

INTERNAL AND EXTERNAL CHALLENGES TO CULPABILITY

Stephen J. Morse*

“Even a dog distinguishes between being stumbled over and being kicked.”¹

The thesis of this chapter is simple: As long as we maintain the current folk psychological conception of ourselves as intentional and potentially rational creatures, as people and not simply as machines, mental states will inevitably remain central to ascriptions of culpability and responsibility more generally. Not only will this continuity be inevitable, it is also desirable. Nonetheless, we are in a condition of unprecedented internal challenges to the importance of mental states in the context of mental abnormalities and of external challenges more generally to personhood and agency based on the new behavioral neuroscience and genetics. The latter challengers argue that the central role the criminal law gives to mental states is deeply misguided.

I begin with the law’s conception of the person as a folk psychological agent who can potentially be guided by reason. Then I canvas the internal challenges to the importance of mental states, rooted in a trilogy of United States Supreme Court opinions, *Montana v. Egelhoff*,² *Clark v. Arizona*,³ and *Kahler v. Kansas*.⁴ In each, the Court permitted limitations on the extent to which defendants could introduce evidence of mental abnormalities to avoid conviction. I conclude that all these decisions are misguided and that such limitations should not be adopted legislatively or judicially even though it is constitutional to do so. My goal is not to criticize the formal legal reasoning of the opinions, although I do so in passing, but to engage at the level of policy. The action is now in the courts and legislatures.

Next, I address the newer, more general challenges to personhood, agency and responsibility that are fueled by alleged advances in behavioral neuroscience and genetics. Some of these are quite radical. They may even turn out to be correct, but at present, there is no conceptual or empirical reason to believe that they are true. Moreover, there is certainly insufficient reason to jettison notions of criminal responsibility that have been developing for centuries and to adopt instead the proposed, radical conception of justice. I do not argue for any particular categorization or hierarchy of mens rea terms,⁵ or for any particular form of an affirmative defense of legal insanity. Nevertheless, culpability, as expressed in mental state requirements (including action), is central to our value as moral agents.

* Ferdinand Wakeman Hubbell Professor of Law, Professor of Psychology and Law in Psychiatry, and Associate Director, Center for Neuroscience and Society, University of Pennsylvania. Judge Morris Hoffman co-authored the original mens rea material from which the present version is derived and Professor Richard Bonnie co-authored the original insanity defense material from which the present version is derived. Ed Greenlee and Andy Lang furnished fabulous research assistance.

¹ OLIVER WENDELL HOLMES, JR., *THE COMMON LAW* 3 (Dover Publications, 1991) (1881).

² *Montana v. Egelhoff*

³ *Clark v. Arizona*

⁴ *Kahler v. Kansas*

⁵ See Kenneth W. Simons, *Rethinking Mental States*, 72 B.U. L. REV. 463-64 (1992) (providing a critique of the present hierarchy and a new proposal); see also Kimberly Kessler Ferzan, *Don’t Abandon the Model Penal Code Yet! Thinking Through Simons’s Rethinking*, 6 BUFF. CRIM. L. REV. 185 (2002). See generally Paul H. Robinson, *A Brief History of Distinctions in Criminal Culpability*, 31 HASTINGS L.J. 815 (1980).

The Challenge to Personhood, Action and Responsibility

As I type the words of this chapter, I have an experience that virtually every neurologically intact human being takes for granted: the subjective experience of first person agency, the experience of mental causation, that my bodily movements and thoughts are caused, roughly speaking, by my intentions.⁶ To the best of our knowledge, only human beings potentially have a fully developed capacity to act for reasons. This description sounds like Cartesian dualism—the notion that we have an immaterial mind or soul that is somehow in causal relation with our physical body and that causes it to move as the mind directs. But I fully accept that we inhabit a thoroughly material, physical universe in which both all phenomena are consistent with physical laws and causal closure. The latter simply means that every phenomenon in the universe is or is derived from physical matter and there are no mysterious entities. In particular, human action and consciousness are produced by the brain, a material organ that works according to biophysical laws. At present, however, we do not have a clue about how the brain enables the mind, or about how action and consciousness are possible.⁷ Understanding how the brain enables the mind would revolutionize our understanding of biological processes and the nature of personhood,⁸ but such understanding may not be possible.⁹

Although action and consciousness are scientific and conceptual mysteries,¹⁰ they are at the heart of both common sense, “folk psychology,” and the conception of the person inherent in judgments about responsibility and culpability. Folk psychology sounds like a pejorative term, but it is not. It is a generic term for all psychological theories that in part explain human action by mental states such as desires, beliefs and intentions. No folk psychological theory thinks human action is fully causally explained by mental states, but all agree that mental states are central to full causal explanation of human action. The capacity for intentional movement and thoughts—the capacity for agency—is a central aspect of personhood and is integral to what it means to be a responsible person. We act because we intend. Responsibility judgments depend on the mental states that produce and accompany our bodily movements. This is how we think about ourselves, and this is the concept of the person that morality and law both reflect. Law and morality as action-guiding normative systems of rules are useless, and perhaps incoherent, unless one accepts this view of personhood. This explains why the law is and must be a thoroughly folk-psychological institution.¹¹

Virtually everything for which we deserve to be praised or blamed and rewarded or

⁶ I am not suggesting that all bodily movements and thoughts are so caused. Many bodily movements are simply mechanistically caused, such as reflexes, and many thoughts simply spring to mind without any conscious intention to produce them. Some behavior, such as habitual gestures or verbal “tics,” does not seem intentional, but neither is it purely mechanistically produced. One can intentionally bring such movements under conscious intentional control. Such behavior is reason responsive.

