simply another fact about human beings that some people act in response to external or internal hard choice threats to which persons of reasonable firmness might yield and most people, most of the time, are not in such situations when they act. This is true even if determinism is true and even if people could not have acted otherwise.

---

37. Stephen J. Morse, *Reason, Results and Criminal Responsibility*, 2004 U. ILL. L. REV. 363, 438–39.

38. *Id.*

39. *Id.* at 440–42.

40. *Id.* at 443–44; *see* James Lenman, *Compatibilism and Contractualism: The Possibility of Moral Responsibility*, 117 ETHICS 7 (2006).

41. *See* Morse, *supra* note 37, at 441.

For a specific example, consider again the doctrines of criminal responsibility. Assume that the defendant has caused a prohibited harm. Prima facie responsibility requires that the defendant's behavior was an act and performed with a requisite mental state. Now it is true that some bodily movements are intentional and performed in a state of reasonably integrated consciousness and some are not. It is also true that some defendants possess the requisite mental state and some do not. The truth of determinism does not entail that actions are indistinguishable from non-actions or that different mental states may accompany action. These facts are true and make a perfectly rational legal difference even if determinism is true. Determinism is fully consistent with prima facie guilt and innocence.

Now consider the defenses of insanity and duress. Some people with mental disorders do not know right from wrong; others do. In cases of potential duress, some people face a hard choice that a person of reasonable firmness would yield to and most people do not. Once again, these differences make perfect sense according to dominant retributive and consequential theories of punishment. A causal account can explain how these variations were caused to occur, but it does not mean that these variations do not exist. Determinism is fully consistent with both the presence and absence of affirmative defenses. In sum, the legal criteria used to identify which defendants are criminally responsible map onto real behavioral differences that justify differential legal responses.

In their widely-noted paper, Joshua Greene and Jonathan Cohen take issue with the foregoing account of the positive foundations of legal responsibility and with my claim that it is the "fundamental psycholegal error" to claim that causation of behavior is per se an excusing condition for that behavior.[42] They suggest that, despite the law's official position, most people hold a dualistic, libertarian view of the necessary conditions for responsibility because "vivid scientific information about the causes of criminal behaviour leads people to doubt certain individuals' capacity for moral and legal

42. Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y B: BIOLOGICAL SCI. 1775, 1778 (2004).

responsibility . . . ."[43] To prove their point, they use the hypothetical of "Mr. Puppet," a person who has been genetically and environmentally engineered to be a very specific type of person.[44] They correctly point out that Mr. Puppet is no different from any other person accused of a crime,[45] say, an identical person I shall call Mr. Puppet2, who became the same sort of person without intentional intervention. Yet most people might believe that Mr. Puppet is not responsible. If so, should Mr. Puppet2 also not be responsible because he is also a product of a gene/environment interaction? Would it not then follow, as Greene and Cohen claim, that no one is responsible?

Green and Cohen are right about ordinary peoples' intuitions, of course, but people make the fundamental psycho-legal error all the time. This is a sociological observation and not a justification for thinking causation or determinism does or should excuse behavior. After all, if causation were an excusing condition in a world of universal causation, no one could ever be responsible. This theory is inconsistent with the positive doctrines and practices of law and morality. Moreover, if Mr. Puppet and Mr. Puppet2 are both rational agents, my argument suggests that they are both justifiably held responsible. The lure of purely mechanistic thinking about behavior when causes are discovered is powerful, but should be resisted.

At present, the law's "official" position—that conscious, intentional, rational and uncompelled agents may properly be held responsible—is justified unless and until neuroscience or any other discipline demonstrates convincingly that humans are not the creatures we think we are. That is, if humans are not conscious and intentional creatures who act for reasons that play a causal role in our behavior, then the foundational facts for responsibility ascriptions are mistaken.[46] If it is true, for example, that we are all automata, then no one is an agent, no one is acting and, therefore, no one can be responsible for action. This challenge, which is powerfully fueled by stunning advances in neuroscience, is empirical and in principle capable of resolution. Let us therefore turn to the problem of the

43. *Id.* at 1776, 1779.

44. *Id.* at 1780.

45. *Id.*

46. Greene and Cohen make this claim as well, which I discuss in Part IV *infra*. *See* Greene & Cohen, *supra* note 42, at 1784.

allegedly "disappearing person."

