ELIMINATIVE MATERIALISM, NEUROSCIENCE AND THE CRIMINAL LAW

ANDREW E. LELLING†

TABLE OF CONTENTS

INTRODUCTION .................................. 1472
I. FOLK PSYCHOLOGY AND THE CRIMINAL LAW .......... 1477
   An Attempt to Avoid Folk Psychology .................. 1480
II. THE PROBLEM IN CONTEXT ........................ 1484
   A. Sigmund Freud and B.F. Skinner ................... 1485
   B. Biology, Behavior, and Blame ...................... 1489
III. ELIMINATIVE MATERIALISM .......................... 1492
   A. What Happened to Dualism ......................... 1493
   B. The Reality of Neuroscience ....................... 1495
   C. The "Content Problem" ............................. 1497
      1. Propositional Attitudes ......................... 1498
      2. Sententialism .................................... 1503
   D. The Argument from History ......................... 1506
   E. The Argument from Progress ......................... 1507
   F. Can Folk Psychology Accommodate Materialism? ...... 1508
      1. Identity Theories .............................. 1509
      2. Functionalism .................................. 1511
   G. An Eliminative Replacement ......................... 1516
   H. An Illustration: Benjamin Libet
      and the Delay of Conscious Intention ............... 1520

† B.A. 1991, State University of New York at Binghamton; J.D. Candidate 1994, University of Pennsylvania. I dedicate this Comment to my father, Dr. Irwin J. Lelling, who gave me the blocks with which I build.

I am also indebted to more immediate supporters: This project would not have been possible without the patient guidance of Professor Michael S. Moore and the editorial wisdom of G. Michael Harvey. The arguments presented here have been improved by other hands, namely: James Bramson, Ari Burstein, Amy Selena Fisher, Silvestre Fontes, Marianne Hoffman, Leo Katz, Joel Mick, Stephen Morse, Lani Anne Remick, Noah Silverman, and the Associate Editors of the Law Review.

(1471)
IV. ELIMINATIVE MATERIALISM, NEUROSCIENCE AND THE CRIMINAL LAW .............................................. 1526
   A. Eliminative Materialism and Criminal Responsibility .... 1527
      1. Fictionalizing Intentionality .............................. 1527
      2. Redescribing the Law .................................... 1530
         a. The Character Theory of Responsibility ............ 1531
         b. Economics-Based Theories of Behavior .............. 1535
         c. Utilitarianism and Avoidance of Mental States ........ 1536
   B. Specific Legal Accommodations of Libet and Dennett .... 1539
      1. Eroding the Intentional ................................ 1542
      2. Rethinking Our Defenses ............................... 1544
         a. Provocation ......................................... 1544
         b. Force of Habit ...................................... 1545
   C. The Ultimate Implication of Eliminativism: Our Selves in Doubt .......................... 1549
      Making Sense of Multiple Selves .......................... 1551
         a. Multiple Personality Disorder and the Fragmented Self .............. 1552
         b. Collective Responsibility .......................... 1557

V. BETWEEN HERE AND THERE: POSSIBLE OUTCOMES AND THE LAW'S ATTITUDE TOWARD NEUROSCIENTIFIC CHANGE .... 1560
   A. We Hit a Dead End ........................................ 1561
   B. We Answer the Question ................................ 1562
   C. The Legal Attitude Toward Gradual Neuroscientific Change .......................... 1562

CONCLUSION ........................................ 1564

INTRODUCTION

Any system of laws that purports to assign blame on the basis of moral responsibility inherently relies upon some view of psychology. Laws generally reflect the moral sensibilities of a community, condemning those acts considered evil or damaging to others. In turn, that assessment of immorality partially springs from an underlying understanding of psychology. For example, those who do not "intend" the immoral results of their acts are less morally culpable than those who do "intend" them. An individual is morally deserving of social retribution only if she is responsible for her illegal actions, and the assessment of her responsibility is made by
referring to the accepted notions of cognition and behavior in her society.¹

The final layer, underlying a society's notions of cognition and behavior, is empirical science. Legal mechanics, if moral, reflect the view of psychology deemed most accurate by progressive brain research, since otherwise there would be misapplications of blame—the law would be guided by a morality based on a faulty view of behavior, leading to punishment of persons not necessarily deserving of retribution. The veracity of our current psychological views, however, is actually a monumental assumption; the psychological theory motivating our assignments of blame is not an outgrowth of sober research, but the culmination of thousands of years of unquestioned supposition. What if the rapid progress of modern cognitive science² were to veer away from the psychological theory upon which we base our morality and our law? If great leaps in scientific research are connected, in chain-link fashion, to shifts in morality and then legality, is society willing to go wherever that research leads us? These questions comprise the backdrop for the dilemmas contemplated in this Comment.

A defined psychological stance is certainly a prerequisite to accepting the precepts of our criminal law. Initially we assume that most individuals are rational, practically reasoning creatures, and this assumption is the cornerstone of our definition of legal personhood. If we did not take this step, it would be impossible to judge individuals against any kind of behavioral standard, since such a standard would be undefinable.³ Based on an assumption of

---

¹ Professor Joshua Dressler notes this relationship in his discussion of the insanity defense. See JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 296 (1987).

² Cognitive science is an umbrella term for disciplines that seek to understand the human brain. Rebecca Dresser has summed up its components as:

[N]euroscience—neurophysiology, neurology, neuropsychology, neurobiology—[where] the problem is approached from the "bottom up," by studying how the brain's microstructure and microactivity affect behavior[; and] cognitive psychology and artificial intelligence, [where] researchers examine the mind from the "top down," by considering what is already known about human mental activity and behavior, and then exploring what internal processes could possibly explain these phenomena.


³ See DAVID HODGSON, THE MIND MATTERS 184-85 (1991) ("[W]e base our actions to some extent on our beliefs about the beliefs, desires, intentions, etc. of ourselves and other people."). An assumption of rationality is also necessary for behavioral predictions, since "without rationality any behaviour is compatible with any beliefs and desires." Jerry A. Fodor, Fodor's Guide to Mental Representation, in THE FUTURE 1993]
rationality we establish, via procedures such as trials, the level of culpability of the person accused of performing a socially unacceptable act. We will, for example, punish her severely if she “intentionally” performed the action, and less so if she “negligently” performed it.4

This system of assessment is not specific to the criminal law. It is merely a legal manifestation of a deeply ingrained view of psychology. We regularly explain the actions of ourselves and others by reference to “a network of principles which constitutes a sort of common-sense theory about how to explain human behavior,”5 and this network constitutes our main “vehicle of . . . interpersonal commerce.”6 We seem to have beliefs and desires motivating our actions, and we judge the actions of others by referring to them. If we know X to be a mostly rational7 agent, that is, a creature whose behavior is systematically caused by, and explainable in terms of, her beliefs, desires and related propositional attitudes,8 and X decides to walk to the grocery store, we intuitively explain her behavior in terms of those concepts. We assert that X is probably hungry, or in extended format, that (1) X desires food (for whatever motivating reason); (2) believes that food is available at the grocery store; and so (3) intends to walk to the grocery store. We further assume that (4) X willed her body to travel to the store (as opposed to being animated by some external force), and so that is why we see her walking toward it.9

4 See MODEL PENAL CODE § 2.02 (Official Draft & Revised Comments 1985).
7 Upon reflection, few would assert that we are perfectly rational creatures. Some would assert that we are not mostly rational creatures. See, e.g., Fodor, supra note 3, at 26-27.
8 A “propositional attitude” is simply an attitude toward certain propositions. For example, “I fear that the sky is falling.” Fear is the attitude I have toward the proposition expressed. The term “propositional attitude” is a convenient way of referring to the numerous mental states we attribute to human actors. The concept is explored further in Part III.C.1.
9 We regularly employ thousands of generalizations based on our traditional understanding of mental states. For example, most would agree that if a person eats, she is hungry; if a person is tired, she cannot concentrate; if a person wants something, and believes she has the capacity to get it, and has no conflicting desires or concerns, she will attempt to get that thing. Numerous “rules” such as these allow us to make sense of our experiences. For further examples, see PAUL M. CHURCH-
It may be apparent by now that our common-sense explanation of behavior makes several assumptions. Within the assumption of rationality, we make the related assumption that there is such an object as a "belief," "desire," or "intention." There is a definite implication to our mind-view that humans undergo "mental events" such as these, and that these events have content: When we state that X believes there is food in the grocery store, we are implying that there is some objectively verifiable element or group of elements inside her skull that roughly corresponds to the proposition "there is food in grocery stores." Finally, we also assume that there is a relationship between these mental states and bodily movements, and that the former somehow initiate the latter. Apparently, our theory of human minds and behaviors not only makes many assumptions, but presumes a bit of knowledge about neurophysiology.

This presumption is interesting, considering that the theory has been in use for thousands of years.\(^{10}\) There is no definable, watershed date in history when human beings decided to attribute rationality to themselves and judge behavior in light of contentful mental states. The system simply evolved as a foundation of our social constructs. There is no evidence for its scientific reality, since it was firmly established before humans possessed any knowledge of cerebral processes, but this system of explaining behavior does seem to make a remarkable amount of sense. All of us adhere to its tenets in our day-to-day lives, and its use is so ingrained that we rarely contemplate its mechanics.

The above description is a brief sketch of what has been labeled our "folk psychology."\(^{11}\) Although entrenched, some have begun to question folk psychology's scientific currency, noting on one hand its mostly unquestioned dominance since antiquity, and pointing to startling advances in neuroscience\(^{12}\) with the other.

---

\(^{10}\) See, e.g., ARISTOTLE, Ethica Nichomachea, in INTRODUCTION TO ARISTOTLE 308, 478 (Richard McKeon ed., 1947) (describing voluntary acts in traditional psychological terms).

\(^{11}\) See, e.g., John D. Greenwood, Folk Psychology and Scientific Psychology, in THE FUTURE OF FOLK PSYCHOLOGY, supra note 3, at 1, 5 (defining folk psychology as "a body of causal-explanatory theoretical references to contentful psychological states employed by layfolk and scientific psychologists"). "Folk psychology" seems a vaguely disparaging phrase, since it implies a lack of theoretical sophistication on the part of its advocates. It is employed here, with apologies to its proponents, for sake of simplicity.

\(^{12}\) See, e.g., NEWSWEEK, Apr. 20, 1992 (special issue on "Mind and Brain"); SCI.
Scientific and philosophical restlessness has borne a new school of thought called *eliminative materialism*, which denies the accuracy of our most elemental psychological assumptions, insisting that the "common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced . . . by completed neuroscience."13 These thinkers are "materialists" because they conceive of the mind as a biological manifestation within the brain, and they are "eliminativists" because they seek an "elimination" of those psychological phenomena that rely on the existence of beliefs, desires, and other "mental states." As neurophysiological explanations of human behavior and cognition expand, time-worn psychological notions are in danger of being supplanted by a radically different, but empirically verifiable scientific substitute. This attack on folk psychology is also an assault on all institutions that have come to rely on its mechanics, including the criminal law.14

This Comment explores the claims and challenges of eliminativist materialism, and their potentially devastating effect on a legal system that relies on folk-psychological notions. Part I further reveals the criminal law's reliance on common-sense notions of behavior, while Part II briefly notes two brain/mind theories that have threatened folk psychology in the past and then looks at the relationship between biological explanations of behavior and the criminal law. Part III explores eliminative materialist claims about the falsity of mental state psychology. That section also assesses folk-psychological accommodations of modern philosophical/scientific advances and presents an eliminative materialist model of the brain. It concludes with an illustration of the eliminativist attack on folk psychology through the findings of Dr. Benjamin Libet, whose conclusions about the underlying neurophysiological components of "voluntary acts" may render the modern legal view of intentional action severely outdated. Part IV presents potential legal recharacterizations and compromises in response to Libet and

13 Churchland, *supra* note 6, at 67.  
14 Psychology influences many areas of the law, but legal reliance on psychological theory is most obvious in the criminal law, which usually depends on the "mental state" of an individual to determine her level of culpability for the act committed. *See Model Penal Code § 2.02.*
eliminativism generally. Finally, Part V considers the ramifications of accepting or rejecting an eliminative materialist future.

I. FOLK PSYCHOLOGY AND THE CRIMINAL LAW

It should not be surprising to learn that the criminal law is beholden to folk-psychological concepts. Indeed, one might assert that it could not realistically be based on anything else.\textsuperscript{15} The criminal law's assumptions about the human mind are so fundamental that one need not look far to uncover glaring examples.

The mens rea requirement of the prima facie criminal case is the obvious starting point. Generally, a person is not guilty of an offense "unless [s]he acted purposely, knowingly, recklessly or negligently, as the law may require."\textsuperscript{16} Under the Model Penal Code, a person acts "purposely" when it is her "conscious object," or intention, to engage in conduct of a particular nature or to cause a certain result.\textsuperscript{17} She acts "knowingly" if she is "aware" of the nature or consequences of her action,\textsuperscript{18} and she is "reckless" if she "consciously disregards" substantial and unjustifiable risks.\textsuperscript{19} Finally, she is "negligent" if she creates substantial risks through deviation from the standard of care that a reasonable person would observe,\textsuperscript{20} meaning a person who has the "normal capacities, physical and mental, for doing what the law requires."\textsuperscript{21} We punish people for negligent conduct in hopes of deterring such carelessness in the population at large—people will be careful because they believe they will be punished for damage negligently caused.\textsuperscript{22} These various assessments of mental state greatly influence the severity of punishment meted out to a particular actor.\textsuperscript{23} For example, homicide constitutes murder when performed purposely, knowingly, or so recklessly as to manifest

\textsuperscript{15} But see infra text accompanying notes 37-52 (discussing Posner's assertion that mental state psychology is not a required component of the criminal law).

\textsuperscript{16} MODEL PENAL CODE § 2.02(1).

\textsuperscript{17} See id. § 2.02(2)(a)(i).

\textsuperscript{18} See id. § 2.02(2)(b)(i).

\textsuperscript{19} See id. § 2.02(2)(c).

\textsuperscript{20} See id. § 2.02(2)(d).

\textsuperscript{21} MICHAEL S. MOORE, LAW & PSYCHIATRY 83 (1984) (quoting HERBERT L. HART, PUNISHMENT & RESPONSIBILITY 152 (1968)).

\textsuperscript{22} See MOORE, supra note 21, at 82.

\textsuperscript{23} See, e.g., GA. CODE ANN. §§ 16-5-1 to -2 (Michie 1992) (designating the death penalty or life imprisonment for those convicted of murder, but only one to 20 years of imprisonment for those convicted of voluntary manslaughter).
"extreme indifference to human life." Further, a homicide that would normally be considered murder but is committed "under the influence of extreme mental or emotional disturbance" may be considered manslaughter, a lesser offense. The law is littered with the mentalistic terms of a folk-psychological framework, and those terms are a guide to culpability in most crimes.

The actus reus requirement also betrays a debt to folk psychology. A "voluntary act" is usually required to establish liability for a criminal action, and that very concept is a folk-psychological artifact. We commonly assert that bodily movement is achieved through the use of volitions—we "will" our bodies to move, and thus they move. These volitions are an integral component of the essential folk-psychological formula: mental states lead

24 Model Penal Code § 210.2(1)(a)-(b) (emphasis added).
25 Id. § 210.3 (emphasis added). Note that this qualification of the offense of homicide cements the law's reliance on folk psychology's concept of intent, just as mens rea reveals a reliance on beliefs, desires, and rationality. Planning an offense, as opposed to impulsively committing it or being provoked to commit it, entails an "ongoing intention" to commit that offense in the future, and so is more culpable. See Steve Walt, Some Problems of Pragmatic Jurisprudence, 70 Tex. L. Rev. 317, 344 (1991) (book review). This continuous intention constrains the formation of other intentions over time:

To accomplish my plan to pocket Jones's watch, I need to deliberate about how to do so. My plan enters into my deliberations as an input. The outcome of deliberation, an intention, is about selecting means to effectuate the plan. Intentions incompatible with pocketing the watch are excluded. If a selected means proves infeasible, my plan is frustrated. Alternative means effectuating my plan will be selected. Treating planning simply as a series of desires cannot account for the consistency constraint that plans place on the formation of intentions. Resort to intentions is needed.

Id. at 344-45.
26 Mens rea is not an issue in "absolute liability" offenses, but these violations are generally malum prohibitum, such as parking violations, as opposed to malum in se. See Model Penal Code § 2.05. But see United States v. Park, 421 U.S. 658, 674 (1975) (discussing the liability of corporate officers for corporate criminal wrongdoing of their subordinates of which the officers had no knowledge). Strict liability is narrowly employed in the criminal law, supported only by public welfare concerns. Most crimes of moral significance require an assessment of mental state, so that society can determine if the actor intended the evil she created.

27 See Model Penal Code § 2.01.
28 Id. § 2.01(1). This term might be considered redundant, since an act will always involve some kind of volition. See, e.g., Oliver W. Holmes, The Common Law 45-46 (1881) ("An act... imports intention... A spasm is not an act. The contraction of the muscles must be willed.").
29 There is generally no criminal liability for movements that are not actions. See Model Penal Code § 2.01(2). This distinction is "embedded... in our morality, our law, and our sense of what a self or a person is." Michael S. Moore, Act & Crime (forthcoming June 1993).
to intentions, which are executed through volitions.\textsuperscript{30} In our previous example, after X consulted her beliefs and desires, she developed an intention to walk to the grocery store, and this in turn became a volition, a willing of her body to move down the street. A folk psychologist might characterize the volition as a “mediating state,” between “our motivations and our intentions, on the one hand, and our actions, on the other.”\textsuperscript{31} The criminal law agrees with this conception, and does not impose punishment for thoughts, or “mental states,” alone. Thoughts must be accompanied by “voluntary acts” in order to incur liability.\textsuperscript{32}

The presence of folk-psychological presuppositions in elemental legal concepts such as mens rea and actus reus results in folk-psychological entanglements throughout other facets of the criminal law. The defenses of necessity\textsuperscript{33} and duress\textsuperscript{34} both require a folk-psychological viewpoint, and the insanity defense\textsuperscript{35} exists because of a legal assumption of rationality in the average actor. Our legal system has sprouted from an accepted picture of proper, or rational, conduct, and an individual excusable by insanity presumably does not see the picture or cannot adhere to its design.\textsuperscript{36}

\textsuperscript{30} MOORE, supra note 29.
\textsuperscript{31} Id.
\textsuperscript{32} See MODEL PENAL CODE § 2.01(1); see also DRESSLER, supra note 1, at 64 (“[T]he criminal law does not punish persons for thoughts, but only for actions that result from thoughts.”).
\textsuperscript{33} See MODEL PENAL CODE § 3.02(1) (“Conduct which the actor believes to be necessary to avoid a harm or evil to himself or to another is justifiable [in enumerated situations].”) (emphasis added). This defense thus assumes that there are cerebral occurrences called “beliefs” that motivate persons to act as they do.
\textsuperscript{34} See id. § 2.09(1) (allowing an affirmative defense when confronted with “the use of, or a threat to use, unlawful force . . . which a person of reasonable firmness . . . would have been unable to resist”) (emphasis added). The only reason to be “reasonably firm” in the face of lawful force is if one fears for her life or believes that harm will come to her if she does not perform a certain act.
\textsuperscript{35} See id. §§ 4.01-.03.
\textsuperscript{36} Tests for removal from criminal responsibility due to mental condition are controversial, and there is no universally accepted standard. See DRESSLER, supra note 1, at 289, 296-304; see also Michael S. Moore, Causation and the Excuses, 73 CAL. L. REV. 1091, 1109-10 (1985) (discussing the relationship between the insanity defense and a causal theory of excuse).
An Attempt to Avoid Folk Psychology

The above discussion presents the criminal law as a paradigm of folk psychology, but this claim has been partially disputed. The arguments of Judge Richard Posner are worth noting, since they attempt to demonstrate that the criminal law does not rely on folk-psychological descriptions. His tact is representative of economic analyses of the law, but Posner takes the added step of claiming to eliminate successfully mental state considerations from the modern criminal law. If true, his assertions would mitigate the threat of "completed neuroscience," since the fate of the law would now be somewhat divorced from the fate of common-sense psychology. Judge Posner argues that, at some level, the law understands that mental entities such as desires and intentions are of "dubious ontology," and that the legal system is "much less mentalistic than legal semantics imply." In so doing, he invokes the work of another distinguished jurist, Oliver Wendell Holmes, to support the notion that as civilizations and their legal systems mature, criminal responsibility becomes increasingly "external," that is, more a matter of conduct than of intent. The focal point for assessments of liability shifts from evaluations of mental state to evaluations of observable conduct, with criminals eventually likened to "unreasonably dangerous machines."

Posner is skeptical of the existence of mental states, or even the mind, and seeks to justify current legal mechanics without reliance on mentalistic explanations. He does acknowledge that beliefs, desires, and rationality—"the mental element"—are integral to the criminal law, but declares that the law does not require a "concept of mind in which intentions and free will figure." Posner equates, by economic analysis, what a person intends to do with what it is rational for her to do. Consequently, legal descriptions are altered from a mental/internal to a behavioral/

38 Id. at 167.
39 Id. at 168. Judge Posner refers to Holmes, supra note 28, citing the diminishing role of mental states in light of advancing scientific knowledge as a "major theme" of that work. Id.
41 See Posner, supra note 37, at 162.
42 Id. at 176.
43 See id. at 170.
external account of human action.\textsuperscript{44} X can only guess what Y intends, but she can probably deduce fairly accurately what it would be rational for Y to do in a given situation. Posner uses the example of seeking a college education:

\begin{quote}
[O]ne might begin by suggesting that some people have a “taste” for obtaining a college education but then show that this taste is instrumental to a more general goal; call it income maximizing. The propensity to attend college will now be seen as a function of the cost of college and of the effect of college in raising one’s lifetime income. Ideally, one could predict whether people would go to college even if one knew nothing about their thoughts on the subject; and then one might stop talking, in analytical work at least, about people “wanting” to go to college or “thinking about” going to college. People would still have desires and thoughts, but these would be strictly epiphenomenal.\textsuperscript{45}
\end{quote}

Posner is thereby able to remove consideration of mental states from an explanation of human behavior. Actions, including crimes, are performed to maximize resources, and are not the result of particular states of mind.

The cherished concept of free will gets similar treatment. Posner initially observes that “a person who acts in accordance with compelling reasons is normally thought to be acting freely.”\textsuperscript{46} Her deliberation, based on an existing set of beliefs, desires, and available information, progressively narrows the range of choices until only one remains. If enough information is available to eliminate uncertainties concerning consequences, the proper choice suddenly seems “predetermined,” since only one most-rational alternative can possibly exist.\textsuperscript{47} A subject’s beliefs and desires, combined with objective data about the situation, yield a most-rational path for her to follow. Free will need not become an issue. Instead of consciousness being the root of free will, it is now an agency enabling individuals to collect enough information to discover an already-determined “best choice.” As Posner puts it, “we have desires, and we have beliefs—formed with the help of consciousness—about how to fulfill those desires; and the conjunction of the desires and beliefs determines, without need to posit a

\textsuperscript{44} See id. at 171 & n.17.
\textsuperscript{45} Id. at 170-71 (footnote omitted).
\textsuperscript{46} Id. at 172.
\textsuperscript{47} See id.
faculty of free will, our volitional behavior."\(^{48}\) Conscious choice becomes cost-benefit analysis, as the potential criminal weighs the costs of arrest and incarceration against the potential benefits to be conferred by, for example, stolen merchandise. Her beliefs and desires, weighed against her information and knowledge of legal realities, are all that determine her conduct.