⁷ Adolph; MATTHEW COBB, *THE IDEA OF THE BRAIN: THE PAST AND FUTURE OF NEUROSCIENCE* 4 (2020); PAUL R. MCHUGH & PHILLIP R. SLAVNEY, *THE PERSPECTIVES OF PSYCHIATRY* 11-12 (2d ed. 1998).

⁸ MCHUGH & SLAVNEY, *supra* note 7, at 12.

⁹ See generally COLIN MCGINN, *THE MYSTERIOUS FLAME: CONSCIOUS MINDS IN A MATERIAL WORLD* (1999) (arguing that understanding consciousness is impossible for creatures with our limited intellectual capacities).

¹⁰ See generally ROBERT AUDI, *ACTION, INTENTION AND REASON* 1-4 (1993) (describing the “basic philosophical divisions” in each of the four major problem areas in action theory).

¹¹ Katrina Sifferd, *CRIM. L. & PHIL.*

punished is the product of mental causation and, in principle, responsive to reason. Machines may cause harm, but they cannot do wrong, and they cannot violate expectations about how we ought to live together. Only people can violate expectations of what they owe each other, and only people can do wrong. Machines do not deserve praise, blame, reward or punishment. Machines do not deserve concern and respect simply because they exist. They do not have a sense of past, present and future. These concepts apply only to potentially acting, intentional agents.

Suppose, however, that our conscious or potentially conscious intentions are not genuinely causal or seldom are so. To use the title of a recent book by an eminent psychologist, suppose that our “conscious will” is just an illusion.¹² Ordinary notions of action and agency are allegedly under attack from psychology and neuroscience,¹³ a critique that some legal scholarship has begun to embrace.¹⁴ If this is correct, the potential normative implications are profound. Most centrally, if conscious will is an illusion, then concepts of responsibility and desert may be equally illusory or at least inapplicable in most cases of human activity. Perhaps no one really deserves anything and human beings are morally indistinguishable from machines. Although many people think that the implications of a thoroughly physical worldview are profound,¹⁵ I shall argue below when considering the radical challenges to personhood, agency and responsibility that one can fully and consistently accept a material, matter-first worldview and also accept traditional notions of personhood, action, responsibility. and desert. But first let us consider the internal challenges to the importance of culpability in the context of mental abnormalities.

I. THE INTERNAL CHALLENGES: MENTAL ABNORMALITY, MENS REA AND CULPABILITY

This section discusses three doctrinal contexts in which the emphasis on mens rea and responsibility generally are challenged: the use of voluntary intoxication to negate mens rea, the use of mental disorder to negate mens rea, and the existence of an independent, affirmative defense of legal insanity. It uses the Supreme Court jurisprudence in these areas to frame the issue.

Voluntary Intoxication

Montana v. Egelhoff is the leading Supreme Court precedent.¹⁶ I start with the factual background because the drama of the facts often drives policy. In July 1992, while camping in the woods of Montana, James Allen Egelhoff met and became friends with Roberta Pavola and John Christenson, who were also camping. On July 12, they spent the day and evening drinking heavily together. Around midnight, officers of a sheriff's department discovered Christenson's station

¹² DANIEL M. WEGNER, *THE ILLUSION OF CONSCIOUS WILL* (2002).

¹³ Denise Park introduces a group of papers concerned with scientific study of will and writes that the premise of all of them is:

There are mental activations of which we are unaware and environmental cues to which we are not consciously attending that have a profound effect on our behavior and that help explain the complex puzzle of human motivations and actions that are seemingly inexplicable, even to the individual performing the actions.

Denise C. Park, *Acts of Will?* 54 *AMERICAN PSYCHOLOGIST* 461 (1999).

¹⁴See, e.g., Deborah Denno, *Crime and Consciousness: Science and Involuntary Acts*, 87 *MINN. L. REV.* 269 (2002).

¹⁵See JANET RADCLIFFE RICHARDS, *HUMAN NATURE AFTER DARWIN: A PHILOSOPHICAL INTRODUCTION* 15-23 (2000). Richards argues that the thoroughly material view of people exemplified by Darwin's theory appears to challenge traditional notions of what is most distinctive about humans.

¹⁶ Note 2 supra.

wagon in a ditch along a highway. Pavola and Christenson were dead in the front seat, each with a single gunshot in the head. Egelhoff was lying in the rear of the car, yelling obscenities. The officers found Egelhoff's .28 caliber handgun on the floor near the brake pedal. There were four loaded rounds and two empty casings. Egelhoff had gunshot residue on his hands. More than an hour after Egelhoff was taken into custody, his blood alcohol content was .36.