IV. THE DEATH OF FOLK PSYCHOLOGY AND THE DISAPPEARING PERSON

Criminal law's view and the ordinary, common-sense view of action is that it is performed by an agent or person who acts for reasons that cause and explain the agent's conduct. Whether one explains action causally or holistically, the theory of action presupposes that it is a person that acts based on the person's desires, beliefs, and intentions.[47] Agents are praised and blamed, rewarded and punished. Because it is an agent who acts, it makes sense to ask that person to give an account of his or her behavior and to be held accountable. Asking a creature or a mechanistic force that does not act to answer to charges does not make sense. The core of agency as the capacity to act for reasons is accepted as foundational for responsibility. In this Part, I first address the conceptual and speculative issues concerning agency, then I consider the empirical evidence for thinking that we are not agents. Next, I turn to the implications of potentially accepting the death of folk psychology. I conclude that there is little reason at present to believe that we are not agents.

A. CONCEPTUAL ISSUES

The law's fundamental presuppositions about personhood and action are open to profound objection. Most fundamentally, action and consciousness are scientific and conceptual mysteries.[48] We do not know how the brain enables the mind[49] 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

47. *See* ROBERT AUDI, ACTION, INTENTION AND REASON 109–78 (1993).

48. *See id.* at 1–4 (describing the "basic philosophical divisions" in each of the four major problem areas in action theory); COLIN MCGINN, THE MYSTERIOUS FLAME: CONSCIOUS MINDS IN A MATERIAL WORLD (1999) (describing the immense difficulty of explaining consciousness and doubting the ability of human beings to do so).

49. PAUL R. MCHUGH & PHILIP R. SLAVNEY, THE PERSPECTIVES OF PSYCHIATRY 11–12 (2d ed. 1998).

another.[50] How can such tenuously understood concepts be justifiable premises for legal practices such as blaming and punishing? 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.

What if agency and folk psychology are an illusion? What if all of the contending conceptions depend on a mistake about human activity? What if, for example, reasons for actions and intentions, agents' conscious understandings of their world and themselves do not explain actions but are simply post-hoc rationalizations that "make sense of" the bodily motions or non-motions that brains produce? Some people, including many psychologists and neuroscientists, think that new discoveries about the causation of behavior are leading inexorably to a purely mechanistic view of the link between the brain and behavior, and thus to a purely mechanistic view of human behavior. The assassin did not shoot that gun; it was his finger that pulled the trigger, his peripheral neurons that caused his finger muscles to contract, central neurons that caused the peripheral neurons to fire, and so on. The assassin's desires, beliefs, and intentions did no genuine work in explaining his action. These are the thoughts that terrify many thinking people about scientific advances in the understanding of human behavior. This is a real challenge.

Many investigators in psychology and the neurosciences increasingly assert the challenge to agency. 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.[51] Here is an extensive quotation from the conclusion, which is necessary to obtain the tenor of the assertion and to evaluate if it is internally logical:

---

50. It is almost impossible not to talk "dualistically" in ordinary speech and writing. Every time a monist neuroscientist uses a personal pronoun in speaking or writing, for example, he seems to imply that there is a genuine him that is somehow distinguishable from his brain activity. This does not mean, however, that the neuroscientist (or anyone else) is really a crypto-dualist. It is simply an inevitable feature of current language, and perhaps it always will be.

51. DANIEL M. WEGNER, THE ILLUSION OF CONSCIOUS WILL (2002); *see also* Daniel M. Wegner, *Précis of The Illusion of Conscious Will*, 27 BEHAV. & BRAIN SCI. 649 (2004). The *précis* is followed by open peer commentaries and a response from Professor Wegner.

> Sometimes how things seem is more important than what they are. This is true in theater, in art, in used car sales, in economics, and—it now turns out—in the scientific analysis of conscious will. The fact is, it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. Although it is sobering and ultimately accurate to call all this an illusion, it is a mistake to conclude that the illusory is trivial. On the contrary, the illusions piled atop apparent mental causation are the building blocks of human psychology and social life. It is only with the feeling of conscious will that we can begin to solve the problems of knowing who we are as individuals, of discerning what we can and cannot do, and of judging ourselves morally right or wrong for what we have done.[52]

Alternatively, to take another example from Greene and Cohen, suppose that "neuroscience holds the promise of turning the black box of the mind into a *transparent bottleneck*."[53] They mean that the brain is the final mechanistic pathway through which all types of explanations of behavior must ultimately operate and that neuroscience will be able to demonstrate that brain mechanisms, not mental states, are doing all the work.[54] They speculate that we may someday possess "extremely high-resolution scanners that can simultaneously track the neural activity and connectivity of every neuron in the human brain" and, that with the help of computers and software, can help people see the neural events that are alone causally responsible for their behavior.[55] If such mechanistic understanding and knowledge were available and widespread, Greene and Cohen are probably correct that notions of responsibility would wither away because most would believe that it was the brain that "did it," not the agent, and we do not hold brains morally responsible.