According to Posner, what progress the criminal law has made in guiding behavior has come from the steady replacement of folk-psychological notions with behavioral ones such as those above. The law is an "instrument of social control,"\(^ {49}\) a system of rewards and punishments that guides a rational, resource-conscious populace along the path of acceptable behavior. If the known costs of certain behaviors are high enough, those behaviors are deterred, without need to refer to mental motivations. If the law were truly designed with intentions and free will in mind, says Posner, then the allocation of criminal liability would not stand as it currently does. For example, under modern law a person who eagerly kills in self-defense is excused, while the person who would never have killed but for a harsh upbringing is punished. This distinction makes sense from a social perspective: punishment would be a useless deterrent in the first instance, but useful in the second, since "[s]ociety has no desire to license people who have had a bad upbringing to kill."\(^ {50}\) Similarly, the strict liability aspects of the law also support the argument that though rationality needs to be assumed, assessments of intent and free will are not crucial to our legal mechanics.\(^ {51}\) Posner sees two important justifications for avoiding intent and free will: first, we are unable to "peer into people's minds" to monitor activity there, and second, even if we could, we are not quite sure that we would find the mental entities posited by folk psychology.\(^ {52}\) Thus we are better off adopting the social/behavioral approach that the law currently follows in many areas, avoiding the necessity of building the law on shaky mentalistic grounds.

\(^{48}\) Id. at 173.
\(^{49}\) Id. at 176.
\(^{50}\) Id.
\(^{51}\) For example, an individual who has sexual intercourse with an underage female whom he believes to be of age is guilty of a crime, despite his intention to have intercourse with a consenting adult. Id. Also consider the felony-murder rule, in which an unintended killing during commission of a felony is treated as murder. See DRESSLER, supra note 1, at 463.
\(^{52}\) See POSNER, supra note 37, at 177.
Economic explanations like Posner's are ultimately not enough to remove the law from the neuroscientific threats that now face folk psychology. Initially, Posner concedes the necessity of rationality, beliefs, and desires, because without them actors could not be deterred by the threat of punishment. A potential criminal is only deterred because she believes she will be punished if she commits a crime, and she fears that eventuality. This alone would be enough to render the law vulnerable to the scientific revisionism proposed by the eliminative materialists. However, even the propositional attitude of intention is far more important to the nuts and bolts of the law than Posner asserts. Consider burglary: the only difference between that offense and criminal trespass might be the presence of intent in the former. The law of criminal attempt also relies upon intention, where an attempt involves a legally permissible action coupled with a culpable intention, and abandoning that intention operates as an affirmative defense to liability. The desire may remain, but the intention has been abandoned, and so the distinct element of intention is required.

The related concept of free will is also not so easily excised. While the Model Penal Code considers reflexes or convulsions to be "acts," they are not deemed "voluntary acts," revealing the need for some theory of volition above and beyond the economist's allowance of mental states such as beliefs and desires. For example, consider Thom, who is driving home from work and sees Len crossing the street in front of him. Thom happens to desire, more than anything in this world, to run over Len and kill him. Thom derives such excitement from savoring fantasies of Len's demise that
he begins to shake uncontrollably, sending his vehicle careening into the unsuspecting Len, who subsequently dies. Thom's beliefs and desires caused Len's death, but Thom did not murder Len. He will be charged with a lower level of culpability (perhaps manslaughter) than he would have been had he willfully maneuvered the car into the poor pedestrian (which would be murder). Legal responsibility is thus influenced by the volitions, or "free will" of the actor, and consideration of them cannot be removed from a description of our criminal law without altering that law.

Posner's theory presents a typical example of the inability to separate the criminal law from a deeply rooted web of psychological ingredients. Since the law does not easily escape an identification with the concepts of folk psychology, one must address the scientific threats—past, present, and future—to that venerable system of ordering the world in which we live.

II. THE PROBLEM IN CONTEXT

Legal practitioners and scientists do not think the same way. The former look backward and the latter look forward. While lawyers attempt to justify their assertions within a framework of previously decided caselaw, researchers seek to introduce new models in place of the old. There are obvious reasons for this difference. Certainty and stability are worthy goals for a legal system, since an unpredictable jurisprudence would have negative effects on society. In contrast, psychological and biological research presumably seeks to uncover a more complete picture of the human condition as it currently exists, and so would stagnate if constrained by the above considerations.

This philosophical divergence between law and science can generate tension. Immersed in history and drawing its strength from gradual elaboration, the law is content to change by preeceden-

---

57 This is an example of a "deviant causal chain," used by many theorists in discussions of action and responsibility. See, e.g., MOORE, supra note 21, at 67-68, 72.

58 The more established that case law, the better. This is the essential element of stare decisis. See BLACK'S LAW DICTIONARY 1406 (6th ed. 1991) (defining stare decisis as a "[d]octrine ... grounded on theory that security and certainty require that accepted and established legal principle ... be recognized and followed").

59 More sweeping gestures, such as major legislative reforms and constitutional amendments, effect more rapid alterations, but these events are rare. We are concerned here with the dynamics of the common law—the law encountered by judges, lawyers, and their clients every day.
tions. In our context, the law has contently stood by its folk-psychological notions while psychologists, neurophysiologists, and even philosophers have restlessly sought the empirical underpinnings of our cerebral processes. Perhaps a confrontation was just a matter of time.

A. Sigmund Freud and B.F. Skinner

There have been potential clashes in the past. Any thinker who has critiqued folk-psychological models and presented a sophisticated substitute has also impugned the criminal law. To note two prominent examples from this century, consider Sigmund Freud and B.F. Skinner. Freud, the founder of psychoanalysis and an enormously influential figure in modern psychology, agreed with Thomas Hobbes that human beings, when stripped of social conditioning, are "savage brutes" with the basest of instincts. The only way humans transcended their brutal origins was through a pattern of internal repressions of socially unacceptable urges that began in early childhood. Freud was able to explain various irrational behaviors in his patients by looking to "unconscious" conflicts between natural instincts and societal demands. These conflicts, usually originating in childhood and since repressed, would eventually reassert themselves "through new and devious channels," thereby causing odd or irrational behavior. The subject would be unaware of these submerged motivations unless elicited to consciousness by psychoanalysis.

Freud's conclusions imply that one should not be held responsible for actions caused by suppressed conflicts. The beliefs and desires motivating those actions are not accessible to the actor, whose strange, perhaps criminal behavior is now explainable in terms of unconscious forces. Unavoidable internal struggles have caused her questionable conduct, and therefore the subject must be released from any criminal responsibility, since she could not have exercised the free will necessary for the assignment of moral

---

60 See Henry Gleitman, Basic Psychology 290, 302-03 (3d ed. 1992). What follows is the briefest sketch of Freud's ideas.
61 See id. at 286. Hobbes and Freud differed on the timing of the subjugation of savage instincts in humanity. Hobbes envisioned a political-social contract agreed to in the distant past, while Freud looked to individual psychology. For Freud, each person in every generation had to come to terms with the strictures society placed on her savage instincts, and this was accomplished through self-induced mental repressions. See id. at 287.
62 Id. at 287.
culpability. One need not hypothesize about whether this twist of our folk psychology has actually affected the criminal law. Freud's theories of unconscious motivation have fueled many an insanity defense. For example, in the oft-cited case of Pollard v. United States, an apparently rational and accomplished police officer engaged in a rash of poorly planned bank robberies approximately two years after the brutal killing of his wife and daughter by a drunk neighbor. Though Pollard explained his motivations for robbery in completely rational terms, the three psychiatrists who examined him concluded that he had actually robbed the banks because of an unconscious desire for societal retribution. Pollard apparently felt responsible, on an unconscious level, for the deaths of his wife and daughter, since he was absent from home at the time of their murder. In light of the psychiatrists' testimony, the Sixth Circuit reversed a lower court conviction of Pollard and absolved him of responsibility for the robberies. Cases such as Pollard imply that Freud's findings fill a secure niche within the realm of criminal excuse.

B.F. Skinner, however, had far more radical things to say about the psychological status quo. Building on the earlier works of Ivan Pavlov and Edward Thorndike, Skinner did extensive research in

63 In light of Freud's enormous influence, his ideas are arguably an integral part of the modern folk-psychological view. Many of Freud's assertions have come under fire from cognitive and psychological theorists. See id. at 297-302.


65 282 F.2d 450, mandate clarified, 285 F.2d 81 (6th Cir. 1960).

66 Pollard, about to marry his second wife, stated that he had decided not to lead the "financially insecure life" that had characterized his first marriage, and so decided to rob a bank. Pollard, 282 F.2d at 463.

67 See id. at 454-55 n.2.

68 See id.


70 See Pollard, 282 F.2d at 464. For further commentary on Pollard, see Reisner & Slobogin, supra note 64, at 523-27. Moore also addresses the concept of unconscious mental states excuseing unacceptable behaviors and finds it problematic. See Moore, supra note 64, at 1641 ("If all conscious mental life is determined by unconscious mental states, as many psychoanalysts believe, why is everyone not excused for all of his actions, seemingly the product of his conscious decisions but in fact [] determined by his unconscious mental states?"). See generally Moore, supra note 36, at 1112-48 (critiquing causalist theories of excuse).

71 See GLEITMAN, supra note 60, at 72-73, 77-79. Pavlov was a major contributor to the study of classical conditioning, a method of building associations between
instrumental conditioning, becoming one of the best known advocates of behaviorism.\textsuperscript{72} Behaviorists look only to observable physical activity for an explanation of human behavior, maintaining that folk-psychological talk of "mental states" is just so much unverifiable conjecture.\textsuperscript{73} Skinner and other behaviorists sought to explain all of human behavior as an elaborate causal exchange between organisms and the environment; behavior was dictated solely by the response—reward or punishment—that the organism could expect from his surroundings. Skinner also took issue with the entire realm of cognitive theory, asserting that behaviorism could account for all the phenomena that folk psychology purports to explain, and do so more simply at that.\textsuperscript{74} Pointing to cognitive theorists' unfounded suppositions and general lack of verifiable brain data, Skinner sought to return behaviorism to the mainstream after it was overshadowed by cognitive explanations of human action.\textsuperscript{75}

Previously neutral stimuli and desired responses. Pavlov's well-known dogs would salivate when given food, and they would also salivate when given food accompanied by the sound of a buzzer (a neutral stimulus). After enough trials Pavlov found that the dogs would salivate in response to the buzzer alone. Thus, the dogs had associated a response with a previously meaningless stimulus. See id. at 72.

Thorndike's contribution was to instrumental conditioning, which took the classical model a step further. The subject was conditioned to perform an instrumental response—the subject learned to give the response in order to enjoy a certain effect. Thorndike placed animals in "puzzle boxes," in which the animal was required to perform a simple task in order to be freed, at which point the creature would be rewarded with a morsel of food. The trials were repeated until the animal had mastered, or "learned," the task. See id. at 78.

The difference between classical and instrumental conditioning is important to note:

In classical conditioning, the animal's behavior is \textit{elicited} by [a previously neutral stimulus] . . . [I]n instrumental conditioning the organism is much less at the mercy of the external situation. Its reactions are \textit{emitted} from within, as if they were what we ordinarily call \textit{voluntary}. . . . [The reactions] operate on the environment to bring about some change that leads to reward.

\textit{Id.} at 79. Instrumental conditioning can also be "aversive"—a common example is spanking a child for unacceptable conduct, thereby discouraging that behavior with the threat of punishment. See id. at 85. This rationale is easily applied to the incarceration of criminals, which is presumably for deterrent as well as retributive purposes.

\textsuperscript{72} Behaviorism may be "the single most influential school of psychology in the English-speaking world." CHURCHLAND, \textit{supra} note 9, at 88.

\textsuperscript{73} See id.


\textsuperscript{75} See id. at 300.
By removing mental states and the peculiarly human trait of consciousness from consideration, Skinner dehumanized psychology, and if his theories had come to dominate popular thinking the criminal law would have suffered profound alterations. Despite Skinner's assertions, however, it is by no means clear that cognitive theories are unworkable, and the current scientific focus betrays a cognitive approach. Skinner actually elucidated a philosophy as much as a scientific theory—most of his proposals were not wrong per se, but merely ignored the cognitive path that researchers later chose to follow. Behaviorism was the "hard science" approach Skinner felt would legitimize the field of psychology, but it flourished by the good graces of neurophysiology's infancy—scientists simply did not know enough about internal brain functions effectively to replace purely behavioral explanations of behavior.

Neither Freud nor Skinner caused devastating alterations to our legal system, although Freud's findings have had some impact. The ideas of these two men are presented as examples of the potentially large gap that can exist between progressive psychology and the legal thinking that is presumably based on an accurate view of that discipline. Since Freud's views have been assailed on various grounds, and Skinner's views have been neglected in favor of cognitive interpretations, it is obvious why the law has had little difficulty with them. But another reason exists for the shallow effect on legal concepts: neither psychological view is legitimized by an intimate knowledge of human biology. Freud presumed no new knowledge of cerebral processes, and Skinner's thesis was

---

76 See Churchland, supra note 9, at 89.
77 See generally B.F. Skinner, Beyond Freedom and Dignity (1971). Skinner completely redefines personal freedom in behaviorist terms, asserting, for example, that "[w]anting is not . . . a feeling, nor is a feeling the reason a person acts to get what he wants. . . . Freedom is a matter of contingencies of reinforcement, not of the feelings the contingencies generate." Id. at 37-38. In short, human action for Skinner is determined by the environment, and not beliefs or desires generated sua sponte. Under this rubric, attributions of moral responsibility are impossible, since all behavior is caused by external forces, and no free will is involved.
78 Skinner's behaviorism does encounter some problems, however, that make it a less powerful and less credible explanation of behavior than its cognitive competitor. For example, it is extremely difficult to explain the phenomenon of speech generation as a result of environmental factors. A person may construct any one of hundreds of verbal responses to a given situation or to no situation at all. It is hard to envision how that behavior could be the direct result of, and only the result of, environmental stimuli external to the actor's brain.
79 See Churchland, supra note 9, at 88.
completely based on observations of external behavior, ignoring both subjective assessments of mental state and analysis of internal brain physiology. This is an important distinction when one considers that the law is prone to following what is called the "medical" model, a view that puts much stock in verifiably biological explanations of human behavior. This relationship between law and biology is further explored below.

B. Biology, Behavior, and Blame

In a very general sense, there are several ways of envisioning human behavior. Besides the "medical," or biological, model, one might adopt a psychological model which considers one's personality to be a product of "patterns established during youth through interactions with one's family or family substitutes."81 There is also the social model, which relies upon larger social groups such as friends or schools as the major influences on later behavior,82 and finally the behavioral model just described, which ignores mental phenomena and looks only to specific reinforcing or aversive elements in the subject's environment.83

There are strong reasons to link the criminal law to a biological approach, which places the origins of motivations and behaviors within the human body. For one, the biological view pre-dates the other models and so forms the foundations of the legal view of personhood.84 But the biological conception is also the only one with a completely "endogenous premise".85 Behavior is attributed to elements inside the body, and so assigning blame to individuals is not problematic. If the law relies on social forces or other elements external to the actor as causes of her behavior, the notion of personal moral responsibility becomes far less intuitive: one can more credibly claim that the locus of blame should be whatever external operant(s) "forced" the actor's hand, instead of the actor

80 See Reisner & Slobogin, supra note 64, at 327 (quoting Aaron Lazare, Hidden Conceptual Models in Clinical Psychiatry, 288 NEW ENG. J. MED. 345, 346 (1973)). Reisner and Slobogin describe the medical model as one in which "one's mental state results primarily from organic or chemical conditions within the body." Id. They employ these models as the basis for discussing mental disorders, but their comments apply equally well to general legal conceptions of behavior.

81 Id.

82 See id. at 328.

83 See id. at 327.

84 See id. at 328.

85 Id.
herself.\textsuperscript{86} Finally, the biological view is the only one that is somewhat verifiable. Instead of being based solely on observation or interpretation, it is able to point to chemicals, hormones, and bodily processes as the “answers” to our behavioral questions. It possesses an air of legitimacy arising from its reliance on empirical physiological elements. Legal willingness to believe biological stories as opposed to psychodynamic, behavioral, or social ones seems to reflect a broader societal acceptance of medical explanations. To most of us, those explanations appear plausible and far more credible than other, more exotic interpretations.\textsuperscript{87}

The law is thus more likely to entertain excuses based upon physiological evidence, and the courts continue to hear ingenious arguments from defense attorneys making use of the latest advances in genetics and brain science.\textsuperscript{88} New information fashions poten-

\textsuperscript{86} Reisner and Slobogin note: “Endorsement of the behavioral or social models . . . could broaden enormously the scope of defenses based on lack of criminal responsibility, since these models suggest that exogenous factors, such as rewards for aggressive behavior or the impact of poverty and social class, [account for motivations and behavior].” \textit{Id.}

The above concerns arise from a strong determinist philosophy, which would excuse behaviors attributable to external forces. Those holding a “compatibilist” position on the question of causation and personal responsibility, however, do not have a problem reconciling personal responsibility with attributing behavior to external causes. A compatibilist would readily admit that some behavior might be caused by forces external to the actor, but that that external causation does not by itself remove personal responsibility for the behavior. Determinism and compatibilism are explained further \textit{infra} note 95.

\textsuperscript{87} Note that courts have not encountered defenses based on behaviorism, for example. The use of biological explanations is an important element in the power of eliminative materialism. \textit{See supra} text accompanying note 13.


The law currently recognizes insanity defenses, defenses based on diminished capacity, and the provocation defense to murder, among others. See Dressler, supra note 1, at 299-304, 325-27, 487. These have been accepted for some time, however, and all of them rely on cognitive or volitional deficiencies in the defendant.

This direct causal relationship is the major stumbling block for biology-based defenses. It is difficult to prove that a physiological condition was not just an influence, but some kind of element forcing the subject to perform the particular crime. It is a purely causal approach. See infra note 95.
III. ELIMINATIVE MATERIALISM

"There is nothing more horrible than the murder of a beautiful theory by a brutal gang of facts."91

Eliminative materialism presents just such a combination of revisionism and hard science. Eliminativists attack the most basic presumptions of common-sense psychology, declaring them an elaborate fiction waiting to be destroyed by a completed neuroscientific explanation of human behavior. This view also represents an important intersection of philosophical and scientific evolution. It is the point at which philosophers have encountered a science finally sophisticated enough to support or reject their theories of the mind, and it relies on neurobiological research to back radical philosophical claims. To appreciate the legal ramifications of eliminative materialism, it is necessary to understand its origins and its relation to competing theories that would preserve mental-state psychology. Cognitive theory has swung from dualism, the view that the mind and the brain are separate entities, to materialism, the view that the “mind” is simply the product of complex brain processes, just in time to derive support from neurophysiological research that legitimizes the latter conception. Unfortunately, the most recent child of this philosophical-biological marriage is an unruly one. Eliminative materialists not only agree with other materialists that explanations of cognition must derive from the physical sciences; they insist that an accurate view of those sciences entails the total elimination of folk-psychological constructs. Where Freud posited unconscious reasons for conscious beliefs and desires,92 where Skinner sought to shift our focus from beliefs and desires to behaviors,93 and where neurophysiology provided potential causal excuses for beliefs, desires, and intentions,94 eliminativists assert that there are simply no such things as beliefs, desires, and intentions, or any other mental states, and that we all operate under profoundly incorrect views of human cognition and behavior.95

92 See supra notes 60-70 and accompanying text.
93 See supra notes 71-80 and accompanying text.
94 See supra notes 81-90 and accompanying text.
95 Note, especially from the viewpoint of legal responsibility, that while eliminative materialism does imply determinism, the assertion that all behaviors are caused by biological and/or environmental factors, it does not require hard determinism. A
A. What Happened to Dualism

Dualism, roughly the conception that the mind is a distinct and nonphysical entity in relation to the body, plays a major role in our common psychological parlance,96 if not in progressive cognitive theory. We speak loosely of someone's "mind" as if it were an object separate from any physical (or physiological) moorings.97 If pressed, however, most of us would not insist that there is a completely nonphysical mind hovering somewhere about the physical brain. As science begins to explain the deeper secrets of hard determinist would assert that since all of our actions are caused by biological and/or environmental forces that are not under our control, no one justifiably can be held responsible for actions her body performs, since these actions are not voluntary in any real sense. The hard determinist philosophy is integral to the biology-based criminal excuses we noted earlier—since the defendant's actions are explainable as a series of biological causes, she should be free from responsibility for them, since biological mechanics are not "freely" chosen or controlled. Denno supports a form of "degree determinism," in which varying amounts of free will or determinism exist in relation to any one action. See Denno, supra note 88, at 660, 664. By an assessment of environmental and biological factors influencing the actor to perform a criminal act, a place can be established for her on a determinism continuum: if she falls sufficiently high on the scale, her behavior is excused. Moore has roundly criticized this view. See Moore, supra note 36, at 1115-16 ("It makes sense to say that we are determined or that we are free, but to speak of being partly determined or partly free makes as much sense as to speak of being partly pregnant.") (footnote omitted). Moore himself adopts a compatibilist position, agreeing readily that "determinism is simply true," in that all actions have biological or environmental origins and causes, id. at 1113, but insisting that individuals are still personally responsible for their actions. See MOORE, supra note 21, at 360-61. For example, if A mentions something to B that prompts B to tell an extremely tasteless joke, A (an environmental factor) has caused B to tell the joke, but B is still responsible for telling it.

Eliminative materialists can also adopt this compatibilist stance, thus avoiding hard determinism. While a hard determinist would maintain that all of our mental states are physiologically caused, an eliminativist would assert that there just are not any mental states at all. Eliminativism may eventually become a hard determinist attack on legal responsibility, but that will depend on where neuroscience takes us; it is not necessitated by the eliminative philosophy, which admits only that cognition and behavior are rooted in neurophysiological processes.

96 René Descartes is the oft-mentioned culprit behind modern dualism. Convinced that our conscious reasoning abilities could not be explained by reference to physical matter, he asserted that the activity known as "thinking" involved a nonspatial material coexisting with the physical body. See RENÉ DES CARTES, Meditation II: Of the Nature of the Human Mind; and that it is more easily known than the Body, in A DISCOURSE ON METHOD AND SELECTED WRITINGS 93, 93-103 (John Veitch trans., E.P. Dutton & Co. 1951) (1641). For a detailed description of Descartes's theory of anatomy, see RENÉ DESCARTES, A Discourse on Method, in id., at 1, 35-51.

97 One can find this tendency in various cliches. To wit: "Are you out of your mind?"
nature, it becomes increasingly difficult to attribute a phenomenon to some ethereal, inscrutable ingredient and leave the matter at that.98 Many of us would willingly accept some sort of physical explanation for consciousness and most philosophers have rejected dualism because of its difficulties.99 Apart from relying on an unverifiable premise, dualism encounters certain logical problems, such as in explanations of physical movement. How does that ethereal mental force cause your physical actions? Positing mental causation of physical movement violates the established principle of conservation of energy.100 Since energy is neither created nor destroyed, but only transferred between forms, a mental state cannot initiate a physical movement, since that would entail the creation of physical energy by the nonphysical mind for use by the physical body.

Dualism still has supporters,101 but their ranks are thinning. The concept seems to have officially lapsed into obsolescence after Gilbert Ryle's devastating characterization of Descartes's dualism as the "dogma of the Ghost in the Machine."102

---

98 Behaviorism embodied a certain degree of modern frustration with dualist explanations, and so relied only on observable phenomena. See Dresser, supra note 2, at 31.

99 See, e.g., MARIO BUNGE, THE MIND-BODY PROBLEM: A PSYCHOLOGICAL APPROACH 5 (1991) ("[D]ualism is not scientifically viable. Hence it is unacceptable to a science-oriented philosophy."); RODNEY COTTERILL, NO GHOST IN THE MACHINE 3 (1989) ("[M]y text goes against dualism and what I feel is too simplistic a view of the will."); DENNETT, supra note 40, at 33, 105-06 (explaining why dualism is in decline); David K. Lewis, An Argument for the Identity Theory, 63 J. PHIL. 17 passim (1966) (positing a theory that maintains folk-psychological concepts within a materialist framework).

100 See Kevin W. Saunders, Voluntary Acts and the Criminal Law: Justifying Culpability Based on the Existence of Volition, 49 U. PITT. L. REV. 443, 461 (1988) (presenting the conservation of energy criticism); see also BUNGE, supra note 99, at 16-21 (presenting a cogent review of the major criticisms of dualism); CHURCHLAND, supra note 9, at 9-10; DENNETT, supra note 40, at 35-36 (analogizing the logical incoherencies of dualism to those of "Casper the Friendly Ghost").