Egelhoff was charged with deliberate homicide, which Montana defines as "purposely or knowingly" causing the death of another human being.¹⁷ "Purposely" and "knowingly" are both subjective. Egelhoff claimed that as the result of extreme intoxication, admittedly voluntarily induced, he lacked the ability to commit homicide. Egelhoff was permitted to introduce evidence of his intoxication at trial. But pursuant to a Montana statute prohibiting a defendant from using intoxication evidence to negate mental states,¹⁸ the trial court instructed the jury that it could *not* consider Egelhoff's intoxication in determining whether he purposely or knowingly killed the victims, as required by the definition of the crime. Egelhoff was convicted, and he appealed on the ground that the intoxication statute violated due process because it prevented the jury from considering relevant evidence and thus relieved the State from proving all the elements of the crime beyond a reasonable doubt. The Montana Supreme Court agreed that the statute violated due process and reversed. Montana appealed to the U.S. Supreme Court.

If James Allen Egelhoff killed when he was genuinely in a state of alcohol-induced "unconsciousness," he did not actually kill purposely or knowingly as Montana law defines these *mens reas*. Perhaps he killed recklessly if he was consciously aware before or during his drinking binge that there was a substantial and unjustifiable risk that he would become homicidal if he drank to the stage of unconsciousness. Or perhaps he was not aware that his drinking created this risk, but he should have known that this risk existed. If so, he is guilty of killing negligently. But if Egelhoff did not kill purposely or knowingly -- a factual issue usually left fully to the jury -- then according to standard principles of culpability and desert, he does not deserve to be punished for killing with one of these two particularly heinous mental states. If Egelhoff's claim about unconsciousness was true, however, he nevertheless demonstrated that he is capable of multiple homicide when drunk. Egelhoff is undoubtedly a dangerous agent.

Montana's statute, which prohibits defendants from using evidence of voluntary intoxication to rebut an allegation that a crime was committed with a required, subjective *mens rea*, expresses moral condemnation of behaving badly when drunk. Aristotle, for example, thought that a person who did harm when drunk was undoubtedly culpable.¹⁹ But getting drunk is one wrong, and whatever else an agent does while drunk is another. With the notable exceptions of felony-murder and certain forms of accomplice liability, the common law does not allow the *mens rea* for one crime to substitute for the *mens rea* required for a second crime. Thus, in one famous case,²⁰ a thief broke a coin-operated gas meter to steal the coins and thereby caused a victim to be exposed to a frightful cloud of poison gas. Charged with both theft and the exposure of the victim to the gas, the court held that the intent to steal the money could not substitute for the required *mens rea* concerning exposing the victim to the gas. To be guilty of the exposure, the defendant had to intend that the victim be exposed or at least be consciously aware of a substantial risk that

¹⁷ MONT. CODE ANN. § 45-5-102 (1995).

¹⁸ MONT. CODE ANN. § 45-2-203 (1995).

¹⁹ NICOMACHEAN ETHICS, Bk. III, ch. 5, 27-29.

²⁰ Regina v. Cunningham, CT. CRIM. APP. [1957] 2 Q.B. 396.

she would be exposed. The *mens rea* for the two crimes had to be proven separately and independently. The exceptions to this rule already noted are highly controversial precisely because they permit strict liability. For example, in felony-murder prosecutions, the *mens rea* for the underlying felony is sufficient to support a charge of murder if death results, even if the felonious defendant lacked the usual *mens rea* concerning death that is typically required to prove murder: purpose, knowledge, or recklessness under circumstances manifesting extreme indifference to the value of human life.²¹

The influential Model Penal Code tries to have it both ways about intoxication.²² While rejecting strict liability generally, the Code provides that a voluntarily intoxicated defendant may use evidence of such intoxication to negate purpose and knowledge but not to negate recklessness. The Code thus equates the culpability for becoming drunk with the conscious awareness of anything criminal that the agent might do while drunk. This "equation" permits the state to meet its burden of persuasion concerning recklessness without actually proving that the defendant was ever actually aware that getting drunk created a grave risk that the defendant would then commit the specific harm the statute prohibited. As an empirical matter, however, this equation is often preposterous. An agent will not be consciously aware while becoming drunk that there is a substantial and unjustifiable risk that he or she will commit a particular crime when drunk, unless the person has a previous history of becoming unconsciously involved when drunk in the creation of great risk of committing this specific crime. If such a prior history or other circumstances indicating previous conscious awareness exists, then the prosecution is capable of proving and should be required to prove the existence of previous awareness. The prosecution should not be able to rely on what is, in effect, the conclusive presumption that becoming drunk demonstrates the same culpability as the actual conscious awareness of a substantial and unjustifiable risk that the defendant would commit the specific harm.

The Montana statute goes even further toward strict liability than the Model Penal Code, of course, by providing that a defendant cannot use evidence of voluntary intoxication to negate purpose or knowledge. One interpretation of the statute -- rejected by Montana's own Supreme Court, but adopted by Justice Ginsburg -- is that the intoxication provision simply works to redefine the mental state element for murder to include an objective *mens rea*: negligence. Ever since the Court's opinion in *Patterson v. New York*,²³ it has been clear that the states have the federal constitutional authority to effect such a redefinition, but this was not Montana's interpretation of its own law.²⁴ More important for my analysis, this redefinition undermines the standard view that culpability is hierarchically arrayed depending on the blameworthiness of the

²¹ See Morse, Reason, Results for a thorough critique of all forms of strict liability.

²² MODEL PENAL CODE, § 2.08(2) (Official Draft 1962).