This picture of human activity exerts a strong pull on the

---

52. WEGNER, *supra* note 51, at 341–42. In more recent work, Professor Wegner appears to have softened the radical interpretation of his claim, which is that we, as persons, as agents, are not really "controllers" whose mental processes cause action. Daniel M. Wegner, *Who is the Controller of Controlled Processes?*, *in* THE NEW UNCONSCIOUS 19, 32 (Ran R. Hassin et al. eds., 2005) ("This theory is mute on whether thought does cause action."). On the other hand, Professor Wegner seems ambivalent and loathes fully giving up the radical interpretation. See *id.* at 27 (arguing that the "experience of conscious will is normally a construction" and referring to mental causation as "apparent"). This apparent ambivalence is present in the work of others.

53. Greene & Cohen, *supra* note 42, at 1781.

54. *Id*.

55. *Id*.

popular, educated imagination as well as on the theorizing of scientists. Consider the following example. In an ingenious recent study,[56] 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. Activity in these regions predicted immediately subsequent purchases "above and beyond self-report variables."[57] As we shall see in the next sub-section, this does not mean that the person's weighing of preferences and prices and the final decision played no role. Activity in the nucleus accumbens, the insula, and the mesial prefrontal cortex is not "weighing" and "deciding." The latter are the activities of people, not brains.[58] The findings interestingly, although unsurprisingly, suggest, however, that specific brain regions play a crucial role in particular types of psychological processes.

This study was reported in the Science Times section of the *New York Times* by John Tierney.[59] Here is how the story was "spun," beginning with its title: *Findings: The Voices in My Head Say "Buy It!" Why Argue?* The shopper is simply the hapless puppet of brain processes and plays no role as an agent in the purchase process. The decision is not up to the shopper; it is up to his or her brain. The conclusion considers how the study might help us deal with feckless consumerism.

> You might remove the pleasure of shopping by somehow dulling the brain's dopamine receptors so that not even the new Apple iPhone would get a rise in the nucleus accumbens, but try getting anyone to stay on that medication. Better the occasional jolt of pain. Charge it to the insula.[60]

In addition to getting the study wrong—insula activation was associated with excessive prices and the decision not to

56. Brian Knutson et al., *Neural Predictors of Purchases*, 53 NEURON 147 (2007).

57. *Id*. at 147.

58. M.R. Bennett & P.M.S. Hacker, *Philosophical Foundations of Neuroscience: An Excerpt from Chapter 3*, *in* NEUROSCIENCE & PHILOSOPHY: BRAIN, MIND & LANGUAGE 15, 18–23 (Maxwell Bennett et al. eds., 2007) (describing ascription of psychological attributes to the brain as "senseless"). *But see*, Daniel Dennett, *Philosophy as Naive Anthropology: Comment on Bennett and Hacker*, *in* NEUROSCIENCE & PHILOSOPHY: BRAIN, MIND & LANGUAGE, *supra* at 73, 86–88 (claiming that it makes sense to attribute "attenuated" sorts of psychological attributes to parts of the brain).

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

60. *Id*.

purchase[61]—it betrays 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.

If accounts such as these from both scientists and the media are correct and their implications were properly understood, rationality would require either that we abandon agency-based conceptions and practices of responsibility or that we learn to live with the illusion that we are agents. The rich explanatory apparatus of intentionality is simply a post-hoc rationalization we hapless homo sapiens construct to explain what our brains have already done. We are just mechanisms, although the illusion of conscious will may play a positive role in our lives.[62] Let us call this the "No Action Thesis" (NAT).

If these doubts about folk psychology and agency are accurate, compatibilism cannot save responsibility because determinism is consistent with either of two inconsistent views of human behavior. The truth of determinism is consistent with the existence or non-existence of agency, with the causal role or non-causal role of mental states in explaining behavior. Responsibility depends on agency, on the causal role of mental states, and the new discoveries arguably deny the possibility of agency as it is traditionally conceived.

Before turning to the actual evidence for NAT, let us consider some conceptual difficulties, using Greene and Cohen's "transparent bottleneck" argument. I will assume that the scanning and computing abilities that the argument employs are possible, although the brain has 10^{11} cells and at least 10^{15} connections.[63] The real problem with the argument is not that

61. *Id*.