B. The Reality of Neuroscience

When it was stated above that most people, if challenged, would not defend a purely dualist conception of cognition and behavior, it was not revealed what view they would support. Our society's prevailing attitude is actually a materialist one. The most common modern conception of the brain is of a complex organ containing the central processing unit for all of our emotions, mental states, sensory inputs, thoughts, and volitions. That unit is the location in which "it all comes together," for sorting, storage, or action.

There is no such location. The above view is what the philosopher Daniel Dennett calls "Cartesian materialism." Descartes believed that there was a "pineal gland" straddling the two lobes of the brain, acting as the point of exchange between the physical brain and the nonphysical mind. In order for an input, command or emotion to reach "consciousness," it had to pass through this gland. The pineal gland does not exist, but Dennett invokes Descartes to describe the widely held view that there must be some spot in all that grey matter acting as a coordinating force. This notion is intuitively powerful. Consider the example of eyesight: light from the environment travels through the lens of the eye, into the retina, up the optic nerve, to . . . where? To you? To some Central Viewer inside your brain that is peering outward? Dennett points out that if your optic nerves are severed, you are unable to see, even though your eyeballs are intact. Notice the

---

103 DENNETT, supra note 40, at 107.
104 Id.
105 See id. at 34.
106 See id. at 105-06.
107 See id. at 107.
108 Those knowledgeable in brain anatomy might respond that the input travels from retinal ganglion cells to neurons in the lateral geniculate nucleus, and from there it is mapped onto the primary visual cortex. See, e.g., Carla J. Shatz, The Developing Brain, SCI. AM., Sept. 1992, at 61, 62. This statement is true, but it does not account for consciousness. Where is the "seeing" occurring? Where are images created, and where are they "projected" for your conscious involvement?

Dennett provides numerous thought experiments to highlight our cognitive illusions. For example, he asks the reader to imagine, in detail, a purple cow, and then asks several questions about the cow's physical features and the direction it is currently facing. DENNETT, supra note 40, at 27. Considering that there are no purple, or cow-shaped, neurons in your skull, it seems as if there must be some sort of distinct "viewer" of your imaginings, since you "see" the cow. The nature of brain events, however, makes it obvious that there is no "witness" watching the goings-on, just as nothing ever "witnesses" the events in your liver or stomach. Id. at 28-29.
109 See DENNETT, supra note 40, at 108.
"you" in that sentence. One would be hard-pressed to find a recognizable "you" in all the "neural machinery" of your brain. There is no internal "screen" on which images are projected, so where, exactly, does "seeing" occur?\(^{111}\)

The bald truth is that if you analyze a human brain, you do not find any central processor,\(^{112}\) just as you cannot find recognizable "beliefs," "desires," or "intentions."\(^{113}\) What you do find are 100 billion interlocking neurons,\(^{114}\) and no specific area for the storage of those contentful mental states. In fact, PET scans show that our very ideas are not held in specific sections of the brain but are fragmented and dispersed among several areas.\(^{116}\) The process which draws these elements together into a cohesive "picture" is unknown.\(^{117}\)

One need only skim the most recent literature in neurophysiology to glimpse the startling and sweeping changes taking place in that science. Our knowledge of aging,\(^{118}\) sensory input,\(^{119}\) learning,\(^{116}\) and the nature of consciousness have changed dramatically in recent years.\(^{107,13,108,11,14}\)

\(^{110}\) Id.

\(^{111}\) Dennett employs the powerful metaphor of the "Cartesian Theater" to elucidate the absurdity of our common views. The Theater is a central "stage" in the brain, across which all of your sensory inputs, etc., parade for "your" viewing pleasure. Anything not appearing onstage is not consciously perceived. See id. at 39, 107-13. For a technical account of the visual system and its implications for consciousness, see Francis Crick & Christof Koch, \textit{The Problem of Consciousness}, Sci. Am., Sept. 1992, at 153, 153-59.

\(^{112}\) Researchers have pinpointed regions of the brain that seem to specialize in certain tasks. For example, the hippocampus is necessary for memory retrieval. See Sharon Begley et al., \textit{Mapping the Brain}, NEWSWEEK, Apr. 20, 1992, at 68, 68. But this same research has debunked the notion of a central work station in the brain for sensory input, language, learning, and other tasks. See id. at 69-70 (describing research demonstrating that neither hearing, visual learning, nor language skills have specific locations within the brain).

\(^{113}\) See infra part III.C.


\(^{115}\) "PET" stands for positron emission tomography. See Begley et al., supra note 112, at 68. These tests monitor brain activity by measuring the use of radioactive glucose by areas of the brain. See id.

\(^{116}\) This point can actually be used to support a folk-psychological representation of mind. See infra note 193 and accompanying text.

\(^{117}\) Begley et al., supra note 112, at 66. It has been hypothesized that there are "convergence regions" where data is collected and re-assembled from simultaneous firings by groups of spatially distributed neurons. See Antonio R. Damasio & Hanna Damasio, \textit{Brain and Language}, Sci. Am., Sept. 1992, at 88, 91.


\(^{119}\) See, e.g., Semir Zeki, \textit{The Visual Image in Mind and Brain}, Sci. Am., Sept. 1992,
ing, 120 and language 121 is constantly being revised in light of new discoveries. The salient point for our purposes is that these alterations draw scientists progressively farther away from traditional conceptions. 122 Launched from a materialist presumption, the expanding neuroscientific picture is fueling the eliminative materialists' endeavors to redefine psychology with neurophysiology, thus altering the traditional view of cognition and the social constructions that depend on it. 123

C. The "Content Problem"

The crucial eliminativist claim arising from the philosophical and neuroscientific developments above is for the absolute nonexistence of beliefs, desires, and other mental states. This assertion is what distinguishes eliminativist philosophy from its mainstream philosophical cousins. Its underlying argument focuses on what can be called the "content problem," or the difficulty of explaining how the mental states of folk psychology can possibly

at 69, 69-76 (describing the history of research in the area of visual sensory systems and new information gained from studies of visually impaired subjects).


121 See Damasio & Damasio, supra note 117, at 89.

122 An excellent example of this is the fascinating research being done in variable intelligence. Scientists now believe that intelligence may be a matter of "neural efficiency." In researching the process of "neural pruning," in which neuron use in thought processes steadily decreases from age five to the early teens, neurophysiologists have speculated that intelligence may simply be a matter of cerebral streamlining—more intelligent people become so because of greater neural pruning in their formative years. The more intelligent brain may know what brain areas not to use, thereby increasing efficiency of thought. See Begley et al., supra note 112, at 68-69. The seeming correlation between genius and madness can now be accounted for by "hyperpruning"—pruning so great as to increase intelligence enormously, while at the same time affecting basic cognitive systems. See id. at 69. This model is not even vaguely related to any traditional conception of intelligence.

Researchers have also discovered qualitative differences between the cerebral processes of men and women. See Doreen Kimura, Sex Differences in the Brain, Sci. AM., Sept. 1992, at 119, 119-25. If further gender or racially based neurological differences are discovered, fundamental equal protection doctrines could be discredited. See, e.g., THE DECLARATION OF INDEPENDENCE para. 1 (U.S. 1776) ("We hold these truths to be self-evident, that all men are created equal . . . .").

123 See, e.g., David Gelman et al., Is the Mind an Illusion?, NEWSWEEK, Apr. 20, 1992 at 71, 71 ("[T]he advent of new scanning technology, with the promise of actually 'seeing' thoughts in action, has prompted a fresh examination of mind theories.").
contain content or meaning—how a human brain can internally possess or represent the proposition “dogs have fur” or “my coat was expensive.” Many philosophers maintain that the concept of contentful mental states is logically incoherent; the point is not that neuroscience won’t vindicate beliefs and desires, but the stronger assertion that it can’t, because of inherent difficulties with the concept itself. This argument is a key element in revealing the precarious stance of folk psychology (and thus its legal manifestations), and so is considered at some length.

The content problem was once only a philosopher’s dilemma, but the recent emphasis on cognitive science has made it a psychologist’s dilemma as well. Attempts to shift the common perception of beliefs and desires from a philosophical puzzle to an empirical theory of the brain has placed more strain on that traditional view than it was meant to bear.124 If contentful mental states are to play a role in any credible scientific explanation of the human brain, certain difficulties, tolerated by philosophers for the sake of argument, must be resolved by psychologists and physiologists who adhere to the stricter standards of science. If problems persist as cognitive science evolves, the criminal law will eventually be faced with the dilemma as well, since folk psychology has a strong voice in the dispensation of blame by contemporary law.

1. Propositional Attitudes

A “belief” is more accurately referred to as “the proposition believed,” like the belief that birds can fly.125 The same can be said of other states, such as desires and fears: we fear that [there will be a nuclear war], or we desire that [class will be cancelled]. All of these states are referred to as “propositional attitudes,” because each expresses a distinct attitude toward a specific proposition.126 The problem with relying on believed propositions as the foundation of modern psychology and legal assessments of culpability is that it is not at all obvious that propositions exist as currently conceived. We commonly view them as “logically distinct, observer-independent, language neutral morsels of fact (or fiction),” as “atoms” used in the construction of our psychological endeav-

125 Id. at 119.
126 See Churchland, supra note 9, at 63.
ors. Under present day assumptions, beliefs are, as the term atom implies, seemingly inscrutable, elemental objects, which remain completely unanalyzed in psychological circles. A bit of reflection reveals that folk psychology's claims concerning the nature of propositional attitudes themselves and their relationship to human psychological states are not as easily defined and defended as that theory needs them to be.

Concerning the assumption that propositional attitudes like beliefs and desires are actual psychological phenomena, consider the world of Bob the frog. As he sits on his lily pad, Bob's brain is receiving massive amounts of potentially important information through his sensory receptors. You approach Bob with a gun. When Bob sees you, he turns to the right and hops frantically away. Now consider Bob's behaviors and the frog-brain states that initiated them. Did Bob see you approach and then want to hop away? Did he believe that you were approaching him directly, and then turn right and flee because he thought that that was the best avenue of escape?

The issue is whether Bob (or even other frogs) have mental states. As Dennett points out, it's not likely that frogs "believe that whales are mammals" or "could want a pizza." Most would say that frogs do not entertain propositions at all. This presents an interesting line-drawing problem. If frogs do not have mental states, do birds? Attribution of folk psychology to sparrows does seem counterintuitive. We might find attributing propositional attitudes to dogs a bit more plausible (dogs do appear to dream, for example, and if they react physically to a dream while asleep, what could they be reacting to if not a mentally conjured proposition?), but there would certainly be some debate. As we progress on the evolutionary scale, why should it suddenly be obvious that human beings enjoy and employ propositional attitudes? There is no apparent biological line to be crossed into the "charmed circle" of mental state psychology, and yet we seem to know that of course humans have beliefs, and many other animals, equally obviously, do not.

---

127 DENNETT, supra note 124, at 206.
128 See id.
129 This example is loosely adapted from the works of Dennett, who adapted it from Stephen Stich, an eliminative materialist. See id. at 106-10.
130 Id. at 107.
131 Id. at 110.
There are stronger points to be made about beliefs as actual psychological phenomena beyond the circumstantial argument above. Assume that Bob does have some kind of plausible mental state, such as the desire to consume food. There is no way to render the exact content of this desire. When he looks for flies, is he looking for flies as flies, for small, dark, flying objects emitting a certain sound, for objects evidencing a certain pattern of movement, or for something less specific or even totally different? No one knows, but it is useful and efficient to simply say that Bob really wants a fly. The attribution of mental state is a rough and ready characterization for an enormously complex neurophysiological system. If we could monitor every single neurological event in a live frog, we could predict when and why it sought a fly, and what it thought it was stalking. Since we cannot do that in frogs (or humans), we stick to propositional attitude labels, even though they are, at best, an extremely distorted description of neurophysiological processes and thus psychology.

One might protest the analogy to amphibians by asserting that humans are the only creatures employing a finely-tuned verbal language, and so are able to exactly describe, via use of sentences with lengthy strings of modifiers, exactly what they “believe” at a given moment. But there is no reason to suppose that language accurately relates the true nature of a brain state. When we are asked what we believe, we internally construct a sentence to be spoken moments later. While this is indisputably one of the best ways to determine what someone “thinks,” it is not “revealing any psychologically important underlying . . . beliefs, . . . [it is] just an artifact of the environmental demand for a particular sort of act [that is, production of a suitable speech fragment].” Like attributing obviously inaccurate psychology to Bob, we employ language as a more sophisticated label, but as a label just the same. We are required to transform complex internal states into verbal communication, and the result is only as accurate as verbal language allows. Indeed, there is nothing to show that human belief is different from frog belief:

In both cases behavior is controlled by a complex internal state that can be alluded to more or less effectively by the everyday folk practices of belief attribution and desire attribution. If in one case

132 See id. at 109-10.
133 See id. at 109.
134 Id. at 114.
it seems that the beliefs are much more finely and precisely individuated by their contents, this may be because we are attending not to the "individual beliefs" themselves (there may not be any such scheme of [neurophysiological] individuation), but to the products of linguistic behaviors controlled by those complex internal states, which products are *ipso facto* as distinct as the discriminations in that language allow.\(^{135}\)

There is no support for the strong assumption that attributions of belief, desire, etc. parallel actual psychological states. There exist no facts at all about what frogs or humans "really" believe,\(^{136}\) and so language (and folk psychology) become our best approximation.

Focusing on the fundamental nature of propositions themselves reveals why mental states simply *cannot* have the propositional content supposed by common-sense conceptions. Propositions possess certain characteristics, such as a "truth-value" (if X is true and Y is false, X and Y must be different propositions), and the essential nature of delineating some aspect of the world.\(^{137}\) But Hilary Putnam's well-known Twin Earth scenario\(^{138}\) reveals that psychological states cannot contain set propositions. In Putnam's thought experiment there is a parallel Earth that is almost an exact duplicate of our Earth, and on it exists a duplicate of myself and all other people. The only difference between our Earth and the Twin Earth is that "lakes, rivers, clouds, ... [and] living tissues ... contain not H\(_2\)O, but XYZ—something chemically different but indistinguishable in its normally observable macro-properties, from water, that is to say, H\(_2\)O."\(^{139}\) Inhabitants of Twin Earth call their clear, tasteless substance "water," as we do, but it is quite different from the substance we encounter, XYZ being chemically distinct. Since my replica and I are physically identical, we are also psychologically identical\(^{140}\) (we think the same about all things, like our opinions on water), and thus possess the same psychological states. But where my neurophysiological representations of belief are about "water," that is, H\(_2\)O, my replica's representations, though identical,

\(^{135}\) *Id.* at 115.

\(^{136}\) *See id.* at 116.

\(^{137}\) *See id.* at 124-26.


\(^{139}\) DENNETT, *supra* note 124, at 127.

\(^{140}\) This position is logically required if one operates from a materialist assumption. If there is no ethereal mind and two beings are biologically identical in every respect, then they must "believe" the same things.
are not about "water," but about XYZ. We "believe" completely different propositions—the belief I would express as "water is H₂O" is about water, and is true, while my replica's belief, which looks and sounds exactly the same, is not about water, but about XYZ, and is false (to me).

In explaining Putnam's example, a folk psychologist would say that the beliefs of my replica and myself are just different, that is, he believes water is XYZ and I believe it is H₂O. The problem with this simple explanation is that since the replica and myself are physical and psychological clones, our psychological (and neuro-physiological) states are exactly the same, so there is no internal differentiation to mark the "difference" in "belief." The individuation between XYZ and H₂O exists only as empirical phenomena in the external world, and not in the brain. The actual attributes of an object, be it water, frogs, or anything else, are real, constant, and distinct, but the psychological representations of them are indiscernible. That is, one cannot "find" distinct beliefs inside the human head. The folk psychologist assumes that differing beliefs link up to differing neurological representations, but Putnam's scenario implies that this is not the case. Thus, we are forced to conclude that psychological states cannot possess propositional content. The attributes of objects, and so the propositions that belong to them, do not correspond with psychological states, showing that propositions cannot be "grasped" and "held" internally.¹⁴¹ The proposition itself is independent of any psychology, or as Putnam expressed it, propositional meaning "ain't in the head."¹⁴²

A less exotic example might make the case more clearly: If over a five-year period I believe that New York City contains exactly eight million people, my psychological state concerning New York City's population will not change during that time. The supposed corresponding proposition, however, will change in truth-value many times during those five years, as the actual population of New York City fluctuates. Therefore, if my psychological state is one constant belief, it cannot be a "propositional" attitude, since it does not change as the truth-value of the proposition changes.¹⁴³ It would be implausible to claim that psychological, behavior-influencing states are dependent on "causally remote features"¹⁴⁴ such as

¹⁴¹ See DENNETT, supra note 124, at 127.
¹⁴² 2 PUTNAM, supra note 138, at 227 (emphasis omitted).
¹⁴³ See DENNETT, supra note 124, at 129 n.3.
¹⁴⁴ Id. at 129.
the migration or emigration of persons from a major metropolis. Those events are not linked to your cerebral processes. The way folk psychologists want to individuate mental states and objects is by how they refer to things in the world, but they cannot be in your head because they are in the world.

We normally think that the way to describe someone’s mental states is to describe that person’s propositional attitudes (beliefs and desires). If we want to build an accurate psychological picture, it seems we must be able to specify the correct and specific propositions believed or desired. That is no longer possible. If propositions cannot be attributed to internal attitudes, folk psychologists cannot assert that propositional attitudes, or “beliefs,” determine psychological states.

2. Sententialism

But it does seem as though some kind of information storing and ordering apparatus must be contained in our heads. Even if one cannot literally grasp a proposition that exists in the external world, a person must be grasping something concrete, some psychologically relevant, empirical structure in the brain. Despite a lack of verifiable beliefs, it seems there should be neuronal structures corresponding to what we can only, inadequately, label “thoughts.” Building on this intuition, many folk psychologists propose a “sentential” attitude theory to replace the propositional one. It entails mental representations that operate in much the same manner as language constructs—propositions are represented internally by arrangements of neuronal “tokens” adhering to an organic system of syntax. Combinations of neurons and neuronal structures into limitless mental “sentences” account for the thoughts a human might entertain and physiologically grounds some surrogate for the beliefs and desires we want humans to have. The “meaning” of a mental token is derived from noting its syntactic relation to other tokens—since every sentence is unique, its combination of relationships with other sentences will be unique, and one may thus infer the meaning of a sentence by analyzing how it is situated in a network of neuronal structures. For example, the correct linkage of neurons X, Y, and Q might “stand for” the belief

145 See id. at 130.
146 This is sometimes called a “language of thought” hypothesis, and is essentially the view of Jerry Fodor. See infra note 193.
that [frogs croak], and looking for this unique combination in relation to other neuron combinations would reveal this representation in a person’s brain. A compendium of varying neuronal shapes and relations allows the folk psychologist to maintain some “belief-like” explanation of psychology and keeps her firmly within the physiological confines of the human brain.

There are difficulties with sententialism, however, that belie folk psychology’s reliance on that theory. Though flexible and seemingly plausible, sentential theorists cannot explain some of the assumptions and ramifications of their view. For one, the origin of the brain’s syntactic system is not revealed, so we cannot determine how the brain links neuronal structures into mental sentences.\textsuperscript{147} Also, how do we know what neuronal shapes “count,” as beliefs or whatever, when analyzing a person’s psychology? As Dennett points out, “[p]sychology is not literary hermeneutics; the ‘text’ is not given.”\textsuperscript{148} The strong implication of sententialism is that since beliefs are represented by mental tokens arranged in a mental language, we must all possess the same “language of thought” in order to believe the same thing.\textsuperscript{149} Now that the propositional attitude theory is forsaken, we cannot make the useful assumption that all persons who “believe” the same thing are just “grasping” a neutral proposition existing in the real world. Under sententialism, each person, in order to believe the same thing as someone else, must have the same exact neuronal components arranged in relation to each other in exactly the same way. Initially, this does not seem plausible, and there is no way to dispositively answer the empirical question of whether all humans construct thought-representations in exactly the same way. Far from convincingly supported, sententialism rests on that unrealistic foundation.

There are other difficulties with sententialism. How does the brain make a syntactic representation out of sensory input? For example, an impulse is sent to the brain by a sensor in the retina, and what the sensor “sees” is reduced to an appropriate “shape” in the brain. What “belief” do we associate with the shape? Suppose that I am sitting in front of a table, on which is situated a red object.\textsuperscript{150} The mental sentence sent by the retinal sensor might


\textsuperscript{148} DENNETT, \textit{supra} note 124, at 137.

\textsuperscript{149} \textit{Id.}

\textsuperscript{150} This example is drawn from DENNETT, \textit{id.} at 144-45.
be "there is a small red spot in the middle of the visual field," or it might be "there is a tomato on the table," or "ten retinal sensors are in state G." There is no way to ascertain what input the brain has received by looking only to the "shape" of the neuronal token created. The only way to determine what the structure "means" is to consider all of its syntactic relations and powers of combination with every other neuronal token, eventually parsing out the entire corpus of a person's neurophysiological make-up in order to specify one belief-representation. As mentioned before, since the representation cannot contain a proposition, our only guide to understanding it would be its relationship to everything else.

Finally, some phenomena cannot be "sententialized." Dennett presents the example of Mike, who "has a thing about redheads." He has idiosyncratic, derogatory views of redheads that influence his attitudes concerning them. This "thing" might be an actual bit of neurological machinery that is "about" redheads, in that it will influence the calculus whenever the topic of discussion is related to redheads, altering the various parameters of the language of thought so that, for example, positive hypotheses about redheads are less likely to be accepted in that particular situation. This "thing" about a certain real object (a redhead) may not be representable as a proposition ("Mike believes that . . .") no matter how many qualifiers and conditional clauses are tacked on. Many of our attitudes are influenced by implicit prejudices, and yet such idiosyncratic "background beliefs" are unquantifiable in the sententalist's mental language.

This discussion of the content problem is meant to expose the unanalyzed assumptions on which folk psychology rests. The principle methods of explaining the basic concept of belief within folk psychology, propositional attitude theory, and sentential attitude theory, are incoherent or implausible. At best they are rationales for dealing with phenomena that we previously could not explain otherwise, and it is not at all obvious that the dictates of common sense psychology are correct, despite their lengthy term in office. The eliminative materialists' critique of content shifts the "burden of proof" of accuracy in cognitive descriptions from the eliminativist challengers to some point in between those thinkers and the theorists who cherish traditional psychological notions. Still, there

\footnotesize

151 Id.
152 Id. at 148.
153 See id.
are even other aspects of the debate that lend support to the eliminativists’ claim that folk psychology, now robbed of its ability adequately to explain beliefs and desires, is woefully incorrect.

D. The Argument from History

From a broader angle, any conceptual system that has remained fundamentally intact for thousands of years must be at least suspected of obsolescence, and folk psychology’s obvious deficiencies as an explanatory tool seem to further that suspicion. As one eliminativist has pointed out:

We do not know what sleep is, or why we have to have it, despite spending a full third of our lives in that condition . . . . We do not understand how learning transforms each of us from a gaping infant to a cunning adult, or how differences in intelligence are grounded. We have not the slightest idea how memory works, or how we manage to retrieve relevant bits of information instantly from the awesome mass we have stored.\(^{154}\)

Eliminativists remind us of past revisionary propositions that seemed equally unfounded, but were eventually proven accurate. Humans have historically been duped by a litany of myths concerning a wide range of natural phenomena, and have been traditionally stubborn in forsaking closely held, yet erroneous beliefs. For example, in the eighteenth and nineteenth centuries it was common knowledge that heat was a weightless substance—“caloric”—contained in the human body (or other materials) in varying amounts (therefore, varying degrees of warmth).\(^ {155}\) Various explanations of nature were built, plausibly, on this premise. It is now clear that heat is not a substance at all, but the result of excited molecules. This new conception is much more successful at explaining phenomena than the old “caloric” conception, and since that substance has never been discovered, we have discarded the theory. Similarly, an educated individual before Copernicus would hardly have doubted that the heavens constituted a huge, slowly turning sphere, with Polaris as its axis.\(^ {156}\) We now

---

\(^{154}\) CHURCHLAND, supra note 9, at 45-46. Churchland wrote this passage in 1984, and science has since made some progress in these areas. See supra notes 118-20.