²³ 432 U.S. 197 (1977) (permitting New York to place on the defendant the burden of persuasion on the issue of "extreme emotional disturbance," New York's analog to the provocation/passion doctrine, which traditionally reduces murder to manslaughter, and permitting the state largely to define, as it wishes, the elements of crime that the state must prove beyond a reasonable doubt).

²⁴ Montana could, if it wished, and without constitutional hindrance, have defined the *mens rea* for murder as negligence. But such a definition would have been a similar abandonment of culpability and objectionable for precisely the same reasons I criticize in the *Egelhoff* opinion. Once again, my quarry is society's response, not constitutional doctrine.

various mental states.²⁵ Our society's dominant morality simply does not accept, and with good reason, that negligent harmdoing is as blameworthy as committing the same harm purposely or with conscious awareness. The latter mental states indicate that the agent is consciously lacking in concern for the interests and well-being of an identifiable victim or class of victims, an attitude toward moral obligations that is more blameworthy than lack of awareness. Few except Oliver Wendell Holmes think that objective and subjective blameworthiness ought to be equated. Characterizing a negligent killer as a murderer does violence to our ordinary notions of culpability and desert.

With these observations in mind, consider Egelhoff's culpability again. First, assume that as the result of voluntary intoxication, James Allen Egelhoff was actually in a mental state that would meet the law's requirement of unconsciousness when he killed Pavola and Christenson. It is not unthinkable morally to condemn drinking oneself purposely or recklessly into a state of unconsciousness, but this behavior is not a crime per se. Criminal law theorists dispute the basis for the exculpatory effect of unconsciousness, but all agree that it does exculpate. Thus, if one believes Egelhoff's claim that he was legally unconscious, or to put it more accurately, if the prosecution were unable to prove beyond a reasonable doubt that he was legally conscious, then Egelhoff is not guilty of purposely or knowingly killing.²⁶ Moreover, there is no evidence that Egelhoff was consciously aware when he was drinking that he would become homicidal when drunk. Thus, he did not kill recklessly, even if one looks back to his earlier mental states to find culpability. Once again, Egelhoff might be fully responsible for becoming unconscious, but without proof of the mental states usually required, it is a form of strict liability to hold him fully accountable for anything that he did while unconscious. He culpably caused the condition that would negate the prima facie case, but not with purpose, knowledge, or recklessness that he would be exculpated.

Egelhoff is a dangerous agent, and it is undeniable that the State might have great difficulty proving beyond a reasonable doubt on these facts that he was legally conscious and thus guilty of purposely or knowingly killing. If he was legally conscious at the time of the killings, of course, the precision of the executions suggests that the most sensible inference is that he killed purposely or knowingly, even if one believes his claim that he did not remember the homicides. Without the crutch of strict liability, however, the State might be able to convict only for negligent homicide, typically graded as involuntary manslaughter, which carries a substantially shorter term of years than murder. But our fear of Egelhoff and revulsion at his deeds should not be allowed to prove too much. The Constitution's requirement that in criminal cases the state must prove each element of the crime charged beyond a reasonable doubt almost always makes it more difficult for the prosecution to prove its most serious charge. Our society bears this risk because, except in a small number of inevitable cases, we believe that it is unacceptable to convict a legally innocent person. Concern with culpability thus almost always conflicts with concern for public safety.

²⁵ See Douglas Husak, *The Sequential Principle of Relative Culpability*, 1 LEGAL THEORY 493 (1995) (defending a qualified version of the claim that culpability is hierarchically arrayed depending on the relative blameworthiness of particular mental states).

²⁶ The alternative theories that would support that result are that unconsciousness negates the act requirement, negates *mens rea*, or provides an affirmative defense.

Egelhoff signals weakened commitment to the importance of culpability. Negligent homicide is not the same as intentional killing, and the culpability of becoming drunk and unconscious is not the same as the culpability for murder. The Supreme Court's acceptance of the equation of these morally distinguishable cases is disquieting. Alcohol abuse, whether or not it is associated with a condition properly called addiction, is a serious social problem that must be addressed. But weakening mens rea requirements and indirectly criminalizing non-criminal behavior or treating minor criminal violations as major is an illiberal and surely ineffective remedy.

Mental Disorder, Intellectual Disability & Mens Rea

Doctrines that permit defendants to present mental disorder evidence to negate mens rea are often misleadingly termed “diminished capacity,” suggesting that these doctrines are some kind of mitigation or partial excuse. This is incorrect. They refer to a straightforward denial of a requisite element, akin to a denial that one acted or to a mistake of fact.²⁷ They are not a lesser form of the insanity defense. The failure to recognize this clear distinction often produces faulty reasoning about whether a defendant should be permitted to introduce mental disorder evidence to negate mens rea because courts wrongly believe the defendant is requesting the court to create a “mini” insanity defense.²⁸

Mental disorders relevant to mens rea most frequently produce disordered cognition, such as hallucinations or delusions, or untoward, sometimes strong, desires, such as the desire to have sexual relations with minors or the desire to set fires or to take controlled substances. In other words, mental disorder produces crazy desires or crazy beliefs about reality, but it virtually never prevents a defendant from meeting the law's criteria for intention, knowledge, conscious awareness, and other mens rea terms.²⁹

Likewise, mental disorder seldom disables a defendant from having the *capacity* to form a mens rea.³⁰ Modern inquiries into state of mind, both academic and judicial, seem obsessed with the vexed question of whether a defendant had the *capacity* to form mens rea, but either a defendant had the requisite mens rea or he did not. Mental disorder may in some cases demonstrate that the defendant did not form the mens rea at the time of the crime. Not having a mens rea or having a mental state inconsistent with the requisite mens rea does not mean, however, that someone was incapable of forming it. If an agent lacks the capacity to do something, it follows that the agent did not do it in fact. Thus, evidence about the defendant's capacity to form a mens rea is logically and factually relevant to whether it was actually formed.