62. This claim should not be confused with the apparently similar claim that "personhood" is an illusion. *See* Martha J. Farah & Andrea S. Heberlein, *Personhood and Neuroscience: Naturalizing or Nihilating?*, 7 AM. J. BIOETHICS 37, 40 (2007) (claiming that our construct of "personhood" is simply the illusory product of innate and automatic brain systems that is "projected" onto the world). There are many problems with the logic of this claim, but even if it is correct, it does not deny that creatures like us have mental states, such as desires and beliefs, that can be causally explanatory. Most charitably interpreted, it simply denies the explanatory usefulness of the normative concept of a "person."

63. Is it really likely, however, that the computer would predict what precise sentences we would speak? At present, of course, the speculation is pure science fiction and, in my opinion, is likely to remain so.

it assumes a (barely) plausible computational ability, but that it appears to assume the validity of a complete reduction of mind to mental states at the level of (apparently) neural networks. Such reductivism is controversial, however, even among monists who believe that the brain produces the mind, which is realizable in the brain.[64] Indeed, the complete post-Enlightenment project of reducing all phenomena to the most basic physical building blocks is also controversial and almost certainly a chimera. Until we have the science to demonstrate that such reduction is possible and that it is the best explanation of mental states, there is no reason to foresee the end of responsibility.

It is possible, of course, that our ability to predict behavior using new neuroscience techniques may become so successful that we will abandon responsibility concepts and practices not because responsibility is impossible, but because the consequential attractions of the potential for social engineering are so great. In this sense, the practical use of responsibility practices may hang by a technological thread.[65] But this is distinguishable from abandoning responsibility because we are not agents. After all, much of our behavior much of the time is predictable—such as being polite at professional meetings—but this does not mean we are not responsible for such predictable behavior. It is also possible that when we do discover how the brain enables the mind (assuming that this is possible) it will so profoundly alter our understanding of ourselves as biological creatures that all moral and political notions will change.[66] Nevertheless, this argument is different from claiming that we are not agents, that our mental states do no explanatory work.

B. THE EVIDENCE FOR THE "NO ACTION THESIS"

The real NAT question is whether scientific and clinical

64. *See* Carl F. Craver, *Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience*, 36 STUD. HIST. & PHIL. BIOLOGICAL & BIOMEDICAL SCI. 373, 375 (2005) (claiming that reduction models have "shortcomings" and that non-reductive physicalism is now a "standard view in the philosophy of mind").

65. Stephen J. Morse, *Neither Desert Nor Disease*, 5 LEGAL THEORY 265, 294–303 (1999).

66. *See* CARL F. CRAVER, EXPLAINING THE BRAIN: MECHANISMS AND THE MOSAIC UNITY OF NEUROSCIENCE 1 (2007) (suggesting that if mysteries such as consciousness are solved, "[I]t will revise our self-conception as radically as . . . Darwin's humbling of our origins.").

investigations have shown that action 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 seem to have their biological substrate in specific regions of the brain. None of these types of evidence offers logical support to NAT, however.

Just because a person may not be aware of all the causes for why he formed an intention does not mean that he did not form an intention, that he was not a fully conscious agent when he did so, and that his intention played no causal role in explaining the person's behavior. Even if human beings were never aware of the causes of their intentions to act and of their actions, it would not necessarily follow that they were not acting consciously, intentionally and for reasons that make eminent sense to anyone under the circumstances.

Human consciousness can undeniably be divided or diminished by a wide variety of normal and abnormal causes.[67] We have known this long before contemporary scientific discoveries of what causes such states and how they correlate with brain structure and processes. Law and morality agree that if an agent's capacity for consciousness is non-culpably diminished, responsibility is likewise diminished. Some suggest that it is diminished because bodily movements in the absence of fully integrated consciousness are not "actions."[68] Others believe that apparently goal-directed behavior that is responsive to the environment, such as sleepwalking, is action,

67. *See* JEFFREY L. CUMMINGS & MICHAEL S. MEGA, NEUROPSYCHIATRY AND BEHAVIORAL NEUROSCIENCE 333–43 (2003) (description of dissociative and related states and their causes and treatments); D. Vaitl, et al., *Psychobiology of altered states of consciousness*, 131 PSYCHOL. BULL. 98 (2005).