\(^{155}\) See CHURCHLAND, supra note 9, at 43-44; STEPHEN F. MASON, A HISTORY OF THE SCIENCES 486 (1962).

\(^{156}\) A much simpler form of this conception originated with the Babylonians as far back as 2000 B.C. It was further developed by the Greeks, beginning with Eudoxus in the fourth century B.C. See MASON, supra note 155, at 18-20, 39-41.
subscribe to an astronomy based on a profoundly different framework. The eliminative materialist Paul Churchland notes that, considering folk psychology in light of the incredibly complex workings of the brain, "it would be a miracle if we had got that one right the very first time, when we fell down so badly on all the others."^{157}

E. The Argument from Progress

The final argument, akin to the analysis of history above, is a claim based on the nature of progress. It is best expressed through an example: Aristotle, an intellectual of eclectic tastes, wrote extensively on both morality and physics,^{158} and his ideas greatly influenced the western world. Today, however, while Aristotle's philosophy of morality is still potent, his theory of physics is antique. This obsolescence is due to the numerous discoveries made in the physical sciences since Aristotle's time,^{159} culminating in a modern conception of physics that looks nothing like anything Aristotle claimed. Meanwhile, his moral philosophy fared better because there are no discoveries to make that can render it obsolete. There is no empirical component to morality, or stated differently, there is simply nothing to know. It is not a verifiable science but a discipline based on opinion, and so it cannot be proven incorrect. Such a discipline can potentially maintain its currency for two thousand years.

Aristotle also theorized about psychology,^{160} and his psychological theory, essentially folk psychology, has also not been proven incorrect, and retains its currency today. The implication here is

---

^{157} CHURCHLAND, supra note 9, at 46. Folk psychology has survived this long because of the enormous technical sophistication required to effectively explore the working human brain. See id. The reader should also consider how remote the chances must be that a complete neurophysiological picture will happen to correspond to traditional folk-psychological notions. One of the major folk-psychological theories does argue for this position. See infra notes 166-72 and accompanying text.


^{159} One can trace this progress from Copernicus through Joseph Priestly to Neils Bohr and beyond. See, e.g., Lawrence Bragg, The Atom, in A SHORT HISTORY OF SCIENCE 118, 122-25 (1959); Herbert Dingle, Copernicus and the Planets, in A SHORT HISTORY OF SCIENCE, supra, at 18, 18-26; Douglas McKie, The Birth of Modern Chemistry, in A SHORT HISTORY OF SCIENCE, supra, at 69, 71-77.

^{160} See ARISTOTLE, De Anima, in INTRODUCTION TO ARISTOTLE, supra note 10, at 145, 205-35.
that if folk psychology represents a verifiable science, there should have been at least some discoveries during the past two millennia marking its progress, as there were with every other empirical science. The fact that folk psychology has retained the same nebulous quality as obvious intangibles like philosophy and morality supports the argument that it itself is an intangible—a theory with no basis in the physical elements of the real world. If a science admits no progress from ancient Greece to the present day, one must wonder whether it is a science at all.

An accurate reflection of our true cognitive and behavioral processes may require a "conceptual revolution" during which mental state psychology, and thus the law, will be recognized as a "false and radically misleading conception of the causes of human behavior and the nature of cognitive activity." But before finally exploring an eliminative materialist model of consciousness, one must assess folk-psychological responses to the developments of science and philosophy outlined above.

F. Can Folk Psychology Accommodate Materialism?

Despite the content problem and the circumstantial weight of history, folk psychology may not be doomed. In order to salvage common-sense psychology and the attendant legal apparatuses, supporters of folk psychology must secure the existence of beliefs, desires, intentions, and other propositional attitudes within a materialist framework. This explanation would also have to take account of the scientific advances mentioned above. If folk psychologists reject dualism, the dispute then focuses on various materialist interpretations of scientific developments—the folk mental state interpretation and the eliminativist interpretation. As Churchland puts it, "the question of the relation of mental states to brain states becomes a question of how an old theory (folk psychology) is going to be related to a new theory (matured neuroscience) which threatens in some way to displace it."

---

161 See supra note 159; see also CHURCHLAND, supra note 9, at 43-44 (discussing the decline of incorrect scientific theories). See generally F. SHERWOOD TAYLOR, SCIENCE PAST AND PRESENT (1945) (reviewing scientific progress in various fields).

162 CHURCHLAND, supra note 9, at 45.

163 Id. at 43 (emphasis omitted).

164 This seems to be the current trend. Greenwood notes that "[m]ost ... defenders of folk psychology endorse the materialist assumption that intentional psychological phenomena—if they exist at all—are incarnated in the human brain." Greenwood, supra note 11, at 6.

165 CHURCHLAND, supra note 9, at 61; see also Lewis, supra note 99, passim (arguing...
Theorists present two major versions of folk psychology that manage to operate within the confines of the physical brain, and the continued viability of our present criminal law depends on a successful defense of these theories from eliminativist criticisms.

1. Identity Theories

This first response to the progressive grounding of cognition in physical processes is simple, but elegant. Identity theories posit that, as neuroscientific knowledge increases, we will discover that our mental-state talk directly corresponds to ("identifies" with) certain physical states or processes in the human brain. That is, science will find that "[m]ental states are physical states of the brain." If neurological findings are found to roughly parallel our folk-psychological framework, then current systems can remain intact. Identity theorists seek an "intertheoretic reduction": an instance in which "a new and very powerful theory turns out to entail a set of propositions and principles that mirror perfectly (or almost perfectly) the propositions and principles of some older theory or conceptual framework." The hope is that the old theory, folk psychology, will be reflected perfectly by the new theory, a neurophysiological account of cognition. For example, the mental state of hate will be successfully "reduced" to a neuronal structure physically present in the brain.

Further, if correct, the identity theory allows for a more accurate assessment of mental states in criminal defendants. It would be conceivable to administer a neurological test capable of searching defendants' memories for bits of residue revealing the nature of an actor's beliefs and desires at the time of an incident. The test would simply trace the neurochemical fragments corresponding to certain known mental states.

---

for a theory relating folk psychology to materialism).

166 Churchland, supra note 9, at 26. According to this theory, everyone sharing a certain belief or desire would have the same neuronal state in their brains. If A and B both believe that C is a murderer, then A and B will both possess a similar set of neuronal connections in their heads that represents the belief that C is a murderer. See Dresser, supra note 2, at 31.

167 Churchland, supra note 9, at 26. Identity theory is also known as "reductive materialism" since it contemplates a reduction of mental states to material components. See id.

168 Cf. id. (maintaining that, in the same way that light and sound can be measured in terms of their component parts, mental states can be anatomized into corresponding, measurable physical states).
Identity theory may at first appear to rely on a massive coincidence between our folk notions of cognition and the complex neurophysiological workings of our brains. It does seem unlikely that mature neuroscience will just happen to reveal physiological structures coinciding with the mental states we have attributed to ourselves for thousands of years. But when we consider other examples of this "reduction" from superficial descriptions to concrete anatomy, the idea gains credibility. Consider water—a clear, tasteless fluid of a certain consistency, according to our folk conception. But water is "actually" \(x\) oxygen atoms + \(2x\) hydrogen atoms. If this were told to an ancient Egyptian, she would be baffled and incredulous, even though \(\text{H}_2\text{O}\) is identical to "water." \(^{169}\) A folk psychologist following the identity theory hopes that mental states like intentions are really identical to corresponding, discoverable neuro-anatomical brain states.

Unfortunately, it is still incumbent upon the identity theorist to avoid the content-problem previously revealed. \(^{170}\) Once again, how can a neuron configuration be the-belief-that-frogs-croak, or the-desire-to-be-taller? Identity theorists answer this challenge by avoiding the attribution of content to mental states and subscribing to a brand of sententialism, portraying beliefs and desires as inferential states within a greater network of such states. A sentence, such as "I desire to be taller," is always an interlocking element of numerous sentences to which that sentence bears relation. The sentence in question will conflict with some sentences, such as "I desire to be shorter," and complement others. This pattern of relations will be unique for every particular sentence, and so the proposition will fill "a distinct inferential role in a complex linguistic economy." \(^{171}\) Viewing the brain as a complex network, mental states now need only be a particular configuration of neurons in relation to all other configurations of neurons in order to possess a certain meaning or proposition. The role played by the configuration represents its significance and the counterintuitive concept of inherent "content" can be discarded. \(^{172}\)

---

\(^{169}\) Other examples are legion. Churchland cites several, e.g., light is identical to electromagnetic waves, and lightning is a massive release of electrons between clouds. See id.

\(^{170}\) See supra part III.C.

\(^{171}\) CHURCHLAND, supra note 9, at 31. This description of identity theory networks is derived from Churchland's explanation. See id.

\(^{172}\) The philosopher Donald Davidson has formulated his own alternative version of the basic identity theory, known as anomalous monism. See Horgan & Woodward,
But we have seen why this conception does not work, and despite the support it derives from successful reductions in the past, the identity theory does not seem entirely plausible. It would still be a coincidence of some magnitude for our relatively crude common-sense psychology to correspond, state for state, to the workings of the most complex structure ever encountered by humanity. Identity theorists are unable to answer eliminativist criticisms of mental state psychology and their explanation does not ensure the continuation of folk-psychological systems. Finally, reductionist descriptions of the brain as "nothing but an aggregate of [neurons]" have been labeled extremely simplistic. Most materialists today are wary of the reductive proposition, and prefer either functionalism, discussed below, or eliminativism.

2. Functionalism

This offshoot of the identity theory is far more viable than its predecessor. Instead of insisting on a consistent set of laws connecting, one-to-one, particular mental states to particular neurophysical states, cognitive theorists might claim that a single mental state corresponds to several possible brain states. For example, pain is a mental state definable by its function: one might describe it as an event accompanying structural damage, causing temporary paralysis, and traumatizing its victim. We could also

supra note 5, at 154. He notes that folk-psychological systems and neurophysiological systems are distinct, and "operate under 'disparate commitments.'" COLIN McGINN, THE PROBLEM OF CONSCIOUSNESS 170 (1991). Physical processes correspond to a set of chemical and physical laws, while mental explanations are beholden to a commonly conceived view of rationality. The two are seemingly irreconcilable, leading Davidson to assert that a complete understanding of all the intricate workings of the brain would not be equally revealing of the nature of consciousness and volition. See id.

By this account, the eliminativist mission of exploding the myth of folk-psychological states within the brain is a nonstarter, since a full neurophysiological understanding is unable to look anything like a psychological system. Davidson's view has received much criticism. See, e.g., id. at 170-71 (noting that even if Davidson were correct, there would be no theoretical obstacle to asserting causal laws between mental and physical states, as occurs in basic identity theories).

173 See supra part III.C.2. (discussing sententialism).

174 That is, the brain. Even supporters of folk psychology admit that this is not likely. See Horgan & Woodward, supra note 5, at 154.

175 This is really the crucial criterion for which we search. A theory that could demonstrate the invalidity of eliminativist materialism would thus protect our legal systems from massive revision, or at least obsolescence.

176 BUNGE, supra note 99, at 8.

177 See id.

178 See Dresser, supra note 2, at 31-32.
assert that pain corresponds to certain neurophysiological events, but, since its definition as a folk-psychological state is only functional, a concrete link to a specific neuronal configuration is not required. Martians might feel pain as well as humans, but the Martian variety may have a completely different physiological representation. Because the definition of pain is functionally-based, this physiological disparity does not suggest that the mental event of pain cannot exist. It can easily exist as a consciously felt phenomenon, but be manifested by numerous different brain events.

Functionalism “is probably the most widely held theory of mind among contemporary philosophers, cognitive psychologists, and artificial intelligence researchers,” and it represents the best hope for salvaging a legal system that relies on mentalistic explanations. The essential difference between functionalists and identity theorists is the former’s assertion that “any physical system—organic or inorganic—having the same functionally related internal states as a human being, [has] the same ‘mental’ states that we do.” Functionalists define mental states by the effect they have on other mental states and by their effect on the functions of the body and bodily behavior. Since pain can be functionally defined, any system that can exhibit the required characteristics can feel pain. Functionalists allow for many possible neurological manifestations of any one mental event—the functional state of pain is allowed to be caused by any number of things. Functionalism is thus more flexible than its relative, the identity theory, which seeks but one defined neurophysiological correlate for each mental state. Most functionalists actually agree with identity theorists that neurophysiological research will eventually yield some finite disjunction between brain states and folk-psychological states, but functionalists find it

---

179 This avoids the identity theory coincidence. Notice, though, that our neurological test for mens rea, supra text accompanying note 168, would no longer work, since mental states would no longer have a singular, consistent neuro-anatomical label. See Dresser, supra note 2, at 32.
180 It may occur to the reader that any definition of pain is problematic, since there seem to be so many species of that state. However, the functionalist theory easily accounts for this: pain is not dependent on a specific neuronal pattern, so there can be many different configurations corresponding to the many different varieties of pain humans seem to feel.
181 Dresser, supra note 2, at 32 (footnote omitted).
182 Id.
183 See CHURCHLAND, supra note 9, at 36.
extremely unlikely that the physiological correlates discovered will match with mental states at a one-to-one ratio.

As with Davidson's proposal, functionalism allows for the folk-psychological notions inherent in the law, since its system of discussing mental states is divorced from technical neuroanatomical talk. Definitions of mental events are dependent on their causal characteristics, so quibblings over physical manifestations are irrelevant to their continued existence. Indeed, the functionalist might argue that since mental states are reducible to any number of neurophysiological phenomena, to "search" for them in the human brain is pointless.

But the functionalist account is not without difficulties. Since its definition of mental states is not biologically based, strict functionalism would warrant extending legal personhood to robots and other systems capable of the same functional representations as humans. It is also important to realize that functionalism suffers a criticism similar to the one leveled against the identity theory: There is no reason to believe that mental states, though functionally explained, are an accurate account of the "states" we actually experience. For example, Paul Churchland points out that functionalism would have saved the misguided pseudoscience of alchemy from replacement by molecular chemistry if alchemists had simply asserted that the four "fundamental spirits" of traditional alchemy were easily definable by their functional states. Those spirits were useful in a limited capacity, allowing alchemists to perform limited physics experiments and transformations, but still radically misrepresented physical realities. Since alchemy gave way to the much more useful and powerful discipline of elemental chemistry, we should reconsider our stubborn adherence to mental state psychology before shoring it up with functionalist arguments. We may find new cognitive explanations as much an improvement on old systems as chemistry was over its forebearer.

Functionalism also gives too vague an account of mental states to be satisfying. Since functionalism allows psychological theories

---

184 See supra note 172.
185 See Dresser, supra note 2, at 32.
186 See supra text accompanying note 173.
187 Mercury, sulphur, yellow arsenic, and sal ammoniac.
188 See Churchland, supra note 6, at 78-81.
to look nothing like verifiable sciences,¹⁸⁹ it does not explain the
course of mental phenomena, but avoids the need for such explana-
tions.¹⁹⁰ In this regard it is similar to behaviorism, which sought
to explain behavior solely through environmental stimuli, while
totally ignoring the issue of mental states.¹⁹¹ Functionalism then
must also share behaviorism’s faults—it does not dispose of the
possibility of a pending neuroscientific revision of our common
sense notions. We have considered the rapid advance of brain sci-
ence,¹⁹² and ignoring it in favor of a purely functional account
does not ensure that neuroscientific realities will simply “go away.”¹⁹³

¹⁸⁹ See Dresser, supra note 2, at 32; Fodor, supra note 3, at 30.
¹⁹⁰ See supra notes 71-80 and accompanying text.
¹⁹¹ See supra note 3, at 29.
¹⁹² See supra part III.B.
¹⁹³ Note that just as functionalism improved upon the identity thesis, Professor
Jerry Fodor has presented an interesting improvement on functionalism, which he
describes as a “species of Functionalist psychology.” Fodor, supra note 3, at 31. This
is the “language of thought” theory criticized by Dennett, see supra notes 148-53
and accompanying text. Fodor begins by asking an often overlooked question: How are
new thoughts generated? See Fodor, supra note 3, at 36-37. If beliefs and desires
correspond to neuronal configurations, then how are new beliefs and desires created?
Fodor finds the answer in the structure of everyday sentences—new sentences are
easily generated by the arrangement of constituent parts, such as words and phrases.
Following this example, beliefs, desires, and other propositional attitudes must be
built out of “smaller” elements, prompting Fodor to posit the existence of “mental
representations.” Id. at 36-44. These representations are “tokens,” symbols linked
to constitute representations of propositions. After reducing beliefs and desires to
manipulable tokens, Fodor jumps to computer technology, which cultivates the art of
connecting semantical properties, such as words, phrases, and other symbols, with
causal properties. This is relatively straightforward—a more commonly known term
for the intelligible connection of semantics with causal roles in language is syntax.
Sophisticated computers represent syntax regularly, and by this linkage of mental
representations to computer technology, Fodor is on his way to explaining mental
state psychology within a materialist framework. Mental processes are explained as
causal sequences (sentences, if we analogize to language) of tokens (words and
phrases) that create ideas, beliefs, or desires. Fodor’s thesis is more complex than
presented here, but this sketch is adequate for our purposes. For a lengthy

Fodor is mentioned because his theory of mental state psychology is a slight
improvement on the sententialism discussed earlier, and because it can claim limited
support from cutting-edge neuroscience. As noted earlier, research has demonstrated
that ideas are not contained in one central location of the brain, but are built out of
ingredients drawn from different areas. See supra notes 116-17 and accompanying
text. This would seem to support the concept of thoughts constituted by smaller
representations. In addition, certain disorders have revealed the extraordinary
specificity of representation in certain brain areas, further supporting the thesis that
mental states are created by knitting various subordinate representations together.
For example, individuals who sustain damage to the temporal segment of the left
Eliminativists would assert that neither identity theories nor functionalism can rely on neuroscientific findings to support their claims, and so both are in doubt. Identity theories rely on an improbable psychobiological coincidence, functionalism ignores consideration of empirical neurological processes, and both are unable to account for the logical difficulty of attributing content to brain states. Neither can deny that eliminative materialists are gaining support for their radical predictions, as neuroscientific discoveries continue to add fuel to the fires of cognitive revisionism. The eliminativists' reliance on "maturing" neuroscience lingual gyrus suffer from color anomia, a disorder which prevents the subject from attributing names to colors. See Damasio & Damasio, supra note 117, at 90. The victim can see colors, match colors, and otherwise negotiate different hues unimpaired, but is completely unable to name them. No other mental or linguistic deficiencies are present. See id. at 90-91. Similarly, damage to different regions of the brain will result in distinct, extremely concise deficiencies in language production or understanding, such as the inability to rhyme, or understand common nouns. See id. at 93-94. The diffuse nature of language elements and the specificity with which a particular element can be removed could hint at a rich representational framework within the brain.

But while Fodor's view has the beginnings of empirical support, it suffers the same criticisms as sententialism, and the eliminative materialists have suggested strong reasons to doubt that mental state conceptions are accurate ones. See supra part III.A. Fodor's explanation has some other problems: first, he is unable to tell us how the mental tokens are linked together by the brain. He seems to assume some faculty in the brain that oversees the coherent linking of tokens for use in beliefs and desires, but there is no direct evidence for the existence of such an instrument. Where are the "semantic constructions" performed? It would be easy to fall back on the Cartesian materialist fallacy described earlier. Does Fodor believe in a "central command" coordinating thought creation? Second, Fodor cannot account for the qualitative nature that folk-psychological states such as pain or fear seem to possess. See supra notes 179-80 and accompanying text. He can adequately explain their construction and relation to other states, but cannot explain the sensations that accompany them, at least not without conceding to an identity theory based on physiological events. Reference to an identity theory would allow Fodor to link his representations with physical properties causing physiological changes. Third, Fodor's view may distinguish too finely between psychological states—two individuals with even slightly different syntactic networks in their brains presumably could not have the same beliefs. See Dennett, supra note 147, at 499. Finally, Fodor's view is open to the same criticisms leveled against functionalism generally, such as Churchland's alchemy example. See supra notes 186-88 and accompanying text.

There are those, however, who argue that eliminative materialism is inherently inconceivable to conscious human beings. Lynne Rudder Baker agrees with the eliminative materialists that it is simply not possible to assign content to a system of interlocking neurons. See LYNNE RUDDER BAKER, SAVING BELIEF 86-88 (1987). The difficulty here, as mentioned earlier, is making "a particular array of molecules stand for something like the belief that the rifle is loaded." Dresser, supra note 2, at 33. Baker agrees with Churchland and others that mature neuroscience will not coincide with folk-psychological systems, see BAKER, supra, at 113, but disagrees with the
for its support presents stubborn difficulties for folk psychologists and legal theorists, since both are forced into a slow retreat as science progresses. The law, being gradual in change, could suffer a series of accommodations to psychobiological findings, perhaps finally designing its systems of blame around a model like the one below.

G. An Eliminative Replacement

So what, ultimately, would the eliminativist picture of consciousness look like? Modern advocates present a challenging case. Daniel Dennett recently developed a radical new model of brain processes that severely undermines common-sense notions of consciousness and personal responsibility. Though not specifically designed for the eliminativist cause, Dennett's richly detailed conviction that advancing neuroscience will reveal the "true" nature of our conscious existence. She seems to assert that purely materialist stories must remain incomplete, since one cannot dispense with notions of intentionality and belief and still operate or communicate coherently. See id. at 148. Similarly, she asserts that much of our neuroscientific research, and research generally, operates on assumptions of intentionality, since researchers are consciously looking for something, testing hypotheses, etc., see id. at 171, and therefore mental states such as those are truly indispensable.

Baker seems to lack imagination. She suggests that "denial of the common-sense conception may be a kind of cognitive suicide," since it will remove our method of making reality intelligible. Id. at 148. Put another way, she asserts that language, for example, "can be meaningful only if it is possible that someone mean something." Id. at 140. Baker seeks thereby to demonstrate why contentful mental states are crucial to our lives—we could not do or say anything of significance without them. But this criticism only highlights our current inability to conceive of the appropriate answers to our cognitive questions. Baker notes that if concepts of belief and desire were eradicated, predictions of behavior would become impossible and legal processes would no longer make sense. See id. at 130-32. These dire prophecies, however, need not be fulfilled. An inability to conceptualize a new state of affairs is no reason to declare that new proposals are wrong; Baker would have the eliminative materialists explain the intricate workings of a new science that has not been completely invented yet. Her criticisms would have kept humans out of the sky, and yes, out of molecular chemistry. Admittedly, there is a distinct difference between talk of technological advances and talk of redefinitions of consciousness, but "inconceivability" is hardly a sufficient reason to abandon a new scientific exploration.

Indeed, the only definitive way to remove the threat to folk psychology is to find dispositive neurophysiological proof of its accuracy. This is no small task, but folk psychology does have the advantages of an incumbent, since it is well entrenched and eliminative materialism is the new challenger. The eliminativists have the difficulty of "proving a negative." Thus some will stubbornly adhere to common-sense psychology until it can be shown that such a system does not and cannot exist anywhere in the brain at all.

See generally DENNETT, supra note 40. Dennett does not consider himself an
ELIMINATIVE MATERIALISM

proposal fits well with eliminativist endeavors to shake the hold of antique psychological notions. As noted previously, Dennett first disposes of the stubborn and insidious fiction of the "Cartesian Theater" where conscious events are internally "viewed." He similarly dispels the notion of any cerebral "central command" akin to Descartes's pineal gland, and if modern neuroscience has fully supported any philosophical claim, it is this one. After excising that conceptual clutter, Dennett fills the gap with a "Multiple Drafts" theory of consciousness: Instead of there being a linear, serial progression of thoughts through awareness, there are only multiple neurophysiological subprocesses operating simultaneously, thereby creating numerous "drafts" of consciousness. Most of these drafts remain uncompleted, and so unconscious, but groups of subsystems will rise to temporarily dominant roles, thus jointly creating what we introspectively consider our "awareness" and all of its phenomena. The rapid succession of these groups in aggregate account for our perception of a unitary self, though this is actually a mask for the conflicting subsystems beneath.

 eliminates materialist per se, and scorns the use of such labels. See id. at 459. This is a bit puzzling, since Dennett devotes an entire chapter of another book to debunking folk-psychological myths about the possibility of "beliefs" and other propositional attitudes. See DENNETT, supra note 124, at 117-211. Dennett himself describes the chapter as "a demolitional tour of that philosophical black hole: propositional attitude psychology." Dennett, supra note 147, at 499.