Asking about a defendant's capacity to form a mental state never provides better information than inquiring directly whether the mens rea was formed in fact, which is the ultimate legal question. Resolving questions about capacity requires a counterfactual inquiry that we lack the clinical and scientific resources to answer. When an expert testifies that a defendant lacked the

²⁷ Stephen J. Morse, *Undiminished Confusion in Diminished Capacity*, 75 J. CRIM. L. & CRIMINOLOGY 1, 6 (1984).

²⁸ See, e.g., *Bethea v. United States*, 365 A.2d 64 (D.C. 1976). The Arizona case the Supreme Court considers in *Clark v. Arizona*, 548 U.S. 735 (2006), discussed below, makes precisely this error. *State v. Mott*, 931 P.2d 1046, 1050-51 (Ariz. 1997). Indeed, as we shall see, the *Clark* majority itself blurs the two doctrines. See *infra*.

²⁹ See Morse, *Crazy Reasons*, at 197-98.

³⁰ Morse, at 42.

capacity to form a mens rea, that opinion seldom has a clinical or scientific basis. It was precisely these types of difficulties that led California to bar testimony about the defendant's capacity to form a mens rea, although it did permit testimony about whether the mental state was formed in fact.³¹

The examples of Daniel M'Naghten and Andrea Yates will help demonstrate that even the most delusional or hallucinating person can form the requisite mental state. M'Naghten delusionally believed that there was a conspiratorial Tory plot to kill him, and formed a preemptive plan to kill the Tory Prime Minister, Robert Peel.³² When he shot and killed Peel's secretary, Drummond, believing the secretary was Peel, he surely intended to kill a person.³³ Likewise, Andrea Yates believed that unless she killed her children, they would become corrupt and would be tormented by Satan for eternity.³⁴ She therefore decided to kill her children.³⁵ She knew they were human beings and that human beings are killed by drowning. Ms. Yates surely intended to kill the five children when she drowned them in the bathtub. Likewise, a person suffering from auditory hallucinations who hears God's voice command him to kill surely forms the intention to kill when he kills in response to the hallucinated command.

Admittedly, on rare occasions, psychotic mentation is genuinely inconsistent with the formation of mens rea. In a well-known California case,³⁶ the defendant, Wetmore, was caught in the victim's apartment under conditions suggesting that he intended to steal the victim's property. Charged with burglary, the defendant claimed that he delusionally believed that the apartment and the property belonged to him.³⁷ If he told the truth, he did not intend to enter the apartment of *another* or to commit the felony of larceny, the elements of which include intentionally taking and carrying away the property of *another*.

Note that even if mental disorder does negate subjective mental states such as purpose, intention, knowledge, or conscious awareness of risk, it would never negate the objective negligence standard. The person with mental disorder who is unaware of a risk that a reasonable person should be aware of is by definition unreasonable. Even the Model Penal Code, which individualizes the negligence inquiry somewhat by requiring the decision maker to consider the behavior of a "reasonable person . . . in the actor's situation" before making a finding of negligence,³⁸ would not go so far as to consider mental disorder—irrational behavior—as part of the "situation." The Model Penal Code never gives a clear definition of the "situation," but it does make clear that it wishes to avoid complete subjectivization of the reasonable person standard.³⁹

³¹ CAL. PENAL CODE § 28 (1998).

³² RICHARD MORAN, KNOWING RIGHT FROM WRONG: THE INSANITY DEFENSE OF DANIEL MCNAUGHTAN 10 (1981).

³³ M'Naghten's Case, (1843) 8 Eng. Rep. 718 (H.L.).

³⁴ Phillip Resnick, *Andrea Yates Case: Insanity on Trial*, 55 CLEV. ST. L. REV. 147, 149 (2007); see also Deborah W. Denno, *Who Is Andrea Yates? A Short Story About Insanity*, 10 DUKE J. GENDER L. & POL'Y 1 (2003) (providing a thorough account of the entire background and an analysis of the expert testimony).

³⁵ Resnick, *supra* note 34; see also Denno, *supra* note 34.

³⁶ *People v. Wetmore*, 583 P.2d 1308 (Cal. 1978).

³⁷ *Id.* at 1310.

³⁸ MODEL PENAL CODE § 2.02(2)(d) (1962) (emphasis added).

³⁹ MODEL PENAL CODE AND COMMENTARIES § 2.02 at 242 (1985).

To assess reasonableness from the standpoint of the “reasonable irrational” person would deprive the negligence standard of all objectivity.⁴⁰

In other cases, mental disorder may not necessarily be inconsistent with formation of mens rea, but evidence of disorder may help bolster the defendant’s claim that he did not form it. For example, suppose a psychotically disorganized person gets lost in an empty part of town on a cold winter’s night and cannot find his way home. To escape the cold, he breaks into a building, is caught, and is charged with burglary on the theory that he intended to steal. In this case, he is fully capable of forming the intent to steal, but his mental disorder helps explain why he broke in simply to keep warm.