68. *See, e.g.*, MICHAEL S. MOORE, ACT AND CRIME 49–52, 135–155, 257–58 (1993) (arguing that cases of compromised consciousness should be treated as non-action); *see also* MICHAEL S. MOORE, *More on Act and Crime*, 142 U. PA. L. REV. 1749, 1804–20 (1994).

but that it should be excused because diminished consciousness reduces the capacity for rationality.[69] Let us assume that the former view is correct, because it offers more direct support to NAT and therefore the greatest challenge to traditional notions of individual responsibility. Let us also assume that divided or diminished consciousness is more common than it appears to be. Nevertheless, neither of these assumptions supports the more radical, general NAT thesis.

Demonstrating that divided or partial consciousness is more common than it appears certainly extends the range of cases in which people are not responsible or have diminished responsibility. Such studies do not demonstrate, however, that most human bodily movements that appear intentional and rational (apparently rational actions) occur when the person has altered consciousness.[70] One cannot generalize to all human behavior from genuinely deviant cases or cases in which a known abnormality is present. A model of action (or, we should say, non-action) built on sleepwalking, for example, is hardly a threat to orthodox notions of individual responsibility.

There is substantial empirical evidence to suggest that laboratory manipulations of unsuspecting subjects can cause the subjects to believe that their intentions were producing action when this was not the case.[71] That subjects can be cleverly misled by experimental manipulations hardly indicates that intentions generally play no role in explaining our behavior. Self-deception under laboratory conditions of deceit does not entail that intentions generally do not causally explain action.

Finally, there is accumulating evidence that various

---

69. Stephen J. Morse, *Culpability and Control*, 142 U. PA. L. REV. 1587, 1641–52 (1994) (arguing that clouded consciousness should be treated as an affirmative defense); *see also* Bernard Williams, *The Actus Reus of Dr. Caligari*, 142 U. PA. L. REV. 1661 (1994) (arguing that human activity with clouded consciousness is action).

70. *Accord* John F. Kihlstrom, *The Automaticity Juggernaut—or, Are We Automatons After All?*, *in* ARE WE FREE? PSYCHOLOGY AND FREE WILL 155-173 (John Baer, James C. Kaufman & Roy F. Baumeister eds. 2008) (reviewing the literature of and the explanation for the claim that virtually all behavior is automatic and concluding that the experimental literature on automatic behavior does not support such a sweeping assertion).

71. *See* John A. Bargh, *Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior*, *in* THE NEW UNCONSCIOUS, *supra* note 52, at 37, 51–54 (2005) (reviewing the evidence and concluding that the "will" is not primarily responsible for action).

psychological processes have their biological substrates in localized regions of the brain. We have long known that many behavioral activities are biologically based in highly specific regions. For example, the ability to recognize faces is highly localized in a region of the temporal lobe of the right hemisphere referred to as the "fusiform face area." Should this area become lesioned, the subject loses the ability to recognize faces, a condition called prosopagnosia.[72] Now, however, functional neuroimaging techniques permit the exploration of brain activity during more complicated psychological processes and can identify biological substrates for the processes. I have already discussed the example of brain regions associated with decisions to purchase an object.[73] For another example, a recent study demonstrated that investigators could determine from the region of brain activity which mental process—adding or subtracting—a subject had covertly intended to, but had not yet, performed.[74]

The localization evidence is immensely interesting and suggestive, but it does not indicate that mental states play no role in causally explaining behavior. There must be a biological substrate in the brain for all human behavior. If your brain is dead, you are dead and not behaving at all. Nor is it surprising that particular regions of the brain are associated with particular psychological processes. For example, a leading, albeit controversial, theory of how the mind works suggests that it is composed of different systems that perform different functions.[75] Although we do not know how the brain enables

72. James W. Tanaka, *Object Categorization, Expertise, and Neural Plasticity*, *in* THE COGNITIVE NEUROSCIENCES III 877, 883 (Michael S. Gazzaniga ed., 3d ed. 2004).

73. *See supra* notes 56–58 and accompanying text.

74. John-Dylan Haynes et al., *Reading Hidden Intentions in the Human Brain*, 17 CURRENT BIOLOGY 323, 323–28 (2007). It is important to recognize that the brain activity accurately predicted only which *type* of process the subject had covertly formed the intention to perform. It did not identify the specific content of the intention, such as which two numbers the subject intended to add or subtract. Despite the enormous advances in cognitive neuroscience, we do not know how to read minds using neuroimaging or any other technique. *Cf.*, Martha J. Farah, *Bioethical Issues in the Cognitive Neurosciences*, *in* THE COGNITIVE NEUROSCIENCES III, *supra* note 72, at 1309, 1309–10 (referring to the ability to identify traits and states as "a crude form of mindreading").