197 Dennett devotes his entire book to a meticulous account and defense of his theory. This discussion only touches upon its basic framework.
198 See supra notes 103-04 and accompanying text.
199 See supra notes 105-14 and accompanying text.
200 The most fundamental PET scan would support this assertion, unless you believe in a nonphysical, irreducible, unscannable mind, thus dooming you to dualism. See supra notes 115-23 and accompanying text.
201 See DENNETT, supra note 40, at 253-54.
202 See id. at 254. Dennett calls these groups of subprocesses "coalitions of specialists." A helpful metaphor is to consider consciousness the synergistic effect of a biological "political community" where elements vie for, and temporarily hold, power. The idea does not seem strange when one considers that without the Cartesian illusions, an account must be given for what we experience as our "stream of consciousness," to borrow William James's phrase. Dennett states:

[The brain] performs a sort of internal political miracle: It creates a virtual captain of the crew, without elevating any one of [the groups of subprocesses] to long-term dictatorial power. Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences rather than an interminable helter-skelter power grab.

Id. at 228.
203 As Dennett explains: "[Y]ou don't have any personal familiarity with—any
Some of these neurological subsystems are "hardwired," like the basic operational protocols of a computer, while others are implemented much like software programs.\textsuperscript{204} Hardwired neuronal programs might be reflex systems, while the simultaneous neuronal subprocesses are different software programs constituting different "states of mind." Dennett, borrowing from artificial intelligence research, envisions consciousness as a "virtual machine" superimposed on organic processes.\textsuperscript{205} He asserts that the currently dominant set of subsystems can be envisioned as a currently-running computer program, in that while a program is running, the computer will give the illusion of a device designed for a specific purpose; if one is running a spreadsheet program, the computer will appear to be a machine solely devoted to creating spreadsheets. Since this is an illusion, the computer and its program are really a "virtual" (spreadsheet) machine, and only for the period during which the program runs.\textsuperscript{206} Extending the analogy to consciousness, awareness is really a "structured recipe of hundreds of thousands of instructions that give the [brain] a huge, interlocking set of habits or dispositions-to-react."\textsuperscript{207} The conflicting subsystems or "instructions" are partially individualized by environmental and culture influences,\textsuperscript{208} and as they become dominant, they present the person’s "user interface" with his surroundings—we interact with others as conscious individuals much as we interact with the keyboard of a program-running computer.

If this all seems somewhat bizarre, an example might make it more compelling. Dennett asks us to consider an instance of 'direct access to'—processes happening in a thousand channels at once, even though that is what is going on in your brain. Your only access . . . comes in a sequential ‘format’ that is strikingly reminiscent of [basic computer] architecture." \textit{Id.} at 215. That "sequential format" is what we experience as consciousness.

The assertion that we do not have access to most of our cerebral workings is not controversial. Humans are hardly conscious of most of the processes occurring in the brain at any one time. For example, if you have one of several hundred test copies of this publication, at some point a specially imbedded wire filament will give you a mild electric shock. If you have an unaltered copy, there is no danger.

Now, if you are particularly credulous, a small organ in your brain called the amygdala suddenly became extremely active, but you were completely unaware of this particular organic activity. \textit{See} Begley et al., supra note 112, at 66 (presenting a similar experiment).\textsuperscript{209} See \textit{Dennett}, supra note 40, at 215-16.\textsuperscript{205} \textit{Id.} at 254.\textsuperscript{206} See \textit{id.} at 215-16.\textsuperscript{207} \textit{Id.} at 216.\textsuperscript{208} See \textit{id.} at 254.
language generation, in which a person needs to launch an insult at another.\textsuperscript{209} Dennett notes that our speech acts must develop from a general strategy to a detailed action, perhaps along a string of internal directives:

\begin{quote}
(1) Go on the offensive!
(2) Do something nasty but not too dangerous to him!
(3) Insult him!
(4) Cast aspersions on some aspect of his body!
(5) Tell him his feet are too big!
(6) Say: “Your feet are too big!”
(7) Utter: yer FİT ār tū bīg?\textsuperscript{210}
\end{quote}

Various neurological subprocesses construct competing drafts of the imminent utterance, all of which struggle for expression. All but one will be set aside, as the actor “decides” to say “Your feet are too big!” instead of the more inflammatory “I’m going to blow your brains out!” or the somewhat lame “You big meany!” After the exchange, the actor certainly considers what other things he might have said, those utterances we all conceive of but never dare to express. These are the “rejected drafts” created by other subsystems in the process.

But if no particular set of systems dominates for long, then who “decides” what subprocess “wins” the contest for expression, and which ones “lose”? Dennett’s answer: No one. The decision jointly emerges from the interaction of processes, much like the game of “question and answer”: \(X\) asks \(Y\) a question, and \(Y\) provides one of two answers (yes or no) based on the structure of \(X\)’s inquiries. \(X\) is able to ask whatever she wishes, and from the interplay a story will develop, even though no one party has created it. The process of speech creation or other intentional action is similar—no one subsystem need be the author. All of the brain’s subsystems contribute to the cause.\textsuperscript{211}

We see in Dennett’s model the culmination of neuroscientific advances and eliminativist philosophy. If Dennett’s well-considered predictions on the future of neuroscience result in the above reality, the legal system will face some serious problems, such as what to make of neuronal “programs” instead of mens rea. How could the

\textsuperscript{209} See id. at 234-37.
\textsuperscript{210} Id. at 234-35.
\textsuperscript{211} See id. at 101-11, 238-39. This lack of established “self” has serious legal implications that are explored infra part IV.C.
law even begin to reflect these changes? Similarly, the reader would be right to point out that folk psychologists and eliminative materialists alike are wallowing in a good amount of conjecture, and so for an institution as substantial as the criminal law to reflect changes based on incomplete neuroscientific evidence and some aggressive philosophizing would be, um . . . criminal. A complete revision of the legal institution based on Dennett’s model or any other eliminativist picture may indeed be premature, so a more concrete and verifiable example of eliminativist concerns will be considered first. A practical example of neuroscience mucking up the legal works may help convince the reader that the potential for large-scale legal dilemmas is very real, and that those dilemmas may arise very soon.

H. An Illustration: Benjamin Libet and the Delay of Conscious Intention

To sum up the story thus far, this Comment began by describing the legal system’s reliance on traditional psychological notions, and revealing the potential gap between modern cognitive theories and legal conceptions. It then noted the law’s willingness to entertain biology-based explanations of behavior and explored the potent eliminativist mixture of revisionism and biology-based support. Although the major folk-psychological theories were not prima facie failures, they were unable to remove the eliminativist threat to a legal system reliant on folk psychology, and our foray into cognitive theory ended with a description of Dennett’s cutting-edge vision of consciousness.

Besides a proposal that we completely discard the current legal system, it is difficult to make reformative suggestions for the criminal law based only on the general eliminativist platform. There is, however, current neuroscientific evidence that seems to require some legal revision. An assessment of recent research concerning the “voluntary act” will illustrate the potential difficulties presented by an eliminative approach to law.

We have seen that the “volitional theory” underlying the law is actually an essential element of folk psychology, an element

---

212 Posed rhetorically here, some options are seriously considered infra part IV.B.
213 To recap: Beliefs and desires develop into intentions, and a volition, or “willing,” of bodily movement is the executory state between the intention and physical action. See supra notes 27-32 and accompanying text (discussing actus reus). Note the distinction between “volitions” and “intentions”—you will a bodily
that is deeply ingrained.\textsuperscript{214} There is no lack of legal theorizing on this subject,\textsuperscript{215} but little of it deals with neuroscientific findings in relation to the criminal law's voluntary act requirement.\textsuperscript{216} This is unfortunate, since some fascinating work has been done in this area. Recent emphasis on neurophysiology\textsuperscript{217} has renewed interest in research on cerebral processes pursued by Dr. Benjamin Libet in the late 1970s through 1985.\textsuperscript{218} Libet's experiments were fairly straightforward. He instructed his subjects to sit facing a clock and concentrate on a revolving spot on its face. At random movement; what you intend to do with it is something else entirely. Obviously, volitions are not the same as actual acts. If your left arm is paralyzed and you will it to move, you have exercised a volition, but not an action.

\textsuperscript{214} Folk psychology does not get more basic than the assertion that human beings are responsible for their bodily movements. Legal injunctions based on this assumption are timeless. See, e.g., Exodus 20:13-17 (Ten Commandments). In fact, the Talmud has an interesting passage on the mens rea/actus reus requirements. See BABYLONIAN TALMUD, Seder Nezikin, Tractate Sanhedrin, 91a-91b (Isidore Epstein trans., 1935) (relating a parable in which a lame man, directing a blind man, is thus able to steal some figs); see also ARISTOTLE, supra note 10, at 478 (discussing acts in accordance with opinions).

For a more romantic meditation on the moral significance of personal choice, see Steinbeck's East of Eden:

\begin{quote}
The American Standard translation [of the Old Testament] orders men to triumph over sin . . . . The King James translation makes a promise in "Thou shalt," meaning that men will surely triumph over sin. But the [original] Hebrew word, the word \textit{timshel}—"Thou mayest"—that gives a choice. It might be the most important word in the world. . . .

. . . Why, that makes a man great, that gives him stature with the gods, for in his weakness and his filth and his murder of his brother he has still the great choice. He can choose his course and fight it through and win.\textsuperscript{215}
\end{quote}


\textsuperscript{216} Saunders does skirt some cognitive philosophy, but a dualist assumption underlies his entire proposal on the volition-act connection. See Saunders, supra note 100, at 454. Moore has given a recent and more thorough review of the subject. See MOORE, supra note 29.

\textsuperscript{217} See generally SCI. AM., supra note 12.

\textsuperscript{218} This Comment makes use of Libet's most recent article on his research. See Benjamin Libet, \textit{Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action}, 8 BEHAVIORAL & BRAIN SCI. 529 (1985) [hereinafter Libet, \textit{Cerebral Initiative}]. Libet later refines the conclusions reached in that article in response to peer commentary. See Benjamin Libet, \textit{The Timing of a Subjective Experience}, 12 BEHAVIORAL & BRAIN SCI. 183 (1989) [hereinafter Libet, \textit{Subjective Experience}]; Benjamin Libet, \textit{Are the Mental Experiences of Will and Self-Control Significant for the Performance of a Voluntary Act?}, 10 BEHAVIORAL & BRAIN SCI. 783 (1987) [hereinafter Libet, \textit{Mental Experiences}].
times after the beginning of the test, the subject spontaneously flexed the fingers or wrist of his right hand, paying close attention to the spot's position at the onset of the desire or urge to act. Simultaneously, Libet monitored the onset of the cerebral processes that initiate voluntary acts via electrodes attached to the scalp of the subject. Libet then compared the time of initiation for cerebral processes with the time the subject designated as the first moment he became aware of desiring to act. What Libet found after more than forty trials was that onset of cerebral processes preceded conscious intentions by almost a third of a second (300 milliseconds). In plain English, Libet found that the brain initiates action "before there is any ... subjective awareness that such a decision has taken place." It begins movement long before you are aware of desiring movement.

After digesting this for a moment, the implications are, for most, quite startling. Libet's straightforward finding has stopped us in our legal tracks, as we are left passively watching the actions our unconscious neurophysiological processes decide to take. Before insisting that Libet must have performed the experiment incorrectly, the reader should know that the procedure is not complex by neuroscientific standards and has been replicated.

---

219 See Libet, Cerebral Initiative, supra note 218, at 529-30.
220 See id. These cerebral processes were measured by an electrophysiological "readiness potential ('RP')," which is a "scalp-recorded slow negative shift in electrical potential generated by the brain." Id. at 529. These RPs always precede voluntary bodily movements. See id.
221 See id. at 532.
222 See id. at 529. Within the framework of chemical reactions, 300 milliseconds is a considerable amount of time.

Also, Libet's experiment targeted "spontaneous" as opposed to "pre-planned" actions, though Libet sees no reason for there to be a physiological difference between the two. See id. at 536. Similarly, the distinction would make little difference for our purposes. For example, a general intention to kill someone must be followed by specific, "spontaneous" intentions, such as when exactly to raise the rifle, and when exactly to pull the trigger. Actions involving no pre-planned general intentions, such as killing in response to provocation or attack, are dealt with infra part IV.B.2.a.
223 Libet, Cerebral Initiative, supra note 218, at 536.
224 I would advise doing so. Our common psychological notions are deeply entrenched, and Libet's pronouncement is extremely counterintuitive. The mental self-discipline required to envision this is on a par with removing Dennett's "Cartesian Theater." See supra note 111 and accompanying text.
225 Recall the previous assertion that most cerebral processes normally occur unconsciously. See supra notes 219-23 and accompanying text.
226 See, e.g., R. Näätänen, Brain Physiology and the Unconscious Initiation of Movements, 8 BEHAVIORAL & BRAIN SCI. 549 (1985). Näätänen even tried to "fool" his own brain by avidly reading a book and then suddenly acting on a completely
So much for the neuroscientific difficulties of the criminal law being mostly "conjecture." Rodney Cotterill has described Libet's findings as "devastating . . . to the dogma of free will," and Libet himself appears to view volitions as epiphenomena: They themselves do no "causing" of bodily movements, but because of the unconscious nature of the initiating cerebral processes, they let you think that they did. Our next question glaringly presents itself: How do we hold individuals legally responsible for their bodily acts if unconscious processes dictate those bodily acts? Are all those determinists right, after all?

Libet, after casting us down, raises us up. He explains that we exercise "veto" power over the volitions initiated by our unconscious brain activities. Once they appear in consciousness, the volitional processes can be permitted to progress unobstructed, or be consciously prevented from resulting in a motor act. Libet bases this finding on a detail of his timing trials. He determined that cerebral initiation of movement begins approximately 550 milliseconds before actual movement, and approximately 300 milliseconds before conscious awareness. This leaves approximately 250 milliseconds between conscious awareness and muscle contraction. The actor uses this interval to assess the action about to

spontaneous decision to move about. His brain was not deceived. Initiation of cerebral processes preceded his unpredictable outbursts by a sizeable period. See id. at 549. Näätänen points out that consistent monitoring of a subject's readiness potentials would reveal in advance when the subject was going to "experience an intention." Id.

Libet's 1985 article is followed by 19 pages of commentary by his colleagues in the fields of physiology, psychiatry, neuroscience, and philosophy. The commentary ranges from positive to negative, and deals with various methodological issues beyond the scope of this paper. None of the criticisms are fatal to Libet's findings, and Libet himself responds to his critics with a lengthy rebuttal. See Libet, Cerebral Initiative, supra note 218, at 539-64.

227 See supra text accompanying note 73.
228 COTTERILL, supra note 99, at 261.
229 In his discussion of speech generation, Dennett expresses similar sentiments: Our actions generally satisfy us; we recognize that they are in the main coherent, and that they make appropriate, well-timed contributions to our projects as we understand them. So we . . . assume them to be the product of processes that are reliably sensitive to ends and means. That is, they are rational, in one sense of that word. But that does not mean they are rational in a narrower sense: the product of serial reasoning.

Dennett, supra note 40, at 252 (citations omitted).
250 See Libet, Cerebral Initiative, supra note 218, at 537.
251 See id. ("Such a [veto] role is feasible since conscious intention is reported to appear about 150 to 200 ms before the beginning of muscle activation . . . even though it occurs several hundred ms later than the cerebral initiating processes.")
be performed and then to allow or stop it. Based on this finding, Libet claims that notions of free will and individual responsibility need not be rejected, just revised: "Processes associated with individual responsibility and free will . . . 'operate' not to initiate a voluntary act but to select and control volitional outcomes."232 But Libet's empirical evidence for this assertion is not as strong as that supporting his original delay-of-consciousness finding, partly because present technology is not up to the task.233 The veto phenomenon also creates certain neuro-philosophical difficulties, since Libet is left presenting a dualist design for the brain.234 Several commentators have noted this, and since dualism appears to be all but defunct, Libet's attempt to save the concept of individual responsibility may not succeed. Arthur Danto accuses Libet of being "in the grip of a kind of metaphysical hysteria," finding it incoherent that Libet would "at the end of his paper, . . . postulate functions [the veto power] whose existence would be incompatible with everything he had up to that point been at pains to show."235 Indeed, Libet balks, much as Dennett does, when he

The concept of a veto function is apparently not controversial. See Eckart Scheerer, Conscious Intention is a Mental Fiat, 8 BEHAVIORAL & BRAIN SCI. 552, 553 (1985) (noting the discovery of a conscious veto phenomenon, or "flashlike counter-command," as early as 1935).

232 Libet, Cerebral Initiative, supra note 218, at 538. In a legal context, this less significant role for intention would lead to some changes. See infra notes 285-309 and accompanying text.

233 See Libet, Cerebral Initiative, supra note 218, at 538 ("[T]here is presently no technique available for recording and analyzing any RPs that may be associated with . . . spontaneous, irregularly appearing conscious urges to act that do not lead to an actual motor event."). Libet does present some support for the veto proposal, see id., as does Scheerer's findings, see Scheerer, supra note 231, at 553.

234 Despite Scheerer's work, supra note 231, Libet's assertion of a conscious veto power has drawn strong criticism. The presence of unconscious cerebral forces and a conscious veto power combine to create the interesting implication that Libet is a dualist. Though he denies espousing a stance on the mind-brain problem, see Libet, Mental Experiences, supra note 218, at 784, Libet appears logically cornered: urges are unconsciously, neurophysiologically caused, and yet Libet poses no such origins for the "veto" power. This has even been called "double dualism"—one kind of mental event, the intention, emerges from neurophysiological processes and is thus a product of interactionist dualism, and another mental event, the veto power, is completely nonphysically based, and is thus the product of classic Cartesian dualism. See R.J. Nelson, Libet's Dualism, 8 BEHAVIORAL & BRAIN SCI. 550, 550 (1985); see also Charles C. Wood, Pardon, Your Dualism is Showing, 8 BEHAVIORAL & BRAIN SCI. 557, 557 (1985) (insisting on Libet's dualism and recasting Libet's findings into a materialist framework). Both of these authors wield dualism as an accusation—further evidence of dualism's widespread rejection in cognitive and philosophical circles. See supra text accompanying note 102.

235 Arthur C. Danto, Consciousness and Motor Control, 8 BEHAVIORAL & BRAIN SCI.
stumbles upon what might be the inevitable truth of eliminativism. Just as Dennett, after expostulating a sweeping neo-cognitive picture, attempts to save the concept of rationality and responsibility with his "intentional stance," Libet declares his conclusions and then scrambles to save us from determinism with the conscious "veto power." In truth, both Dennett and Libet have landed squarely in Churchland's lap—supporting the cause for a complete rejection of mental state psychology. In fact, far from requiring accusations of dualism, Libet actually provides further support for Dennett's model of consciousness, and thus for eliminative materialism.

Libet's conclusions concerning the delay of awareness of intention need not be startling. The difficulties of Danto, Nelson, Wood, and others who cannot reconcile Libet with their view of reality all stem from the most stubborn of folk-psychological myths: the myth of the unitary mind. This Comment has twice mentioned the pitfalls of Cartesian materialism, and once again it has arisen to confuse us. Those still beguiled (despite several warnings) surely protested Libet on the grounds that the elusive "you" simply must command your own volitions, and not the other way around! But if we embrace a Dennett-like vision of the brain, not only do Libet's findings make sense, it is difficult to envision how else intentionality might operate. Eliminate Descartes's apocryphal "command center," and we are left with a complex machine: the initiation of a potential motor act is begun by coalitions of competing subsystems, representing various action alternatives, and influenced by other subsystems (like sensory input). Within a Dennett-like scheme, conscious willing is the effect of the coalition of neurological subsystems that comes temporarily to dominate the brain. As one of the competing subsystem groupings

---

540, 541 (1985). Danto himself does not dispute Libet's findings, content to agree that "consciousness simply assures us we are in contact with ourselves." Id.

236 This is discussed infra notes 246-52 and accompanying text.

237 See Danto, supra note 235.

238 See Nelson, supra note 234.

239 See Wood, supra note 234.

240 See supra notes 104-23 and accompanying text.

241 The reader might find a quick review of part III.G helpful at this point.

242 Dennett briefly mentions Libet's experiment, and rejects Libet's conclusion that unconscious intentions "present themselves" to consciousness (in the Cartesian Theater?) and then are acted upon if not vetoed. See DENNETT, supra note 40, at 165. However, his alternative interpretation of Libet's data is less empirically supportable than Libet's own thesis. See id. at 164.
emerges, a course of action is "chosen"—subsystems interact, rejecting some drafts and keeping another. The "winning" course of action then "comes to consciousness" as that team of subsystems gains dominant status. Libet's measurements of cerebral activity preceding "subjective awareness" are thus accounted for by the underlying neurological subsystem competition, and what Libet calls the "veto" phenomenon is nothing more than the "awareness" that arises with the winner.

IV. ELIMINATIVE MATERIALISM, NEUROSCIENCE AND THE CRIMINAL LAW

A specific and plausible view of eliminative materialist claims has finally been built. Eliminativists' views are now not as wildly extreme as once suspected, and those views are supported by empirical findings unexplained by common-sense psychology. Now, since advocating abject surrender to hard determinism and scrapping the legal system is not, and never was, helpful, one must explore some possible legal responses to the fall of folk psychology and the findings of Libet. Part A will pursue the theoretical issue of how to rationalize assignments of blame without reference to folk-psychological theories, and Part B will suggest less drastic alterations of the current criminal justice system that might begin to accommodate Dennett's new model of consciousness and the findings of Libet. This Comment will then close with an exploration of the final theoretical obstacle to a post-folk-psychology legal framework and make some predictions about the future of criminal law as neuroscience continues to develop.

243 See the "question and answer" element of Dennett's model of consciousness, supra text accompanying note 211.

244 Dennett would certainly not dispute that there is a "mechanistic" explanation for all of human behavior, but he does believe we are incapable of consciously adopting determinism. See Daniel C. Dennett, Mechanism and Responsibility, in ESSAYS ON FREEDOM OF ACTION 157, 181 (Ted Honderich ed., 1973). The problem is highlighted by a commentator quoted by Dennett: "The motto of the mechanist ought to be: One cannot speak, therefore one must be silent." Id. (quoting N. Malcom).
A. Eliminative Materialism and Criminal Responsibility

There are several methods of reconciling the nonexistence of mental states and volitions with legal responsibility for individual behavior, but all of them require some degree of "cheating"—these methods either create a fiction on which to base responsibility (what lawyers would surely call "constructive" moral responsibility), or circumvent the difficulties presented by Libet and the eliminative materialists by changing the accepted description or rationale of the criminal law. Each approach is discussed in turn.

1. Fictionalizing Intentionality

As is by now obvious, the philosopher Daniel Dennett is no friend to folk-psychological constructs. Interestingly though, Dennett does not trumpet the cause for scrapping mental state psychology. After demolishing common-sensical notions of belief and proposing radical new hypotheses of brain function, Dennett retreats from the obvious implications of his theories to an instrumentalist position: he proposes that we maintain our current system of predicting behavior via reference to propositional attitudes, even though we now know it to be an inaccurate picture of cognition and behavior. This concession would allow us to preserve an easy method of assessing rationality, making behavioral predictions, and assigning responsibility for individual actions.245

Dennett envisions reformulating the folk psychology fiction into three "stances" from which to judge behavior: the physical stance, the design stance, and the intentional stance.246 Prediction from the physical stance involves reference to physical laws, such as "when enough water is poured into a cup, it will begin to flow over

245 See id. at 164-65. Dennett has compiled a book of his essays defending this position. See DENNETT, supra note 124. For a cogent summary of his main arguments, see Dennett, supra note 147, passim.