Again, the crucial issue is to determine what the defendant’s actual mental state was and to compare that mental state to the mental state required by the crime charged. Of course, the lurking problem is that it is sometimes very difficult to determine a defendant’s mental state at the time of the crime. Memories fade or are wittingly or unwittingly shaped by what happens afterwards. Defendants have powerful incentives to lie, but these difficulties arise in all retrospective mental state evaluations, and not just in cases involving mental disorder. As I discuss later in this Article, one of the primary skills we have evolved over the last 100,000 years is the ability to discern the intentions of other humans.⁴¹ Like Holmes’s dog, jurors are quite good at recognizing the difference between a kick and a stumble.

In sum, the mens rea issue is entirely distinct from the legal insanity issue, even if precisely the same evidence would be relevant to adjudicating both claims. People with mental disorder are not automatons; rather, they are agents who act for reasons. Their reasons may be motivated by distorted perceptions and beliefs, but they do form intentions and have knowledge of what they are doing in the narrow, most literal sense. Thus, it is very uncommon for mental disorder to negate all mens rea, even if the defendant is profoundly delusional, as Daniel M’Naghten and Andrea Yates presumably were.

In some rare cases, as we have already discussed, evidence of mental disorder might negate mens rea because the mental state it produces will be flatly inconsistent with the mens rea required by the definition of the crime or because it indirectly helps to explain why mens rea was not formed on that occasion.⁴² In these cases, the same evidence that a defendant was delusional may both

⁴⁰ Failure to recognize this point and the incorrect belief that the Model Penal Code adopts nearly complete subjectivization for negligence are major analytic reasons that the primary contemporary proposal to abolish the insanity defense fails. See Christopher Slobogin, *An End to Insanity: Recasting the Role of Mental Disability in Criminal Cases*, 86 VA. L. REV. 1199, 1239 (2000), reprinted in expanded form, but with unchanged analysis, in CHRISTOPHER SLOBOGIN, *MINDING JUSTICE: LAWS THAT DEPRIVE PEOPLE WITH MENTAL DISABILITY OF LIFE AND LIBERTY* 52-53 (2006) [hereinafter Slobogin, *An End to Insanity*]; see also CHRISTOPHER SLOBOGIN, *MINDING JUSTICE: LAWS THAT DEPRIVE PEOPLE WITH MENTAL DISABILITY OF LIFE AND LIBERTY* (2006) [hereinafter SLOBOGIN, *MINDING JUSTICE*]. Professor Slobogin’s “integrationist” alternative to the insanity defense is discussed *infra* in the text accompanying notes **Error! Bookmark not defined.**-62.

⁴¹ See *infra* text accompanying note **Error! Bookmark not defined.**-31.

⁴² See *supra* text accompanying notes 27-**Error! Bookmark not defined.**

negate mens rea and support a finding of legal insanity, but the questions being answered by the evidence are different.⁴³

With this background in mind, let us turn to *Clark*. In the early morning of June 21, 2000, Eric Clark, a seventeen-year-old resident of Flagstaff, Arizona, was riding around in his pickup truck blaring loud music.⁴⁴ Responding to complaints about the noise, Officer Jeffrey Moritz, who was in uniform, turned on the emergency lights and siren of his marked patrol car and pulled Clark over.⁴⁵ Moritz left the patrol car and told Clark to remain where he was.⁴⁶ Less than a minute later, Clark shot and killed Moritz.⁴⁷

Clark was charged with intentionally killing a police officer knowing that the officer was acting in the line of duty.⁴⁸ He did not contest the shooting and death, but he claimed that as a result of paranoid schizophrenic delusions, he lacked the required mens rea for the crime charged (the intent to kill a person and the knowledge that the victim was a police officer).⁴⁹

Substantial evidence, including Clark's statements to classmates a few weeks earlier that he wanted to shoot police officers, suggested that Clark knew Moritz was a police officer and that he had planned just such a shooting. He had even arguably lured Officer Moritz by driving his truck with its radio blaring in a residential area. On the other hand, Clark presented testimony from family, classmates, and school officials about his bizarre behavior during the preceding year, including rigging his bedroom with fishing line, beads, and chimes to warn him of intruders, and keeping a bird in his car to warn him of airborne poison.⁵⁰ These actions were plausibly a result of his paranoid delusions. Indeed, there was lay and expert testimony that Clark thought Flagstaff was populated with "aliens," including some that were impersonating police officers, that the aliens were trying to kill him, and that only bullets could stop the aliens.⁵¹ The defense expert also testified that Clark may have turned the radio up to drown out auditory hallucinations.⁵²

The operative Arizona rule concerning the admission of evidence of mental disorder to negate mens rea was based on an Arizona Supreme Court decision, *State v. Mott*, which held that psychiatric testimony was inadmissible to negate specific intent and that evidence of mental disorder, short of legal insanity, was not admissible to negate any mens rea element.⁵³

⁴³ For example, if Clark really believed he was killing an alien impersonating a police officer, he lacked knowledge that he was killing a police officer, thus negating the mens rea of knowledge. For the same reason, he did not know what he was doing or that it was wrong.