75. *See, e.g.*, JERRY A. FODOR, THE MODULARITY OF MIND (1983) (providing a strict modular theory).

the mind, it makes sense to assume that specific psychological processes would have brain substrates specific to each individual process. Based on what we already know about localization and on the reasonable assumption that it would be inefficient if all regions of the brain needed equal activation to support all psychological processes, localization is most likely to be true. Even if all this is correct, however, it does not follow that mental states do no causal explanatory work. It demonstrates at most that the neural network substrates for specific mental functions may be located in specific regions of the brain.

What is needed to support NAT is a general and direct demonstration that causal intentionality is an illusion *tout court*, but no such general demonstration has yet been produced by scientific study. The most interesting evidence has arisen from studies done by neuroscientist, Benjamin Libet,[76] which have generated an immense amount of comment.[77] Indeed, many claim that Libet's work is the first direct neurophysiological evidence of NAT.[78] Libet's exceptionally creative and careful studies demonstrate that measurable electrical brain activity associated with intentional actions 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,[79] the basic scientific methodological validity of these studies.[80] The crucial question then becomes whether the interpretation of these

76. Benjamin Libet, *Do We Have Free Will*, *in* THE VOLITIONAL BRAIN: TOWARDS A NEUROSCIENCE OF FREE WILL 47 (Benjamin Libet et al. eds., 1999) (summarizing the findings and speculating about their implications).

77. WEGNER, *supra* note 51, at 54–55 (characterizing the recounting of Libet's results as a "cottage industry" and noting the large and contentious body of commentary).

78. William P. Banks & Susan Pockett, *Benjamin Libet's Work on the Neuroscience of Free Will*, *in* THE BLACKWELL COMPANION TO CONSCIOUSNESS 657, 658 (Max Velmans & Susan Schneider eds., 2007).

79. *See, e.g.*, HENRIK WALTER, NEUROPHILOSOPHY OF FREE WILL: FROM LIBERTARIAN ILLUSIONS TO A CONCEPT OF NATURAL AUTONOMY 250–252 (Cynthia Klor trans., 2001); Jing Zhu, *Reclaiming Volition: An Alternative Interpretation of Libet's Experiment*, J. CONSCIOUSNESS STUD., Nov. 2003, at 61, 61–77.

80. Banks & Pockett, *supra* note 78, at 659–662 (concluding after a careful review of possible artifacts that "[R]eadiness potentials do start before the subject consciously 'decides' to move.").

findings as supporting NAT is valid. It does not follow from this temporal ordering that conscious intentionality does no causal work. It simply demonstrates that non-conscious brain events precede conscious experience. This seems precisely what one would expect of the mind-brain. Electrical impulses move quickly among neurons, but some lag between brain activity and conscious experience seems unsurprising. Once again, if the brain is dead, the person is dead. Prior electrical activity does not mean that intentionality played no causal role. Electrical activity in the brain is precisely that: electrical activity in the brain and not a mental state such as a decision or an intention. A readiness potential is not a decision.[81]

Moreover, Libet does not carefully distinguish between urges or wants on the one hand and decisions and intentions on the other.[82] Indeed, Alfred Mele argues that the experimental evidence is much more consistent with the readiness potential being associated with an urge rather than with an intention or a decision. [83] A perfectly plausible reading of Libet's work is that various non-conscious causal variables, including non-conscious urges, precede action—who would have thought otherwise?—but intentionality is nonetheless necessary for action.

Libet also suggests that people can "veto" the act during the delay between becoming aware of the intention and performing the intended action, which he surprisingly conceives of as an undetermined act. Other researchers appear to have localized the part of the brain that is the substrate for this activity of vetoing.[84] But, in addition to the implausibility of the veto being undetermined,[85] the conceptual foundations of the interpretation that the subjects were exercising a genuine veto are shaky at best.[86] This suggestion undermines the claim that the brain is doing all the work because it is an agent's

81. *See* ALFRED R. MELE, FREE WILL AND LUCK 30–46 (2006).

82. *Id.* at 33; *see also* M.R. BENNETT & P.M.S. HACKER, PHILOSOPHICAL FOUNDATIONS OF NEUROSCIENCE 228–31 (2003) (criticizing Libet's account of action).