Dennett has drawn criticism from all quarters for his compromising nature; Baker devotes an entire chapter to criticizing Dennett's instrumentalist position. See, e.g., BAKER, supra note 194, at 149-66 (insisting that Dennett cannot maintain a materialist view if he asserts that intentional systems are a necessity); see also Paul M. Churchland, Folk Psychology and the Explanation of Human Behavior, in THE FUTURE OF FOLK PSYCHOLOGY, supra note 3, at 51, 65-67 (describing Dennett's intentional stance theory as "arbitrary protectionism" of folk psychology and likening Dennett to a deceptive Monty Python character); Horgan & Woodward, supra note 5, at 151 (criticizing Dennett for using folk-psychological systems but not accepting their veracity).

246 See Dennett, supra note 244, at 163-64.
the sides, because the physical laws of gravity will make it do so.” The design stance predicts behavior by referring to the mechanical nature of an object, such as “when this cursor reaches the right-hand margin, the word processor will automatically transfer it to the left-hand margin on the next line, because the program is designed that way.” Finally, there is the intentional stance, which predicts behavior based on a subjective conception of rationality, such as “when a chess computer is faced with situation X, it will move its king one space to its right because, to me, that would be the best strategic move.” We predict the moves of the computer by referring to what a reasonable move would be (to us), since the program, while “designed,” is too complex for a human to grasp and predict from the design stance. Since we cannot base our prediction on internal workings, the best way to guess is by reference to what we would do in the same situation. Human actions can be interpreted in the same way: we predict the behaviors of others by referring to what we envision as rational behavior, since the human being is a machine too complex for us to make accurate predictions of its behavior on the basis of our current neuroscientific knowledge. This intentional system operates regardless of whether the object in question actually has beliefs, desires, intentions, and so forth. It is a purely practically and observationally based tool.

The use of Dennett’s scheme as a fictionalized account of human behavior clarifies the position of certain groups in relation to the rest of society. Consider the insane: since they are not “rational,” society (and the law) should treat them from a design stance if enough is known of their mental disorder effectively to do so. If the neurological “design” of the insanity is known, that knowledge can be used to predict or perhaps guide the insane person’s behavior, without reference to any mental state language. Similarly, to manipulate another's behavior with the use of drugs would be to treat her from the physical stance, since drugs interfere with specific and physiologically set processes. No reference to mental states is required in that instance either.

247 See id. at 164.
248 See id.
249 See id. at 165. The idea here is to predict from the stance most useful in a particular situation. We attempt to predict the behavior of the mentally ill person from the design stance because her illness stems from a defect in her neurophysiological mechanics, or design.
Dennett's fiction saves the criminal law numerous contortions. It allows that for purposes of assigning blame, society is simply employing a fictional model of humans as beings that entertain propositional attitudes and perform intentional acts. We employ this model of rationality because it is more easily grasped than the actual complex workings of the human "machine," not because we genuinely believe that humans have such attitudes and intentions. Since most humans operate within Dennett's intentional system, that is, they are rational, Dennett's instrumentalism allows the criminal law to remain practically untouched, since it rests on similar presumptions. One could set varying degrees of blame based on whichever "stance" dictated the actor's behavior. The physical stance would imply that the individual's actions were compelled biologically, as with drugs, and that she is therefore not responsible for them. The design stance would imply that the actor's actions were caused by the manipulations of another, such as in situations of necessity or duress, or were the result of submerged neurophysiological processes, such as the efficiency protocol discussed below. Finally, the intentional stance would apply to the standard legal person—the mostly practical reasoners at whom the law is aimed.

Adoption of Dennett's or a similar intentional fiction is tantamount to conceding defeat in the effort to shape legal responsibility around a biological science that is leading in a direction society did not foresee. Dennett asserts that the intentional stance is vital to living in the real world, even though it mimics incorrect psychological assumptions. This may or may not be so, but regardless of that claim, this position in the criminal law amounts to purposeful ignorance of scientific change. Under such a scenario the law is maintaining its own view of human behavior while scientists fill in an increasingly different reality. Professor Moore notes that legal theorists today are prone to accept "as if" views like those of Dennett, but Moore himself is not willing to accept this gap between scientific reality and assignments of blame:

---

250 Note that this system leaves the door open to legal personhood for artificial intelligences.

251 See infra notes 296-304 and accompanying text.

252 See DENNETT, supra note 124, at 108. As Dennett points out, "Try catching frogs without it." Id.
[Lawyers] are apt to be apologetic in the face of deterministic science. To justify our criminal laws, lawyers talk of positing free human actions, even though they admit that scientifically there is no such freedom.

This kind of reconciliation by fiat cannot possibly work. The law demands more than that we pretend people are free and thus hold them responsible as if they were. A just legal system requires people to be truly responsible. Like alchemy, the law may be rendered obsolete by the superior, if nascent, successor of scientific discovery.

2. Redescribing the Law

In contrast to fictionalizing intentionality, recharacterizing the law to avoid reliance on mental state assessments allows us to avoid accusations of stubborn intransigence in the face of progressive science. Moreover, through reappraisals of long-standing legal doctrines, we might be able to maintain much of the current legal system. Professor Kevin Saunders, for example, has concocted an ingenious method of avoiding the old mind-body problem in questions of volition and the criminal law. His proposal is not directly applicable to our endeavor because the problem he identifies only exists if one espouses dualism; nevertheless, his method of redefinition is a good model for changes we might make. Encountering unsolvable philosophical/scientific difficulties in attempting to show that nonphysical mind states are the cause of physical actions, Saunders completely avoids the problem by describing the mind as an entity that "solicits" the body. According to Saunders, even if the nonphysical mind cannot directly cause physical acts, it can be held culpable for attempting to solicit the body to perform criminal actions. The Model Penal Code can be read to support Saunders's proposal:

(1) Definition of Solicitation. A person is guilty of solicitation to commit a crime if with the purpose of promoting or facilitating its commission he commands, encourages or requests another person to

---

253 Moore, supra note 36, at 1121-22.
254 See Saunders, supra note 100, at 467. The problem of the Law of Conservation of Energy is a good example. See supra note 100 and accompanying text. Materialists avoid this problem because for them there is no separate mind and brain: the "mind" is a biological manifestation, not a spiritual one. The phenomenon of physical processes causing other physical processes is obviously not problematic.
engage in specific conduct which would constitute such crime or an attempt to commit such crime or which would establish his complicity in its commission or attempted commission.\footnote{255}

Thus, if we view the mind as merely in contact with the body, minds, and therefore people, can be held liable for physical acts, since the culpable mental state need only command or request action, not actually cause it.\footnote{256}

Saunders's idea is clever, but our task is far more daunting. We must fashion a systemic legal redescription which completely removes the criminal law's manifestations of mental state psychology while avoiding a major upheaval. Two existing legal theories attempt to perform this trick: the "character" theory of responsibility and economics-based theories of human behavior. None are entirely successful, but they may guide the way for more sophisticated attempts at reconciling eliminative materialism and our legal necessities.

a. The Character Theory of Responsibility

\textit{"[Y]ou are being condemned, Robespierre is told, 'because you lack grace."}\footnote{257}

The character theory is generally credited to David Hume, who once stated that "actions are objects of our moral sentiment so far only as they are indications of the internal character."\footnote{258} Under the related legal view, an actor can only be held responsible for her acts if those acts are expressive of her underlying character. Personal character becomes the primary focus of responsibility, and the behaviors an actor exhibits are considered proxies for that character, since they are the only observable indications of the kind of character a person possesses.\footnote{259} While the character theory

\footnote{255 \textsc{Model Penal Code} § 5.02(1), quoted in Saunders, \textit{supra} note 100, at 469 (emphases added by Saunders).}

\footnote{256 See Saunders, \textit{supra} note 100, at 469-75.}

\footnote{257 Michael S. Moore, \textit{Choice, Character, and Excuse}, \textsc{Soc. Phil. \\& Pol'y}, Spring 1990, at 29, 55.}

\footnote{258 \textit{David Hume, An Enquiry Concerning Human Understanding} 108 (Charles W. Hendel ed., The Liberal Arts Press 1955) (1748).}

\footnote{259 See Moore, \textit{supra} note 257, at 29. Moore compares the character theory of excuse to the more widely accepted choice theory (which asserts that actors are responsible for actions they had the capacity or opportunity to choose to avoid) and ultimately rejects the former. \textit{See id.}}
appears to base responsibility on a person's actions, it actually assigns blame on the basis of the underlying bad character traits evidenced by those actions. Moore, within the context of criminal excuse, explains:

We are *prima facie* morally responsible and legally liable for wrongful and illegal acts, but we are only *prima facie* responsible and liable because we do not yet know whether such wrongful acts were truly expressive of our characters. If they are, we are actually responsible and liable; if they are not, we are neither. The excuses, in such a case, serve to filter out wrongful actions for which we are not responsible because they are not expressive of our character. In such a way the excuses reconcile the ultimate nature of moral responsibility—it is for character, not action—with the surface features of our moral and legal ascriptions: namely, that they are for actions and not character.²⁶⁰

A simple example involves a crime performed while under the influence of drugs. If Len is a law-abiding citizen, and is involuntarily injected with a drug that transforms his normally calm personality into that of a homicidal sociopath, it seems intuitively correct not to hold Len responsible for killing Thom while in that state. It seems as if "Len," as in Len's character, did not commit the crime. In other words, we do not punish Len for the act of killing Thom because the act was not at all expressive of Len's true character.²⁶¹ By focusing only on character we are able to avoid relying on mental state suppositions.

The theory applies equally well in the case of provocation. Instead of determining whether a person's "will" was "overborne" by her emotions, courts view *prima facie* bad acts as vital clues to the nature of the defendant's character. If the neurological dispositions provoking these acts are "good" dispositions, in that they fit

²⁶⁰ *Id.* at 41.

²⁶¹ Presumably the relationship between action and character in a particular case would be defined at trial, but first the character theorist must elucidate which aspects of the person count as character traits and which do not. As Moore notes, the distinction is far from obvious. For example, does an actor's character include only those traits she identifies as constituting her personality, or does it include, as Freud might assert, aspects of her "subconscious?" Or should character perhaps be defined only as the sum of a person's past behaviors? *See id.* at 43. Note also the tempting trap of fleshing out a person's character traits by assessing what that person believes. Doing so leads us back to supposing propositional attitudes, a folk-psychological pitfall. This "character constitution" problem is not insurmountable, but a careful delineation of the accepted boundaries of "character" would be an essential part of a refined character theory of criminal responsibility.
some socially set definition of goodness, the actor is determined to have good character and is acquitted; if the tendencies motivating action are "bad" ones (the neurophysiological replacements for what we call hatred, greed, or whatever), this indicates bad character, and the actor is held responsible. This scenario is actually similar to current legal applications of the provocation defense, under which we seem to judge the quality of emotions motivating provoked acts. If the emotion of blind, envious hatred, for example, compels you to kill, no court hesitates to convict you of murder. But if that emotion is grief, righteous indignation, or some other emotion society finds morally acceptable, you might be guilty of only manslaughter. Other defense doctrines lend themselves to reconstruction within a character framework: For example, in cases of necessity and duress, a defendant would be permitted to escape responsibility for her acts when those acts are not indicative of her personal character but instead are induced by external forces that overcome her naturally good character traits.

The character theory is thus able to assess blameworthiness without relying on any psychological pretensions. The quality of the actor's character is presumably surmised through some kind of fact-finding tribunal, and the act performed is determined to instantiate that character or not. No mental state suppositions are required to determine the defendant's culpability. All analyses are based on observable conduct and its connection to whatever internal and behavioral components are found to constitute personal character.

Blaming persons only for evidencing bad character does create some difficulties. An initial theoretical puzzle is whether any of us are truly responsible for our characters. The character theory requires that all persons be held accountable for the kinds of persons they are—for their traits of kindness, hatred, honesty, diligence, and a host of others. Moore does not find this problematic, but recent literature supporting the exoneration of criminal defendants on the basis of a harsh sociological background implies otherwise. For example, if courts acquit defendants who grow up in the inner city without parents because they do not enjoy the internal moral constraints bestowed by safer neighbor-

---

262 See id. at 45 ("I think it is obvious that, in some sense, we are responsible for being the kind of people that we are . . . ").

263 See, e.g., David L. Bazelon, The Morality of the Criminal Law, 49 S. CAL. L. REV. 385, 396, 401-03 (1976) (arguing that social and economic background should be considered as mitigating factors in assessing criminal responsibility).
hoods and strong parental figures, is not their excusal based on character aspects for which they are not held responsible? Proponents of acquittals in these instances would apparently not agree that all persons are responsible for their character traits.

A more startling implication of the character theory is that persons may logically be punished for bad character even in the absence of any act. If character is the core of responsibility, and physical acts are only proxies for that core, then why not directly punish people for having bad character, instead of waiting for the actions that express it? The character theory seems to commit its supporters to this view, and as Moore points out, no one deserves punishment merely for being "a poor specimen of humanity."

Finally, the character theory cannot distinguish between certain morally relevant factors. For example, the doctrine of mistake-in-fact might easily lend itself to reconstruction as an aspect of character-based legal theory, but how can we (or a court) discern between an actor who makes unreasonable mistakes because of a complete indifference to the concerns of the criminal law, and the actor who commits unreasonable mistakes because he is careless, but generally a nice guy? The former individual possesses bad character (a lack of appreciation for the rule of law), and the latter person generally embodies good character, despite the undesirable but innocuous trait of carelessness. The determination might be made easier by a neuroscientific analysis of the actor's brain states, but barring that kind of empirical assessment, making the distinction is a problem.

Moore would presumably argue that this person was simply "morally unlucky," due to her disadvantaged background, and still responsible for her character and the acts to which it led. See Moore, supra note 257, at 46.

See id. at 55. The reader may note that without mental state suppositions and relying on the character theory as stated here, the only way to determine if an individual has bad character is by analyzing actual acts she has performed, and so a determination of bad character cannot even be made without waiting for some illegal action on the part of the person. The answer to this seems to be that an assessment of bad character need not depend on performance of illegal acts, per se. Many behaviors are considered immoral, but not illegal. A sufficient record of such behaviors might cause a societal determination of bad character in the individual, and lead to incarceration without relying on a specifically illegal act.

In addition, advanced science may be able to track neurological and genetic dispositions toward good and bad behavior, making the assessment of character less dependent on an actual act.
Despite these objections, the character theory has potential. With some tinkering and perhaps a bit of Dennett’s intentional fiction to fill certain gaps, it could be successfully employed to put the old wine of the criminal law into a new and mental-state-free bottle. The serious theorist of character-based legal responsibility would have to define the accepted boundaries of personal character for legal purposes, and specify what neurological activity instantiates legally punishable character. She would also have to import an act requirement into the theory, to avoid punishing persons merely for not reaching some required level of personal integrity.

b. Economics-Based Theories of Behavior

Another redescription that would obviate reliance on mental state psychology is one drawn from economic theory. On such a view, what a person “intends” to do is equated with what it is “rational” for that person to do. The legal description is shifted from an assessment of internal states to an external, almost behavioral account. Actions, including crimes, are performed to maximize resources. By following this axiom, the law can distribute punishments and rewards in such a manner as to deter behaviors it deems undesirable—if the known costs of an act are high enough, a resource-conscious populace will be discouraged from performing it. No reference to “mental states” is required.

The above thesis essentially mimics the position of Judge Posner, whose economics-based legal view was detailed more fully earlier in this Comment. It also suffers from the same criticisms previously leveled at Posner’s account, most importantly, that Posner and other economic theorists do not really avoid mental state attributions. Even if we allow the economic theorists to attribute a minimalist form of “rationality” to actors, in that an average actor will behave predictably—for whatever reason—according to certain standards, economics-based legal theory still relies on propositional attitudes like beliefs and desires. When Posner asserts, for example, that tort liability rules will influence future behavior, he does not explain exactly how that is possi-

268 See supra notes 54-57 and accompanying text.
ble. Unless Posner envisions some brute, external force dictating the actions of persons' bodies, he implicitly relies on beliefs and desires. Legal rules presumably influence future behavior because the potential actor believes there will be certain repercussions following certain behaviors, and so does or does not perform a certain act, depending on whether the actor fears those repercussions will damage her resources. Because economics-based legal theory requires a risk-averse populace deterred by the threat of resource depletion, typical economic redescriptions cannot operate without some folk-psychological support.

The similarity between economic theory and behaviorism, however, guides the way to avoiding the cognitive trap ensnaring Posner and other economic theorists. Both rely on deterrence—the calculated dispensation of punishment to prevent the occurrence of undesirable acts. Posner is unsuccessful in his attempt to explain the law without mental state attributions because he relies on a concept of deterrence entailing beliefs and fears. What he does not mention is that behaviorism shows that deterrence requires nothing of the kind. B.F. Skinner elicited complex actions from rats and pigeons through a refined regimen of rewards and punishments, and few would presume that those creatures possess beliefs and fears as we understand them. Skinner was able to condition his animals to behave in a certain manner at specified times through a method of deterrence that makes no pretensions of mental state attribution. He simply rewarded or punished them for performing (or not performing) certain acts at certain times. This conception of deterrence might lend itself to a semi-utilitarian design of the law that avoids folk-psychological elements. This design is discussed below.

c. Utilitarianism and Avoidance of Mental States

Once mental states are removed from the economic theory of law, that theory can no longer preserve the law as it currently stands, and it leads us to a partially utilitarian rationale for avoiding individual psychology. If that step is taken, the law can then progress indefinitely, unmolested by neuroscientific intrusions. The new underlying legal framework would generate a legal system unrelated to explanations of human cognition and volition.

of the civil negligence system).

270 See GLEITMAN, supra note 60, at 80.
We arrive at a utilitarian answer because any theory of punishment that might allow society to avoid psychological assessments cannot entail moral judgments of individual actions as a factor in distributions of blame. As mentioned earlier, any system of laws that purports to assign blame on the basis of moral responsibility inherently relies on some view of psychology, and here we are attempting to skirt that reliance. This eliminates most nonconsequentialist theories of punishment and leaves us to mold a utilitarian explanation for society's new criminal justice system.271

The classical premise of utilitarian legal theory is that "the purpose of all laws is to maximize the net happiness of society,"272 and happiness entails minimizing the occurrence of acts that damage the community. The method of happiness maximization derives economic legal theories—punishment for certain acts is dispensed to persons in the hope that (a) they will be conditioned to avoid a particular behavior in the future, and (b) other persons will also be deterred from the same conduct by fear of suffering whatever befell those who committed the illegal acts.273 For our purposes, the standard utilitarian theme must be slightly altered to remove the second prong of the rationale. The deterrence of Thom by punishment of Len relies unavoidably on the mental state of fear. Even Skinner could not affect the behavior of one rat by punishing another.

One can envision a legal system concerned only with the greater good of the community, defined as general social order, an absence of needless violence, and whatever other factors are considered important. Punishment, as conditioning against future illegal behavior, is also designed to remove the bad actor from the society she has damaged. This segregation ensures that the bad actor does not cause further harm. A correlation exists between the amount of punishment dispensed (probably confinement) and the amount

---

271 The economic redescriptions of law discussed in the previous section actually come close to presenting a utilitarian rationale. Under the economic rubric, behavior is thought of as guided by what goods are to be gained, instead of by moral injunction. The difference between the utilitarian design discussed here and the economic theory of behavior is that the former envisions elimination of all nonutilitarian aspects, since those tend to draw mental state suppositions into the picture. As noted earlier, the standard economic explanation of behavior relies on the beliefs and fears necessary for proper deterrence.

272 DRESSLER, supra note 1, at 4 (citing Jeremy Bentham as the principal proponent of this idea).

273 Dressler calls these goals "specific" and "general" deterrence, respectively. See id. at 5.
of quantifiable damage society has suffered as a result of an individual’s behavior. The more severe the crime, the longer the actor is segregated from the community.

This system avoids mental state attributions by following a Skinnerian conception of deterrence and by measuring punishment only in relation to damage done. The sole issue for a factfinder is the amount of disutility inflicted on the whole, not any morally culpable state in the individual. This shift greatly changes the law as it currently operates, since no internal motivations or states act as mitigating factors—defenses such as provocation and insanity are not available, and there are no varying degrees of punishment corresponding to the mental state of the actor at the time of the offense. Confinement is not for rehabilitative purposes as currently conceived, since that requires folk psychology, but it does condition the behavior of the human animal and prevent further damage to society. Other benefits are gained if neuroscience allows for the precise neurophysiological alteration of inmates to prevent future illegal conduct; since appeal to beliefs and desires is impossible, the bad actor might be chemically altered to remove the threat of further harm. If manipulations of an actor’s neuro-anatomical design are ineffective or impossible, this utilitarian-type legal system could countenance life-long confinement, constant drug-induced subservience, or the death penalty in an effort to avoid further harm to the community.

Such a system seems the least desirable of alternatives. Unlike the intentional stance, the character theory, and supposedly the economic theory, it does not leave the current legal system mostly intact. That goal is pursued in an effort to avoid drastic legal changes and their attendant social upheavals. Further, instead of substituting defunct folk psychology with a new vision of behavior grounded in neurophysiology, the new utilitarian alternative avoids addressing the internal workings of individuals altogether. Eliminative materialists would not support adoption of this system even though it avoids folk psychology, since it also avoids the neuroscience that those philosophers believe will lead to more just

Note that under such a system defenses like necessity and duress might be available, since if an actor commits what would be a criminal act but in so doing prevents a worse damage to society from occurring, the net effect is that he has benefitted society. Also, in such a case the force or person causing the actor to act out of necessity or under duress would be held responsible for the harm inflicted by the actor, since it is she, rather than the actor, who is considered to have actually caused the harm.
An accurate explanation of human behavior would ensure more accurate assignments of responsibility, so completely ignoring the details of physiology is not a recommended solution.

None of the suggestions in this section are problem-free. Dennett's conception of the intentional stance is a fiction of intentionality preserved for the criminal law. The character theory of responsibility is the most useful of the redescriptions suggested, but it has major difficulties that must be resolved. The economic theory is viable if it is influenced by behavioral theory, but by itself it cannot fulfill Posner's claim that an economic recharacterization of the law can both avoid mental state attributions and leave the current law intact. Finally, the utilitarian approach too greatly changes current legal mechanics and avoids all cognitive theory to the detriment of society.

The most desirable alternative might actually be to leave intact, for legal purposes, some degree of intentionality (albeit fictional). From that basis society might seek to partially reconcile the law with profound changes in cognitive science, without discarding thousands of years of legal evolution. In the spirit of reconciliation, the next section reduces the discussion to a more specific level; assuming that the current law already relies on a fictional system of intentionality, it explores potential changes based on recent neurophysiological theorizing—that of Libet and Dennett.

B. Specific Legal Accommodations of Libet and Dennett

The theories of Libet and the eliminative materialists strongly imply that the folk-psychological conception of volition—a process of propositional attitudes leading to intentions and thus physical action—is not accurate. However, despite Libet's findings and the proposed theoretical revisions above, it simply may not be possible to excise notions of personal choice and intentionality as we currently employ them. A fictional account is the easiest alternative. Indeed, philosophers such as Peter Strawson maintain that this is our only alternative. They insist that metaphysical debates over issues such as intentionality, determinism, and free will are of no consequence, since free will is a fundamental element of human

275 Cf. CHURCHLAND, supra note 9, at 45 ("The simple increase in mutual understanding that [eliminative materialism] made possible could contribute substantially toward a more peaceful and humane society.").
nature and we are unable even to contemplate its negation.\textsuperscript{276} For Strawson, the reactions and judgments we make in response to others' actions are so central to our existence that they "neither call[] for, nor permit" rational explanation, and thus commit us to an acceptance of free choice regardless of theoretical disputes.\textsuperscript{277} No matter what one thinks of the fundamental status of free will, legal reliance on a doctrine of intentionality is deeply imbedded and may minimally persist regardless of the arguments mounted by eliminativists and neurophysiologists and whatever clever legal reconstructions are proposed. In light of this entrenched position, perhaps some compromises can be reached between traditional legal doctrines and the eliminativist picture of human behavior that will reflect realistic possibilities for a legal accommodation of eliminativist conclusions. None of these alterations will approach the magnitude of those discussed in the previous section, but they may at least turn the criminal law in the direction that cognitive science appears to be going.