⁴⁴ *Id.* at 2716.

⁴⁵ *Id.*

⁴⁶ *Id.*

⁴⁷ *Id.*

⁴⁸ *Id.* All statements of the facts are taken from the Supreme Court opinion.

⁴⁹ *Id.* at 2717-18.

⁵⁰ *Id.* at 2717.

⁵¹ *Id.*

⁵² *Id.* at 2739.

⁵³ *State v. Mott*, 931 P.2d 1046, 1051 (Ariz. 1997) (en banc), *cert. denied*, 520 U.S. 1234 (1997). The Arizona Supreme Court mischaracterized the use of battered woman syndrome evidence offered to negate mens rea as a "diminished capacity" defense. *Id.* at 1050. It then held, consistent with its mistaken premise, that the Arizona legislature had implicitly rejected that defense when it refused to recognize generalized diminished capacity and ruled that the evidence could be excluded. *Id.*

At a bench trial, the judge permitted introduction of all the lay and expert testimony about Clark's mental disorder at the time of the crime.⁵⁴ The judge specifically found that Clark was suffering from paranoid schizophrenia and had distorted perceptions at the time of the crime. He ultimately ruled, however, that *Mott* barred him from using it to consider mens rea and that he could consider this evidence only to decide the issue of legal insanity.⁵⁵ He found Clark guilty of first degree murder.⁵⁶

Clark appealed on the grounds that the *Mott* rule violated procedural due process. The Arizona court of appeals affirmed the conviction, the Arizona Supreme Court denied certiorari, and the United States Supreme Court granted certiorari to decide "whether due process prohibits Arizona...from excluding evidence of mental illness and incapacity due to mental illness to rebut evidence of the requisite criminal intent."⁵⁷ In addressing this issue, the Court's majority goes quite wrong, not only confounding mens rea and insanity, but forcing psychiatric and psychological evidence into an arbitrary system of classification that is unworkable and that had never been considered or argued by Arizona.

Although the extent of the *Mott* rule's evidentiary exclusion is not entirely clear, it is clear that it prevents defendants from introducing substantial, relevant, and reliable mental health expert testimony concerning whether or not the requisite mens rea was formed in fact. Recall Clark's factual claim that, as a result of delusions produced by mental disorder, he actually believed Officer Moritz was a dangerous "alien" impersonating a police officer.⁵⁸ If this were true, then Clark would be simply not guilty of intentionally killing a person (unless the definition of "person" was expanded to include extraterrestrials), let alone guilty of knowingly killing a police officer acting in the line of duty (again, unless the definition of "police officer in the line of duty" was expanded to include an extraterrestrial whose alien line of duty was to impersonate a police officer).

Recall that the trial judge permitted Clark to introduce expert testimony that addressed the mens rea question, but ultimately decided that *Mott* barred him from considering this testimony in reaching a verdict on the homicide charge.⁵⁹ Thus, Clark was functionally prevented from using relevant and reliable testimony to cast a reasonable doubt about the mens rea for the crime charged.

The Tripartite Evidence Construction

In order to insulate the *Mott* rule from Clark's due process claim, the Supreme Court majority mischaracterized Arizona's rule. Instead of holding that *Mott* excluded expert mental health testimony on the issue of mens rea, the Court ruled that *Mott* actually applied only to two of the three types of distinguishable mens rea evidence: observation evidence, mental-disease evidence,

⁵⁴ *Clark*, 126 S. Ct. at 2717.

⁵⁵ *Id.*

⁵⁶ *Id.* at 2718.

⁵⁷ *Clark*, 126 S. Ct. at 2718.

⁵⁸ *Id.* at 2749.

⁵⁹ *Id.* at 2717-18. A psychiatrist called by the defense not only testified that Clark was suffering from paranoid schizophrenia with delusions at the time of the killing and was therefore insane, but also that because of those delusions, Clark could not have formed the intent to kill Officer Moritz with knowledge that Moritz was a policeman. *Id.*

and capacity evidence.⁶⁰ The Court then upheld the rule against a due process attack simply by announcing that the Arizona courts could not possibly have intended it to apply to the first category.⁶¹

The majority defined the first category, “observation evidence,” “in the everyday sense [as] testimony from those who observed what [a defendant] did and heard what he said; this category would also include testimony that an expert witness might give about [the specific defendant’s] tendency to think in a certain way and his behavioral characteristics.”⁶² The Court pronounced that this first category of mens rea evidence was admissible despite *Mott* and could be presented by either lay or expert witnesses.⁶³

The second category, “mental-disease evidence,” was defined as “opinion testimony that [a defendant] suffered from a mental disease with features described by the witness.”⁶⁴ That is, such testimony provides general information about the mental disorder from which the specific defendant allegedly suffers. The Court pronounced this kind of evidence, which is almost always provided by mental health experts, inadmissible under *Mott*.

The majority defined the third category, “capacity evidence,” as evidence about a defendant’s “capacity for cognition and moral judgment (and ultimately also his capacity to form mens rea).”⁶⁵ This, too, is opinion evidence provided by experts, and, as with category two, the Court upheld *Mott*’s exclusion of this type of mens rea evidence.