83. MELE, *supra* note 81, at 33, 40.

84. *See* Marcel Brass & Patrick Haggard, *To Do or Not to Do: The Neural Signature of Self-Control*, 27 J. NEUROSCIENCE 9141, 9144 (2007) (identifying the part of the brain that is activated when the "veto" is exercised).

85. Banks & Pockett, *supra* note 78, at 667.

86. MELE, *supra* note 81, at 34–35.

mental state, a newly formed intention to veto, that causes the agent not to perform the act. In short, Libet's work presupposes agency at every step in the process.

Finally, Libet's task involved "random" finger movements that involved no deliberation whatsoever and no rational motivation for the specific movements involved.[87] This is a far cry from the behavioral concerns of the criminal law or morality, which address intentional conduct in contexts when there is always good reason to refrain from harming another or to act beneficently. In fact, it is at present an open question whether Libet's paradigm is representative of intentional actions in general because Libet used such trivial behavior.[88]

Libet's work is fascinating, but it does not prove that humans are generally not conscious, intentional agents or capable of employing their conscious intentionality when they have good reason to do so.[89] Even if the work is methodologically valid, various conceptual and interpretive arguments undermine the claim that Libet has demonstrated that NAT is true.

In short, despite the often astonishing findings and impressive advances in neuroscience and allied disciplines, there is no compelling evidence yet that NAT is generally true. Future discoveries may undermine this conclusion, however, so in the next subsection I turn to the implications of NAT.

C. IMPLICATIONS OF NAT

NAT provides no guidance about what people should do next and, in any event, degenerates into self-referential incoherence. Suppose that you were convinced by the mechanistic view that you were not an intentional, rational agent after all. (Of course, the notion of being "convinced" would be an illusion, too.[90] Being convinced means that you

87. Participating in the study and cooperating with the investigator can be rationally motivated, of course. But the experimental task was to move one's finger randomly, for no good reason.

88. Banks & Pockett, *supra* note 78, at 662–63.

89. *See* Jerry Fodor, *Making the Connection*, TIMES LITERARY SUPPLEMENT, May 17, 2002, at 4 (arguing that the new neuroscience rarely has much to contribute when the phenomenon in question is complex social behavior).

90. *See* Daniel C. Dennett, *Calling in the Cartesian Loans*, 27 BEHAV. BRAIN SCI. 661, 661 (2004) (wondering, in response to Professor Wegner, who is this "we" that inhabits the brain).

were persuaded by evidence or argument, but a mechanism is not persuaded by anything. It is simply neurophysically transformed.) What should you do now? You know that it is an illusion to think that your deliberations and intentions have any causal efficacy in the world. (Again, what does it mean according to the purely mechanistic view to "know" something? But enough.) You also know, however, that you experience sensations such as pleasure and pain and that you care about what happens to you and to the world. You cannot just sit quietly and wait for your neurons to fire. You cannot wait for determinism to happen. You must, and will of course, deliberate and act.

If one still thought that NAT were correct and that standard notions of genuine moral responsibility and desert are therefore impossible, one might nevertheless continue to believe that the law would not necessarily have to give up the concept of incentives. Through poorly-understood automatic processes, it is possible that various potential rewards and punishments would shape behavior even if they did not do so as premises in practical reasoning. Such an account would be consistent with "black box" accounts of economic incentives. 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 thought and agents, not a property of brains—play in this "decision"? Once again, the NAT account seems to swallow itself. Moreover, NAT proponents of consequentialism could hardly complain about those who refuse to "accept" what the proponents think rationality requires. The allegedly misguided people who resist are simply the victims of their automatic brain states. They cannot be expected intentionally to use their capacity for reason to accept what the consequentialists believe reason demands. Indeed, the consequentialist's belief is also an illusory mental state or it exists but plays no role in explaining behavior.

Even if our mental states play no genuinely causal role (about which, once again, we will never be certain until we solve the mind-body problem) human beings will find it almost impossible not to treat themselves as rational, intentional agents unless there are major changes in the way our brains work. Moreover, if one uses the truth of pure mechanism as a premise in deciding what to do, this premise yields no particular moral, legal or political conclusions. It will provide no guide to how one should live or how one should respond to the truth of NAT.