The current law, faithful to folk psychology, holds an individual responsible "only for those consequences that are caused by his actions, and not for those things in which his body, but not his acting self, is causally implicated."\textsuperscript{278} If we accept Libet's findings alone, without the backdrop of Dennett's model, we are left with beliefs and desires leading directly to volitions, without the presence of an intention until after the volition is formed and about 250 milliseconds from execution.\textsuperscript{279} With the combination of Libet and Dennett, Libet's findings become a bit more palatable to the conceptually queasy, but we are still left not quite as "intentional" as we supposed ourselves to be. That description is apt in light of Dennett's intentionality fiction—humans are complex machines to which we attribute an "intentional stance," but as we parse the neuro-anatomical workings of voluntary motions, we uncover a design stance, and this is understandably disturbing. Human action

\textsuperscript{276} See P.F. Strawson, Freedom and Resentment, in FREEDOM AND RESENTMENT AND OTHER ESSAYS 1, 3-4 (1974) (arguing that even if determinism is accepted, there will still exist some degree of free will in human behavior).
\textsuperscript{277} Id. at 23.
\textsuperscript{278} Moore, supra note 64, at 1567.
\textsuperscript{279} See supra note 231 and accompanying text. Saunders's discussion of volition and criminal acts roughly hypothesizes the findings of Libet: "If volitions cause acts, the mind is the causal agent. If volitions do not cause acts, the mind might be seen as approving the act but not involved and so, perhaps, not culpable." Saunders, supra note 100, at 454 n.47.
can now be explained as a mechanical occurrence instead of the result of conscious decisionmaking. Dennett realizes this in another context:

Wholesale abandonment of the Intentional is in any case a less pressing concern than partial erosion of the Intentional domain, an eventuality against which there are no conceptual guarantees at all. If the growing area of success in mechanistic explanation of human behavior does not in itself rob us of responsibility, it does make it more pragmatic, more effective or efficient, for people on occasion to adopt less than the Intentional stance toward others. Until fairly recently the only well-known generally effective method of getting people to do what you wanted them to was to treat them as persons. One might threaten, torture, trick, misinform, bribe them, but at least these were forms of control and coercion that appealed to or exploited man's rationality. . . . The advent of brainwashing, subliminal advertising, hypnotism and . . . the more direct physical tampering with drugs and surgical intervention, for the first time make the choice of stance a genuine one.

Dennett wrote this passage twenty years ago, foreshadowing the neuroscience that is now eroding our “Intentional domain” and its legal inhabitants.

Saunders notes: “If the mind, through its volitions, causes the body to act, the mind is culpable for those acts. If the mind is not so directly involved, allowing the mind to play such an active role in the attribution of actus reus is on less firm ground.” No longer secured, we have lost confidence in the fundamental truth of folk psychology, and yet we have a neuroscience too youthful to provide an adequately detailed substitute. Folk psychology appears mostly wrong, and the true workings of the brain mostly unknown, but the need for legal coherence remains. Shaken in our ability to assess intentionality and now relying on a fictional account, can we morally punish “intentional” actors more severely than others? Can we justifiably condemn that which we can no longer fully explain?

280 Dennett, supra note 244, at 182. An interesting consideration: If neuroscience were to enable us to cleanly eliminate any person's desire to kill, should we make use of this power? As Dennett rhetorically asks, “What if mass hypnosis could make people stop wanting to smoke?” Id.

281 Saunders, supra note 100, at 461. Note Saunders's fully dualist stance. He discusses the “mind” as something distinct from the “body.”

282 There are critiques of Libet's work, for example, that are far more damaging to notions of intentionality than anything Libet himself implied, and they further confuse our definition of intention. Building on William James's interpretation of volitions as “anticipatory images” of actions, one commentary gives a different
1. Eroding the Intentional

The Model Penal Code has this to say about intentional actions:

A person acts purposely with respect to a material element of an offense when:

(i) if the element involves the nature of his conduct or a result thereof, it is his conscious object to engage in conduct of that nature or to cause such a result . . . . 283

It would appear that this definition requires some revision, since we now must doubt whether it can ever be our “conscious object” to do particular acts in the manner intended by the Code.284 If our conduct is unconsciously initiated and then consummated if not stopped within a narrow window of opportunity, acting “intentionally” loses the ring of culpability it once possessed. We may not need to eliminate the intentional concept completely, but we are in a precarious position.

account of Libet's discovery:

An unconsciously generated action initiative (signaled by the RP) brings about a corollary discharge image of the action in consciousness at a somewhat later time, which in turn provokes a secondary affective “urge” to match motor behavior to the image. The image and urge together constitute [a volition] . . . . According to this model, however, [a volition] is a conscious correlate of feedforward regulatory processes rather than an intentional decision by a human agent. . . . To claim that [a volition] causes the action would be erroneous in the same sense that it would be incorrect to assert that activation of a radar system causes the airplane to take off just because the pilot generally turns on the radar, for guidance purposes, prior to takeoff.

Ralph E. Hoffman & Richard E. Kravitz, Feedforward Action Regulation and the Experience of Will, 10 Behavioral & Brain Sci. 782, 783 (1987). In other words, Hoffman and Kravitz argue that the volition is not a causal factor at all—it merely emerges from cerebral processes dictating movement and prompts the person to match his actual behavior to the “suggestion” of his unconscious. It has no power, "veto" or otherwise.

283 MODEL PENAL CODE § 2.02(2)(a)(i).

284 Interestingly, the Model Code's psychological view was criticized ten years ago by Professor Harold Edgar:

[T]he psychological model of conduct underlying the Code, the idea that we act by willing discrete behaviors, reflects ancient psychology. Surely it is a fiction; no homunculus lives inside the brain, sending out orders to animate the limbs. Although the law can live happily with fictions, the acceptance of fictions usually marks issues the law seeks to avoid, not ones it confronts.

Harold Edgar, Mens Rea, in Encyclopedia of Crime and Justice 1028, 1038 (Sanford H. Kadish ed., 1983). In part IV.A.1 I will consider the use of legal fictions in dealing with eliminative difficulties.
One option is to expand and weaken the definition of intentionality. Since inaccessible brain events are involved, we could define an intention as not just its narrow conscious component, but as the entire set of neurological processes that culminate in a physical act, thus deemphasizing the conscious elements relevant to our current conception of intent. As a folk-psychological artifact, "intentions" were developed long before anyone had any knowledge of neurological systems, and a refinement of the concept is probably overdue. From the legal standpoint, this change would lessen the moral significance of an act "intentionally" performed, and the hierarchy of punishment would reflect this alteration. For example, the Model Penal Code's definitions of "murder," "manslaughter," and "negligent homicide" would be retained in order to distinguish between homicides with and without conscious components, but the retribution due an "intentional" killer would not greatly outweigh that due a reckless or negligent one. Given our new knowledge, the division of murder into degrees would be even more questionable, since imposition of the death penalty would partially depend on now suspect mentalistic distinctions.

The general point here is to revise the law to reflect cognizance of our biomechanical natures. The decision, however made, to pull the trigger or use the knife cannot be as morally charged as the law currently implies. That decision involves a good dose of passivity, since the consciousness component, far from choosing the action, may only monitor or obstruct volitional outcomes.

Assessments of culpability might also benefit from a limited character-type analysis of responsibility, instead of relying solely on the dubious ontology of intention. Even without the advantage of clearly defined intentionality, the law must still differentiate between the killer who planned the act for three weeks and the one who committed the act spontaneously, or between the actor who killed out of anger and the one who killed accidentally. Since intention can no longer provide a meaningful distinction, analysis of underlying character may fill the resulting gap in attributions of blame.

286 Id. § 210.3.
287 Id. § 210.4.
288 See, e.g., Cal. Penal Code § 190 (West 1988) (designating death penalty or imprisonment for a minimum of 25 years to life for first degree murder but only imprisonment for a minimum of 15 years to life for second degree murder).
289 See supra notes 257-66 and accompanying text.
2. Rethinking Our Defenses

a. Provocation

Our revised view of intentions is also applicable to the area of criminal defenses, although the general theme of "loosening" the legal significance of "intentional" action still applies. Consider the provocation defense, which reduces the level of culpability for a killing from murder to manslaughter. Though some courts have implied otherwise, provoked killings are intentional ones, and our new model of the brain further informs the "provocation" phenomenon. Consider the classic example: Thom stumbles upon Len engaged in sexual intercourse with Thom's wife. Thom, duly enraged, kills Len. Thom will most likely be convicted of manslaughter, having claimed provocation as a mitigating factor. The rationale for that defense might be that Thom succumbed to the human weaknesses that occasionally catch us all, so he is not a "bad character" in need of the most severe social sanction. Or, it might be argued that Thom's deep emotions overrode his ability to respect legal and moral mandates, so he could not control his conduct. What we can now say is that a particularly powerful coalition of subsystems, driven by input from sensory subsystems, "swept into power" and dictated the actions of the body that is Thom. The strength of this grouping of systems arose from both the intensity of the sensory input and a lack of competition from other systems. There may have been only one "possible" course of action for Thom; if there was no struggle among coalitions of neuronal systems for satisfaction, no cerebrally proposed action alternatives besides beating Len to death, did that "unanimous" set of neuronal structures create a force strong enough to counter Thom's remaining brain systems? Perhaps those brain systems,

---

290 Professor Dressier cites Holmes v. Director of Public Prosecutions, [1946] 2 All E.R. 124 (Eng.), as an example of a court finding a lack of specific intent to kill in a case involving a provocation defense. See DRESSLER, supra note 1, at 475 n.13.

291 See id. at 475 ("The provoked killer acts due to anger, not evilness.").

292 See id. at 476 (noting that "deep emotion . . . undermines [the] capacity to respond in a legally and morally appropriate fashion").

293 Perhaps a more illuminating way of phrasing this is to say that Thom could not at that moment "conceive" of doing anything else. He could not "conceive" of another alternative because all parts of his brain were in an "agreement" of sorts.

294 In Consciousness Explained, Dennett does not address the idea of variable power among groups of subsystems depending on their "size" or "support." See Dennett, supra note 40, at 210. If our new schematic for volition is beginning to sound remarkably similar to a democratic election, note that Dennett himself makes use of
or parts of several systems, involved in the storage of moral and social constraints?

This explanation of provocation makes a certain amount of intuitive sense. Couched in traditional terms, a rising tide of mental [neuronal] protestation [subsystems] makes Thom kill Len before he can stop [veto] himself. Put differently, our explanation sounds remarkably like saying, in the language of folk psychology, that Thom’s emotions overcame his capacities as a practical reasoner. We must also remember one important facet of our new brain model: the powerful coalition of subsystems calling for Len’s demise does not “intrude” upon an already existing awareness. Under Dennett’s model, this coalition becomes Thom’s awareness, since consciousness is an outgrowth of series of coalitions dominating in succession—any competition between coalitions for dominance occurs in a preconscious stage. We could even cast Thom as an automaton of sorts, operating purely from the “design” stance, following the instructions of a temporary but overwhelming neuronal program.

Our re-explanation of the provocation defense may warrant a liberalization of its use, depending on what more we learn about the neurophysiological workings involved. This is not a problem, if one assumes that neurophysiological explanations can only lead to a more just assessment of the merits of this defense in a particular case.

b. Force of Habit

To fall into a habit is to begin to cease to be.

Renewed talk of the unconscious might resurrect consideration of a defense based on motions performed habitually. Should there be a “habit defense,” exonerating those who perform criminal acts by way of habitual motions? The Model Penal Code gives a pretty


295 See, e.g., Addington v. United States, 165 U.S. 184, 186 (1897) (allowing mitigation from murder to manslaughter when the provocation renders "any ordinarily prudent person for the time being incapable of that cool reflection that otherwise makes it murder"); Moore, supra note 36, at 1148 ("[E]xcuses are all related to the exercise of the actor's practical reasoning capacities.").

solid "no" to this question, specifically removing habitual movements from its definition of acts not voluntarily performed. Should the Code be revised? It has been said that "[o]ne whose body moves, while the mind is totally engrossed in thought, appears to be lacking volition." Perhaps we can restate this dualist explanation within the framework of our new brain-conception: the brain, hard-wired to prioritize for efficiency, removes some activities from current "awareness" to allow consciousness to pursue other, higher tasks. These activities, surely composed by smaller neuronal subsystem groupings, may work "in the background," like the clock functions of some computers, and never become part of the series of subsystem coalitions constituting awareness. Certain operations are "automatized," like tying shoelaces. As Moore points out, this is a desirable mechanism: "At the level of consciousness, I do not pay attention to most . . . details of motor movement. And it is a good thing, for if I had to focus on each [detail of motor functions], I'd never be able to focus on my larger goals, muse about philosophy, etc." Consciousness can do just so many things at once. Certain actions in certain situations will recede to subconsciousness as we concentrate on larger, or perhaps simply other, tasks.

---

297 See Model Penal Code § 2.01(2)(d) (describing as a nonvoluntary act "a bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious or habitual") (emphasis added). Commentary on the Code does not explicate the drafters' decision to include "habit" as a voluntary act, aside from stating that the provision is "designed to make the requirement of an act a minimal one." Id. cmt. 2.

298 Saunders, supra note 100, at 460 n.70.

299 For a discussion of intelligence as a product of cerebral efficiency, see supra note 122.

300 Gleitman, supra note 60, at 298.

301 Moore, supra note 29. Moore gives the example of learning to play the piano, where the specific strikings of each key are eventually not consciously performed, but are weaved into greater movement patterns that are consciously executed. See id. Dennett has noted: "Most of our intentional actions are performed without any [deliberation], and a good thing, too, since there wouldn't be time." Dennett, supra note 40, at 252.

302 Arthur Danto notes:

Choosings between alternative courses of action, in the preponderance of motor acts we perform, occur as the outcome of deliberations of which we are barely conscious, if at all. A slow-motion film of Matisse shows the artist making countless decisions with his fingers that at normal speeds look like a single confident chalk stroke defining the edge of a leaf. He may or may not have been conscious of each decision, but I suspect that he was conscious only of drawing a leaf.
The Model Penal Code does note that "bodily movement[s] during unconsciousness" are not voluntary acts. Is this consistent with the now-expanded role of nonconscious activity? Our efficiency proposal implies that habitual activity is not conscious at all. The Code Commentary reveals that the term "voluntary" is meant to encompass "conduct that is within the control of the actor." Libet has shown us that even "voluntary" movements may not be initiated consciously. If well-worn neural pathways in the brain predispose the body to a certain repetitive movement without the need for consciousness (since perhaps "awareness" of the movement would decrease efficiency), can we say the movement is voluntary? If repetition has caused a background set of neuron structures to operate without needing to be a member of the neuron group currently "in power," can we hold the person responsible for her structural tendency toward operational efficiency?

These questions would arise in only a narrow set of cases, and the legal alterations might not be major. Driving a car provides an example with which most of us can identify: You are commuting to work and become so absorbed in your thinking that after you arrive at your destination it occurs to you that you remember absolutely nothing of the trip or anything you saw during it. Or, "lost in thought," you drive to an oft-visited location such as your place of work instead of to the grocery store, which is where you originally meant to go. In neither of these instances were you conscious of driving the car. A repetitive and well-practiced behavior, automobile operation drops from consciousness and operates in the background, while awareness is taken over by neuronal subsystems devoted to whatever cognitive endeavor preferred. The driving process can be so completely programmed that if not consciously guided, the "default setting" will bring the driver to a well-known destination, whether intended for this trip or not. The driver is "conscious" of Beethoven's Ninth or his plans for the next day, but

---

Danto, supra note 235, at 541. For an extensive, if somewhat dated, account of volitions that also supports this conclusion, see Gregory A. Kimble & Lawrence C. Perlmuter, The Problem of Volition, 77 PSYCHO. REV. 361, 373-81 (1970) ("Highly practiced acts tend to recede from consciousness, to become routinized and automatic and, in that sense, involuntary.").

303 Model Penal Code § 2.01(2)(b).
304 Id. § 2.01 cmt. 1.
not "conscious" of acceleration, deceleration, or whatever is in the road ahead.

Now say our particular driver, Thom, "daydreaming," runs over Len quite "unintentionally." It is light out and Thom is sober, driving on a straight patch of road with his eyeballs positioned to accept input from directly ahead of the car. Thom, however, does not "see" Len walk into the street, and as a result runs him over. We will avoid the "act/omission" question by asserting that Thom was steadily accelerating when he hit Len. Under current law, the case for Thom "knowingly" running down Len could be plausibly made, since from the viewpoint of a third party it would appear that Thom could clearly see Len directly in front of him on that sunny day. On the other hand, if Thom can convince a court that he really was "daydreaming," he might avoid a conviction for murder. Of course, he would not avoid blame altogether; the case for manslaughter could definitely be made on the grounds that Thom was at least reckless in his conduct—how dare he daydream while driving?

In light of our new considerations, however, it appears that (a) Thom did not knowingly run over Len, because neither his driving behaviors nor his visual input currently held sway in consciousness, and (b) Thom did not recklessly collide with Len because Thom merely fell victim to an "operational protocol" of his standard human body—when a behavior does not require consciousness, remove it, and make room for more complex cognitive endeavors. Thom did not even perform a "voluntary act" in this vehicular homicide, even though he steadily accelerated to Len's position in the road.\(^\text{305}\)

Thom might be held to have acted negligently, since he "allowed himself\(^\text{306}\) to slip off into wool-gathering. Indeed, most would agree that he should have been aware\(^\text{307}\) of Len (most likely daydreaming himself, not conscious of his walking movements or the

\(^{305}\) After the collision, consciousness is obviously returned to driving-related subjects, since (a) obstruction of a habit will return the attendant movement to consciousness, see Kimble & Perlmuter, \textit{supra} note 302, at 376, and (b) the collision is now the most complex cognitive problem demanding Thom's attention.

\(^{306}\) This colloquialism is no longer accurate. Thom is designed to submerge processes not needing his immediate attention. Automatization of behaviors is not a matter of irresponsibility. Such processes would have to be consciously resisted on an ongoing basis in order to avoid Thom's situation. Such resistance to habit forming is probably not possible for a normally functioning human being.

\(^{307}\) \textit{See} MODEL PENAL CODE § 2.02(2)(d) (defining negligence).
curb ahead) crossing the street. But did Thom evidence a “gross deviation from the standard of care that a reasonable person would observe?” Maybe not—we might argue that reasonable persons with efficient brains will, by definition, develop automatized patterns such as car operation. Policy considerations, however, would dictate that we not relieve Thom of all culpability. The need to encourage citizens to resist the automatization process so as to increase public (and pedestrian) safety is strong enough to justify holding Thom negligent.

The goal here, once again, is to decrease the law’s ability to assign intentionality to its subjects, or to attach less moral significance to findings of intention than is currently done. Moore points out that automatized behaviors are “nonetheless accessible to consciousness” if the actor redirects his attention, so Moore would presumably agree with the Model Penal Code’s definition of habitual movement as voluntary behavior. True or not, social policy considerations probably militate in favor of the Code’s viewpoint; safety concerns motivate punishment of habitual actors as negligent in order to promote attentiveness to potentially dangerous activities. We need not specifically narrow the actus reus requirement for phenomena such as habitual movement, but consideration of the automatization process does contribute to the overall necessity of reconsidering intentionality with a skeptical eye.

C. The Ultimate Implication of Eliminativism: Our Selves in Doubt

In the above proposals we see the beginnings of legal reform based on eliminativist claims, the least of which is a loosening of legal reliance on the propositional attitude of “intention.” But analyzing the legal nuances of an expanded definition of intention or even re-explaining legal mechanics to avoid mental states is nit-picking compared to the ramifications of eliminative materialism for the elemental concept of personal identity. We considered the Libet-Dennett combination within the framework of present-day law, law that is grounded in folk psychology. While this was an informative inquiry yielding potential solutions to the dilemmas presented by Libet and Dennett, the broader eliminativist model threatens to sweep the law away. There is a consequence of eliminativism that was previously touched upon but not explicitly stated: by exploding

508 Id.
509 MOORE, supra note 29.
the myth of the unitary mind, eliminative materialism has destroyed the conception of a unified self. Neuroscience today refutes what we have been calling Cartesian materialism, and Dennett's "multiple drafts" model of simultaneously operating neuronal structures seems to be accurate. Since there is no "central command" guiding behavior, how do we blame an entire body for its actions at any one time? Remember the characterization of Dennett's competing neuronal coalitions as a "political community." If consciousness is composed of a chain of neuronal groupings rising and falling in succession, how can the law blame an individual at time $x$ for actions taken at time $y$? Who is it that the law is blaming for the crime?

Dennett's proposal would not be the first to question our belief in a singular self. Freud divided the self into several components, finding an ego, superego, and id to affect the body's behaviors. From a different standpoint, Thomas Nagel, after considering neurophysiological experiments in which severed hemispheres of

---

510 See Dennett, supra note 294, at 171-73.

511 Dennett realizes the implications of an eliminative model: "[T]he self... turns out to be a valuable abstraction, a theorist's fiction rather than an internal observer or boss." Dennett, supra note 40, at 431. Dennett sees all animals as possessing a "functional sense of self" in that they operate to ensure their own survival, but considers the mind or the soul to be "the most tenacious bad idea bedeviling our attempts to think about consciousness." Gelman, supra note 123, at 71 (quoting Daniel Dennett).

It should be noted that different bodies can act independently of one another and yet not possess individual selves. That is, two beings presented with the same situation may react similarly, but they will not react identically. Therefore, are they not exhibiting individual selves, since they react differently to the same circumstances? Not necessarily. According to Dennett's model, the development of neuronal subsystems is affected by environmental factors that make each being differ from others, and this accounts for differing reactions to the same stimuli. However, the problem of assigning blame still remains. When an action is performed, a certain coalition of systems is responsible for it, and there is no way to determine which systems they are. Regardless, if the being is to be punished, we are unable to punish certain neurological systems and not others, and punishing the whole, under Dennett's approach, would seem tantamount to punishing a computer system for actions performed while running a certain program.

512 Moore, supra note 64, at 1633. Moore considers the fragmentation of self implied by Freud's ideas, finding that "the common sense notion of the unified self can withstand any of the insights or phenomena that generated the psychoanalytic subdivisions of self. Indeed, the common sense concept of a unified self better accommodates those insights and phenomena than do the Freudian subdivisions." Id. at 1634.
the brain were made to operate independently, also doubted the truth of the "single mind":

[W]e take ourselves as paradigms of psychological unity, and are . . . subtly ignoring the possibility that our own unity may be nothing absolute, but merely another case of integration, more or less effective, in the control system of a complex organism. This system speaks in the first person singular through our mouths, and that makes it understandable that we should think of its unity as in some sense numerically absolute, rather than relative and a function of the integration of its contents.

But this is quite genuinely an illusion.

Revealing the self as a facade for jostling teams of neurons makes assignments of moral responsibility problematic. Blaming a person for a particular action at a particular time is now akin to blaming a computer for running a particular program. The analogy fails in that an external actor loads the programs the computer runs, while our neurological programs are internally chosen. But under our model they are not "chosen" by anything in particular, emerging instead from a combination of influences involving no master coordinator.

So to whom is responsibility assigned? Our morality requires specific shoulders to bear its moral burdens. Who has rights under the law? Who has obligations and responsibilities? If the law cannot rely on the existence of a unified self, it cannot coherently blame human beings for specific actions. Before any legal revisions can be based on eliminativism, this problem must be solved.

Making Sense of Multiple Selves

Perhaps we can look to an interesting psychological phenomenon for guidance. The law's treatment of individuals suffering from Multiple Personality Disorder (MPD) might shed light on potential methods of assigning moral responsibility to fragmented selves.

---

514 THOMAS NAGEL, MORTAL QUESTIONS 163 (1979).
515 See supra text accompanying notes 204-08, 241.
a. Multiple Personality Disorder and the Fragmented Self

If we conceive of neuronal combinations as constantly shifting "subselves" within the human brain, we might draw a parallel between proposals for legal treatment of MPD and our current dilemma. MPD is a controversial condition, and some psychological commentators doubt its existence.316 The American Psychiatric Association has acknowledged it, however, and defines the disorder as follows:

300.14 Multiple Personality Disorder
A. The existence within the person of two or more distinct personalities or personality states . . . .
B. At least two of these personalities or personality states recurrently take full control of the person's behavior.317

The victim of MPD shifts from personality to personality depending on her current situation, with no "default" or "master" personality to designate as her "self." In the legal context, the obvious question is how to blame body X for criminal acts committed while under personality Y, when body X is now under personality Z. Since there is no singular self, is it fair to incarcerate Z for the acts of Y?

We can conceive of the various personalities or "alters"318 within a body in different ways. They may be viewed as completely different persons,319 different personalities, or facets of one intricate personality.320 The last conception might seem the most useful for our purposes; it describes the "personality states" of the APA description, which sound similar to Dennett's shifting neuronal states or coalitions. But to draw that conclusion is to make once again the "unitary mistake," supposing that there is a single self, except with some remarkably distinct subparts.321

318 This term is commonly used "to refer to [MPD sufferers'] different 'aspects' or 'personalities.'" See Saks, supra note 316, at 386 n.7.
319 See id. at 389, 403-18. Saks is able to make this claim by subscribing to a psychological definition of personal identity, instead of a physical one. This is not the place to debate the point, but the psychological view is not without problems. See generally PETER UNGER, IDENTITY, CONSCIOUSNESS AND VALUE (1990) (making a convincing case for the physically based view of identity through the use of the reader's own intuitions, as elicited by hypothetical situations).
320 See Saks, supra note 316, at 389.
321 Saks makes this assumption in her discussion of MPD and legal responsibility.
tion of MPD most suitable for our needs is actually the second one: the view that alters are "parts or aspects of persons," not possessed of any unitary theme except that of being present in the same body.

Professor Elyn Saks asserts that persons suffering from such a condition should not be held responsible for their criminal acts. To support this claim, she presents the example of Pete, a law-abiding citizen who is drugged and thereby temporarily transformed into Paul, who kills someone. The drug wears off, and Pete finds himself in prison, charged with murder. Saks would exonerate Pete/Paul under a "character analysis" similar to that discussed above. Since Pete can claim that "it was not I who acted," meaning that the act did not reflect his true character, it would not be fair to blame him. Saks then equates Pete/Paul to a sufferer of MPD, who might have two personalities, one "nice" and the other murderous.

The character-based theory of responsibility is, as Saks herself admits, a minority view. But Saks also makes a conceptual error that renders her analysis unhelpful in our search for a relation between the eliminative brain-model and legal responsibility. In the example above, Pete/Paul has a singular self—Pete. He was merely overcome by drug-induced influences that made him kill. His situation is no different than that of a drug addict who commits homicide while in the grips of a delusion. Assessing Pete's responsibility has nothing to do with questions of self. In contrast, the MPD sufferer, like our new model of normal brain function, does not possess a dominant self. She has two equally powerful personality constructs, and thus no locus on which to pin moral/legal responsibility.

See id. at 436-43.
322 Id. at 418.
323 See id. at 419.
324 See id. at 421. Saks also uses the provocation example.
325 Id.
326 See id. at 424 (discussing John/Joe, an MPD sufferer, Saks asserts that "[i]f the character-based theory of responsibility has merit, John would seem to have an excuse for Joe's act, just as Pete would for Paul's: the act does not reflect his character").
327 See id. at 431.
328 Saks attempts to justify her stance:
John [one half of an MPD sufferer, the other half being Joe] cannot say "I am not guilty because Joe did the act," for that would be to say "I am not guilty because another part of me did the act"—and why would that mean no guilt? John can, however, say "I am not guilty because the act of that part
A second proposition for removing MPD sufferers from legal responsibility is more applicable to their actual condition as disunified selves, and thus may be more useful to us. The law excuses an actor if she "did not have the capacity or a fair opportunity not to act," a maxim manifested by the insanity defense. Can the MPD sufferer take advantage of this excuse? Saks notes that courts have presented two views on this question. Some maintain that if any of the MPD victim's personalities did not know of the criminal act, or could not control it, the person should be released. Others hold that the MPD sufferer is innocent only if the particular personality state committing the crime itself meets the insanity test criteria. Saks supports the former view, which would release the MPD sufferer on the grounds that a currently submerged self could not control the act. If correct, this conclusion might lead us to similar findings for normal persons, whom we now see as governed by a series of alternately submerged and then dominant neuron sets.

There are two reasons, however, why this analysis of MPD will not lead to legal impunity for all. The first is that Saks's argument in support of this approach is not convincing. She draws an analogy between the MPD sufferer, who commits crimes while under the aegis of a certain personality, and the sleepwalker, who commits crimes while under the influence of a "‘sleeping self.’" Courts have released defendants for crimes committed while asleep on is no reflection on me—we have different characters." The character-based theory, then, explains why John is innocent of Joe's crime, though they are not different people, provided we can say that they have different characters.

Id. at 424. Saks, whether she intends to or not, suggests that John is the "dominant" partner of the John/Joe relationship, and thus, since John did not commit the deed, "he" should not be punished. But if we remember that John and Joe are on equal footing, the argument loses some of its appeal. Who cares if the crime is no reflection on the character of John? It is a reflection on the character of Joe, who is just as much a part of the body that is John/Joe as John is. It is not as though John has returned for good, as Pete does in the drug example. Joe presumably gets equal cognitive time, and will continue to do so in the future. Do we insist on culpability in every fiber of a defendant's body and experience before we convict her? Do we release a defendant because, under some circumstances, she reveals good character?

329 Id. at 431.
330 See id. at 434.
331 Id.
332 See id. at 435 (citing the famous case of The King v. Cogdon (Victoria 1950) (Austl.) (unreported), in which the court released the sleepwalking Mrs. Cogdon from responsibility for killing her daughter with an axe). Cogdon is discussed in Norval Morris, Somnambulistic Homicide: Ghosts, Spiders, and North Koreans, 5 RES JUDICATAE
the ground that the sleepwalker has a dominant self that the law realizes is not responsible for the crime. The courts treat a defendant's sleep-induced mayhem the same way they deal with Pete's drug-induced mayhem—as the product of forces that overcame the dominant character's legal sensibilities. Thus, both the sleepwalker and Pete are released. Saks's analogy fails because the same reasoning does not hold for an MPD sufferer, whose killer-personality is just as robust and present as any other, more friendly component. As previously noted, MPD sufferers do not have a single dominant self, and neither do we, if Dennett's model is accurate.

But even if the MPD sufferer could take advantage of an insanity defense, there is a qualitative difference between her and a person governed by our new model of consciousness that would prevent the latter from avoiding legal blame. This difference is in the nature of self-fragmentation in Saks's MPD victim and the person with our brain-model. The MPD victim has several discrete personalities or personality states, all of which are separated by an "amnesia barrier." One alter usually has no knowledge of the others, and many MPD sufferers initially come for treatment complaining not of their actual condition, but of "blackout" periods, where they "wake up" involved in some activity of which they have no memory. In addition, each personality has its own extremely distinct characteristics, often being different-handed, or even wearing glasses with differing prescriptions. The alters have already-scripted outlooks and agendas that do not change composition.

The neuron-states of Dennett's model share none of these qualities. The coalitions that might create consciousness are of limitless combination, as subsystems group and regroup in response to environmental stimuli and a host of other factors. More importantly, we do not have consciousness gaps—all neuronal coalitions share in a common neurobiological memory storage, thus

---

29, 29-31 (1951).

333 See Saks, supra note 316, at 397-98. One can easily envision a scenario in which one person seeks out two psychologists—one while under a certain personality, and the other while under another personality. The MPD victim would be unaware of the repetition.

334 Id. at 396.

335 If coalitions of neuron-structures were not variable, we could not account for the breadth of our psychological experience. We touched on a similar point in our discussion of Fodor: that limitless thoughts could only be created if recombinable subparts, or symbols, were used. See supra note 193.
allowing for experiential continuity. The average human being could not claim to have “blacked out” and then killed someone. The smooth readjustments of our neuron processes allow us continuity of thought, thus barring use of an insanity defense.

It appears that analysis of Multiple Personality Disorder will not give us a clear rationale for assigning blame to our new person, one with an ever-adapting virtual machine for consciousness, and no unitary self. Indeed, another distinction between MPD and Dennett’s model is that while MPD sufferers have competing influences that can be considered distinct personalities, or even persons, we do not. Our prior description of neuron combinations as “subselves” may have been misleading. Those groupings do not have the depth of their MPD counterparts—no distinct handedness, eyeglass prescription, emotional disposition, or other person-like traits. They are more accurately described as computer programs, with both the utility and superficiality that term implies. Unfortunately, re-presenting the picture this way leads us to the conclusion that instead of dealing with the legal responsibility of a fragmented self, we are attempting to assign blame to bodies containing no recognizable “selves” at all. This interpretation of Dennett’s model is probably the more accurate one, and leaves us with a completed analogy between humanity and machinery.

As the next section will elaborate, there is a way finally to assign moral responsibility to a human machine consisting of various neuronal programs. Making a legal structure based on eliminativ-

336 See Saks, supra note 316, at 403-09.
337 Remember that Dennett considers the self an abstraction. See supra note 311 and accompanying text. He would agree with our revised assessment.
338 Note that the machine description does not itself force us to forego assignments of blame. Machine-like explanations of human action, or “mechanism,” and rationality can coexist:

[F]rom any particular mechanistic explanation of a bit of behavior it would not follow that that particular bit of behavior was or was not a rational response to the environmental conditions at the time, for the mere fact that the response had to follow, given its causal antecedents, casts no more doubt on its rationality than the mere fact that the computer had to answer “108” [as the product of 18 and 6] casts doubt on the arithmetical correctness of its answer.

Dennett, supra note 244, at 173. We tend, intuitively, to excuse behavior for which we know the cause. Since machines are “causal devices” made up of parts and processes, it may seem as though no responsibility is possible. As Moore points out, we often think “tout comprendre c'est tout pardonner”—to understand all is to forgive all. Moore, supra note 36, at 1092. This fallacy is explored by Moore extensively in his critique of causal theorists. See id. at 1112-48.
ism a possibility entails acceptance of a broadened definition of responsibility. If personal responsibility becomes synonymous with "team" responsibility, the elements of a body's neurological system can be collectively blamed for the actions of any subparts.

b. Collective Responsibility

An expanded conception of responsibility will allow us to blame coherently persons whose consciousness is actually a parade of neuronal coalitions. There is no need to discriminate a blamable self from the background of shifting brain activity. Long-standing legal doctrines provide essential clues to how this can be done: certain legal practices will simply blame persons linked to a criminal act but innocent of its actual commission. Accomplice liability "'extends to acts of the principal in the first degree which were a "natural and probable consequence" of the criminal scheme the accomplice encouraged or aided.'" Participation in a conspiracy will also justify blaming those who did not commit the underlying criminal act. A person who agrees with others to assist in the planning or commission of a crime may be legally responsible for it if any conspirator commits an overt act in furtherance of the conspiracy.

These doctrines evolved from our tendency to assign groups of persons a singular identity, and to identify ourselves with a larger collective in certain situations. For example, every player on a baseball team

is likely to use the first person plural pronoun in speaking about the team's victories or defeats, even if he did not participate in the game . . . . By doing so, the player fully inhabits his role as a member of the team. . . . The actions of other individuals in their capacity as team members become one of the ways in which the particular player authors objects and events. By talking about the victory as "our" victory, the player experiences and enacts what may be called the "collective moment": he subscribes as an author to an event . . . whose intelligibility . . . depends on the existence of a number of similarly situated individuals.

---


340 See MODEL PENAL CODE § 5.03(1)(b)(5).

341 Dan-Cohen, supra note 339, at 986 (emphasis added).
The group activity can develop characteristics that are not assignable to an individual member. As one commentator has quipped, "[t]he charge on San Juan Hill . . . is more than a number of men racing up an incline behind a myopic horseman." An orchestra provides a clearer example: I may only be playing the violin, but the entire group is playing Beethoven's Ninth. I have responsibilities as an individual, but I also have responsibilities as a member of an orchestra.

Legal theories of corporate responsibility further illustrate the potential shift from individual to collective blame. For example, some commentators question the practice of holding corporations criminally liable, since a "corporation" is nothing more than a fiction, an aggregate of individuals in a specific economic condition. Corporations per se, however, do share many characteristics with natural persons that enable them to be deemed "persons" and blamed for bad acts. The key element is to determine the appropriate features shared by legal persons, which seem to be capacities such as rational moral choice more than any biological criteria. As noted previously, an individual is not punished as a full legal person if she cannot think rationally. Like natural persons, corporations arguably have this capacity to make reasonable choices, in that they "choose" between various business strategies and policies. These decisions are not made by any one individual within the organization, and the final choice may not be that which would have been chosen by a single biological decision-maker. The obvious (and folk-psychological) objection to likening corporations to biological beings is that while humans actually have propositional attitudes enabling them to exhibit rational choice, corporations can only pretend to possess them.

We have seen, however, that the attribution of beliefs and desires

---

343 See id. at 846-47. Professor Dan-Cohen gives the example of nationality. See Dan-Cohen, supra note 339, at 987. For example, to be an American is to bear responsibility for specific acts far removed from personal experience, such as the Vietnam War or the space shuttle explosion. Each individual citizen may feel shame or pride with regard to these types of momentous events.
344 See, e.g., MEIR DAN-COHEN, PERSONS, RIGHTS AND ORGANIZATIONS 44 (1986); John Keeley, Organizations as Non-Persons, 15 J. VALUE INQUIRY 149 (1981).
346 See id. at 270.
347 See id.
to humans is a fiction in itself, enabling us to simply codify and explain enormously complex neurophysiological workings. The modern legal conception of corporate liability merely places a similar mask over the mechanics of the corporate structure, and there is no reason why competing intracorporate forces and competing neuronal coalitions cannot both be blamable as groups generating aggregate decisions.\footnote{Walt and Laufer point out that natural persons will sometimes act as “loosely allied decision making units,” giving the impression of multiple selves: [S]ometimes, for some decisions, individuals are like small collectivities. . . . . . . . . . . An individual may adopt strategies intended to limit or preclude choices which otherwise would be made later. An earlier self is strategically manipulating a later self. This can be directly, through pre-commitment, or indirectly, through institutional devices. \textit{Id.} at 271-72 (footnotes omitted).}

It is here we find the kernel of an answer to the problem of assigning blame to people without a unitary mind or self. If the concept of collective responsibility is readily applied to groups of distinct persons, it is even more easily assigned to groups of neurons. Each neuronal coalition is, by definition, linked to a whole, unless the person suffers from MPD or another disorder affecting her continuity of thought. An integral part of all subsystems’ “identities” is their existence as processes within a single body, despite competition with other subsystems and lack of a single coordinator. Every possible coalition is a member of the “team” constituting person X. Thus our cerebral subprocesses acquire “the social identity . . . that can serve as an individual responsibility base for the group’s collective endeavors,”\footnote{Dan-Cohen, \textit{supra} note 339, at 986.} much like Dan-Cohen’s baseball player. Substitute the “body’s” collective endeavors for the “group’s” above, and we are well on our way to justifying assignments of blame to bodies without identifiable selves. The “social fact”\footnote{Flynn, \textit{supra} note 342, at 846. A social fact refers to an essential social ensemble. In this instance, it is the fact that individual practices and events are made more significant by their place in a complex network of social interactions and history.} of collective responsibility is readily envisioned as a biological phenomenon, by which separate brain structures are “synthetically enriched”\footnote{\textit{Id.} (quoting JEAN-PAUL SARTRE, \textit{CRITIQUE OF DIALECTICAL REASON} 378 (Alan Sheridan-Smith trans., 1976)).} by their relation to other structures,
which in combination become an identifiable collective susceptible to blame.

V. BETWEEN HERE AND THERE: POSSIBLE OUTCOMES AND THE LAW'S ATTITUDE TOWARD NEUROSCIENTIFIC CHANGE

The above proposals require further elaboration, but the ideas presented are meant to demonstrate that legal doctrine can begin to accommodate eliminativist ideas and remain understandable. An eliminative future may be a bit strange, but with a rationale for maintaining moral blameworthiness, it can now be possible.

Eliminative materialists make a case for the demise of mental state psychology, and I have made suggestions for filling the legal gaps. If common-sense notions are to be cast aside, a more impressive explanatory framework is required to maintain a just legal system. Perhaps this framework will be "design-oriented," in that it will rely on our Dennett-like picture of neurobiological processes, leading to an eliminative legal system treating the human as a complex machine. Another possibility is the use of the character theory of responsibility, which would assign responsibility based on the collective character expressed by succeeding neuronal coalitions.

But predictions are hazardous, since no one can know what neuroscience will uncover. Churchland bravely speculates concerning social changes in communicative processes, learning capabilities, and the like, but supportable suggestions cannot be made without further empirical evidence. With Libet and Dennett we have looked as far ahead as we can without relying on completely unsupportable assumptions. One firm conclusion is that if the eliminative materialists are correct in declaring folk psychology radically false, a progressive attempt to reform legal codes based on available neuroscientific information can only lead to more just dispensations of moral blame. Dresser imagines a murder trial in which a "mentalometer" is used to retrieve crucial neuronal residue from the defendant's brain, in order to establish accurately the nature of her culpability, however that culpability is assessed in a world free of beliefs and desires. Under Dennett's model, the defendant is presumably blamed for the "bad" neuronal drafting and redrafting traced by Dresser's device, which resulted in an

---

352 See Churchland, supra note 6, at 84-88.
353 See Dresser, supra note 2, at 27.
output that caused an illegal act. We cannot know if this is a realistic proposal, but Dresser has the right idea. The potential legal benefits of tapping the innermost processes of the brain are enormous. A new, sophisticated knowledge of the dynamics of human cognition and behavior could only add credibility to legal pronouncements. What makes contemplation of the changes exciting is that while the benefits may be great, such a new system will probably look nothing like the one we have today.

None of the changes we have discussed will occur overnight. If folk psychology is to become obsolete, this realization will happen only gradually. Development of a neuroscience able to handle the complex evidentiary questions of the criminal law will arise more quickly, but not rapidly enough to require abrupt, systemic alterations within the legal system. There are a few basic paths the future may take concerning scientific progress and the criminal law.

A. We Hit a Dead End

There is the possibility that we will never uncover evidence dispositive of the question of whether mental state psychology is radically false. Baker, who could not conceive of the elimination of belief, and Davidson, who asserted that psychological and biological systems require explanations so different that they are irreconcilable, are representative of this approach. If we are able to convince ourselves of the irreducibility of mental states, we have, a fortiori, removed any problem.

This is not likely to occur. The distinct nature of eliminativism is its marriage to hard science, and scientific research is not likely to stop anytime soon. The chances of uncovering some neural phenomenon “closing the book” on questions of mental state psychology are exceedingly slim. If such a finding is not made, the eliminative materialists will continue to gain ground, since the search for a substitute will be continual.

354 See supra note 194.
355 See supra note 172.
B. We Answer the Question

Another possibility is that further research will vindicate folk psychology as an accurate representation of underlying neurophysiological processes. This is the aforementioned identity theory, which looked to biological correlates for mental states.\(^3\) Like the possibility that we decide mental states are irreducible, such a development would allow us to forego legal reforms. As our discussion has shown, however, it may already be too late for this option. What we know of the brain today is enough to destroy the myth of Cartesian materialism, and neurodiagnostic equipment has not yet revealed identifiable biological correlates for the phenomena of folk psychology. The content problem also continues to plague folk-psychological explanations. Finally, it was previously noted that proving common-sense psychology accurate through neurobiological findings would constitute an enormous coincidence. Folk psychology is quite old, and combined with the content problem, a good argument can be made for its extinction, alongside Aristotelian physics and flat-Earth theories.\(^5\)

C. The Legal Attitude Toward Gradual Neuroscientific Change

Barring the luck required for either of the outcomes above, the law will have to adopt a stance toward neuroscientific change. One option is to ignore studiously the developments of neuroscience and rely on the legal system as it currently exists. If those developments were felt to threaten the moral order severely, such a position would not be unjustifiable. Consider Libet: No matter how solidly one supports his findings of unconscious initiation of voluntary actions, very few of us can fully comprehend the idea of a human being not being the conscious initiator of acts her body performs. As importantly, most persons would consider Libet's conclusions singularly undesirable. Almost one hundred years ago, William James declared:

> If a certain formula for expressing the nature of the world violates my moral demand, I shall feel as free to throw it overboard, or at least to doubt it, as if it disappointed my demand for uniformity of sequence, for example: the one demand being, so far as I can see, quite as subjective and emotional as the other is.\(^5\)

\(^3\) See supra notes 166-78 and accompanying text.  
\(^5\) Churchland, Dennett, and Baker are in agreement that scientific explorations will not vindicate folk psychology. See Dresser, supra note 2, at 34.  
\(^5\) MOORE, supra note 21, at 488 n.33 (quoting WILLIAM JAMES, THE WILL TO
James's position would require some kind of social contract in which the law openly ignores the obsolescence of mental state psychology in favor of a system of blame-assignment in keeping with society's entrenched views on behavior and morality. But when profound neuroscientific discoveries are made, this stance will damage the credibility of the criminal law, since it may not properly reflect what we then know of human behavior. If a convincing redescription that is devoid of psychological pretensions can be developed for the law and its mechanics, the "credibility problem" may never arise at all, but we have seen that this path is rife with potential pitfalls.

The law, being large and unwieldy, will most likely muddle its way into an awkward compromise position. Scientific discoveries below a certain level of significance will be ignored, while those of sufficient credibility, profundity, and notoriety will eventually be reflected in our legal scheme. The historical development of several insanity tests is a lucid example of the law coming to terms with confusing "psychological" phenomena, and the controversy surrounding that effort probably foreshadows further legal-scientific tussles in response to neurophysiological advances. An implicit threshold of acceptability may even be in place today, as the courts accept or reject certain biology-based criminal defenses. An analysis of the factors controlling such decisions would reveal the intricacies of a legal view toward science.

It is unlikely that a formal policy agenda will be instituted with regard to legal adaptation of scientific advances. As has been the case historically, the law will suffer blow by scientific blow, beating a stubborn retreat until it finds itself at a point very different from the one at which it began. It was mentioned at the outset that this incremental evolution is the hallmark of a common law that maintains the stability of society through only gradual alterations. There is no reason to believe that this will not continue, with a steady infusion of neurobiological input slowly molding the law to its contours. Although the eliminativists may succeed in erasing the

BELIEVE 147 (1898)). One wonders whether James could have guessed what brain science would show us. If he were confronted with the profundity of the changes we contemplate, would it solidify his resolve, or prompt him to change his opinion?

559 See DRESSLER, supra note 1, at 289, 299-304.

folk-psychological facade and its legal manifestations, that victory will entail a long and arduous process of legal stubbornness combatting scientific progress.

CONCLUSION

Here ends the tour. This Comment has presented the law in light of its psychological assumptions and explored the neuroscientific progress that threatens them. The struggle between folk psychology and eliminative materialism has been analyzed in an effort to evaluate the credibility of each, and we are left with a few disturbing changes and some educated guesses. In this cognitive-philosophical realm, open-mindedness is a requirement. Humanity has reached the threshold of resolving a few of the Big Questions, and the answers to those questions may not comport with present practice, legal or otherwise. The proposals suggested here are controversial, but what is indisputable is that there are cherished assumptions humanity has nourished in ignorance for thousands of years; generations of forebearers made them a part of us, and we resist their decline with hostility and fear. Society and its legal system must soon confront these assumptions, either in response to the eliminative materialists or other challengers that will surely follow. We can have no idea what further biological truths humanity will uncover, but their discovery is an adventure for philosophers, scientists, legal theorists, and the rest of us alike.