D. REASONS TO REJECT NAT

Answers to the possibility of NAT are rooted in common sense, a plausible theory of mind, 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. As the eminent philosopher of mind, Jerry Fodor, has written:

> [I]f commonsense intentional psychology were really to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're that wrong about the mind, then that's the wrongest we've ever been about anything. The collapse of the supernatural, for example, didn't compare. . . . Nothing except, perhaps, our commonsense physics . . . comes as near our cognitive core as intentional explanation does. We'll be in deep, deep trouble if we have to give it up. . . .
>
> . . . But be of good cheer; everything is going to be all right.[91]

Moreover, the folk-psychological theory has much explanatory power and is capable of scientific investigation.[92] Finally, it is hard to imagine the nature of a scientific study that would prove conclusively to creatures that have created

---

91. JERRY A. FODOR, PSYCHOSEMANTICS: THE PROBLEM OF MEANING IN THE PHILOSOPHY OF MIND, at xii (1987).

92. *See, e.g.*, BERTRAM F. MALLE, HOW THE MIND EXPLAINS BEHAVIOR: FOLK EXPLANATIONS, MEANING AND SOCIAL INTERACTION (2004) (providing a full theoretical account and empirical support). There is also growing recognition within psychology that "mental-state inference is one of the most fundamental tools of social cognition." Bertram F. Malle, *Folk Theory of Mind: Conceptual Foundations of Human Social Cognition*, *in* THE NEW UNCONSCIOUS, *supra* note 52, at 225, 229.

that study and will assess it with mental states that mental states do no work.

The plausible theory of mind that might support such explanations is thoroughly material, but non-reductive and non-dualist. It hypothesizes that all mental and behavioral activity is the causal product of lawful 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.[93]

Moreover, there is a perfectly 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. Practical reason is inescapable for creatures like ourselves who inevitably care about the ends they pursue and about what reason they have to act in one way rather than another.[94] Because we are social 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.[95] As a profoundly social species, it seems apparent that our ancestors would have been much less successful, and therefore much less likely to be our ancestors, if they were unable to understand the intentions of others, not sure they could convert their intentions into action, and were not also equipped with powerful assumptions that that stranger coming over the hill is equipped with the same capacity for harmful intentions as they are.[96] The ubiquitous

93. *See, e.g.*, Searle, *supra* note 6, at 113–14 (terming his position "biological naturalism" about consciousness).

94. BOK, *supra* note 31, at 75–91, 129–31, 146–51 (1998).

95. LARRY ALEXANDER & EMILY SHERWIN, THE RULE OF RULES: MORALITY, RULES AND THE DILEMMAS OF LAW 11–25 (2001) (explaining why rules are necessary in a complex society and contrasting their account with H.L.A. Hart's theory).

96. *See* Justin N. Wood et al., *The Perception of Rational, Goal-Directed Action in Nonhuman Primates*, 317 SCIENCE 1402, 1405 (2007) (demonstrating that the ability to understand the intentions of other creatures evolved in primates 40 million years ago); *see also* Esther Herrmann et al., *Humans Have Developed Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis*, 317 SCIENCE 1360 (2007) (comparing chimpanzees and orangutans to two-and-a-half-year-old humans and discovering that they have approximately equal cognitive skills concerning the physical world, but that

and centrality of mental states suggests that they are very evolutionarily expensive if they play no causal role in peoples' lives.

Human beings have developed extraordinarily diverse ways of living together, but a ubiquitousness feature of all societies is that they are governed by rules addressed to beings capable of following those rules. As Fodor notes, one of the most basic, well-justified assumptions about human nature is that we are consciously intentional creatures that are capable of a great deal of rationality. At the very least, we remain entitled to presume that conscious intentions are causal and to place the burden of persuasion at a very high level on those who wish to substitute another account.

In sum, the allegedly disappearing person is fully visible, well, and continues to act for good reasons, including the reasons not yet to accept NAT.

V. CONCLUSION

There is an avalanche of new neuroscience in both scientific journals and the popular media. The legal and social implications of this work seem very troublesome. As a special report on neuroscience in *The Economist* warned, "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."[97] In a more recent editorial, the same newspaper direly reported that "modern neuroscience is eroding the idea of free will."[98] The editorial argued that free will is necessary for responsibility and that "science will shrink the space in which free will can operate by slowly exposing the mechanism of decision making."[99]

For the reasons I have given, I believe that these warnings are not conceptually and empirically justified, but they clearly represent a particular view that the new neuroscience engenders. Because agency and responsibility are so central to our interpersonal and moral lives, so central to our conception of ourselves, and so tied to notions of dignity and autonomy, I

---

humans have superior cognitive skills for understanding social interaction).

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

98. *Free to Choose?*, *supra* note 4, at 16.

99. *Id.*

hope that we will always have reason to reject the view that we are not agents and responsibility is impossible. At present, however, we are justified in believing that we are agents and can be responsible.