In upholding the constitutionality of *Mott*’s bar to mens rea evidence in categories two and three, the *Clark* majority mistakenly refers to barring these types of evidence on the legal insanity issue, not mens rea: “Thus, only opinion testimony going to mental defect or disease, and its effect on the cognitive or moral capacities on which *sanity* depends under the Arizona rule, is restricted.”⁶⁶ Is this astonishing error typographical, or a deeper reflection of the majority’s own confusion between these two very different principles?⁶⁷

The majority’s unprecedented tripartite construction fails to do the theoretical work necessary to draw a sensible line between which types of expert mental disorder evidence states may and may not exclude. This classification was not part of Arizona law (or any state law I know about) and cannot be found in any Supreme Court precedent. If that were not enough, neither of the parties or amici ever suggested such a construction in their briefs or at oral argument.

⁶⁰ *Id.* at 2724-26.

⁶¹ *Id.* at 2725-26.

⁶² *Id.* at 2724.

⁶³ *Id.* at 2725. It is interesting that the majority constructs these categories of mens rea evidence without being explicit about whether they are meant to describe the actual limits of *Mott* (an odd thing for a federal court to be doing when the rule at issue was invented by a state court) or the constitutional limit of *Mott*’s reach. It must be the latter, but the majority’s reluctance to admit it is puzzling.

⁶⁴ *Id.*

⁶⁵ *Id.*

⁶⁶ *Id.* at 2726 (emphasis added).

⁶⁷ This is not an isolated example of the majority conflating insanity and mens rea. As discussed *infra* in the text accompanying note 71, they do it again later in the opinion.

Moreover, this construction not only allowed the Court's majority to find the artificially restricted rule constitutional, it also allowed it, alternatively, to conclude that Clark failed to preserve his constitutional argument on this point because he never asked the trial court to admit mental state evidence in category one despite the exclusion of categories two and three.⁶⁸ As Justice Kennedy aptly put it in his dissent, "Seizing upon a theory invented here by the Court itself, the Court narrows Clark's claim so he cannot raise the point everyone else thought was involved in the case."⁶⁹

Justice Kennedy, and to some extent Justice Breyer in his concurrence, also recognized that the majority's tripartite classification quickly breaks down in practice and therefore ends up being an unworkable solution to this constitutional problem.⁷⁰ There are clear, core cases of each type, but testimony rarely comes so neatly packaged. For example, sound empirical evidence about the characteristics of people suffering from a particular mental disorder is based on observation and is factual. In many cases, such evidence would help a finder of fact understand the behavior of a defendant who suffers from that disorder, even though the evidence comes from the observations of others. That is to say, much of the inferential expert evidence Justice Souter believes *Mott* may constitutionally exclude is actually observational evidence that could not have been excluded had the observations been about the defendant himself. Would experts be allowed to testify about their own observations of the defendant, but then not be permitted to testify about the features of a recognized diagnostic category that help explain those observations?

All clinical judgments in medicine, psychiatry, and clinical psychology are, by their very nature, informed by the clinician's observation of the particular patient being seen, by the accumulated wisdom of observations of other patients by that clinician and others, and by findings from empirical studies. The Court's fictitious categories of evidence not only bleed into one another, but they also seem peculiarly unable to do the important constitutional work the Court asks of them.

The Court also reasoned that Clark's due process claim depended on the application of the presumption of innocence, the presumption of sanity, and "the principle that a criminal defendant is entitled to present relevant and favorable evidence on an element of the offense charged."⁷¹ The Court noted, as it was bound to under *Winship*, that the presumption of innocence could only be overcome by proof beyond a reasonable doubt of all elements, including mens rea. This is boilerplate.

The Court then observed that the presumption of sanity is universal in some form, but confusingly mischaracterized it as a presumption "that a defendant has the capacity to form the *mens rea* necessary for a verdict of guilt and the consequent criminal responsibility."⁷² Once again, the Court confused mens rea with insanity, this time in the heart of the opinion explaining why Arizona may constitutionally prevent a defendant from introducing category two and three

⁶⁸ *Clark*, 126 S. Ct. at 2727.

⁶⁹ *Id.* at 2738 (Kennedy, J., dissenting).

⁷⁰ *Id.* at 2737-38 (Breyer, J., concurring), 2738-39 (Kennedy, J., dissenting).

⁷¹ *Id.* at 2729 (majority opinion).

⁷² *Id.* at 2730.

evidence to negate mens rea. Indeed, the whole presumption-of-sanity discussion in this section of the opinion is irrelevant to Clark’s mens rea claim.

The Court not only blurred the distinction between mens rea and insanity, but also made wholly inconsistent observations about the relationship between the two. It rejected the argument that mens rea and insanity are “entirely distinguishable,” yet obscurely noted that insanity “trumps” mens rea, suggesting that they are not functionally distinguishable because the former subsumes the latter when both are claimed.⁷³

The majority correctly acknowledged that evidence of the defendant’s state of mind at the time of the crime might indicate the defendant’s actual mental state and the presence of an enduring incapacity to form the requisite mens rea.⁷⁴ Given that acknowledgment and the majority’s recognition of a defendant’s constitutional right to present evidence that negates an element, how did the Court conclude that Arizona could constitutionally deprive Clark of that right? The majority reached this conclusion by claiming that Arizona could permissibly “channel” mental disorder evidence solely into the insanity issue and out of the mens rea issue because Arizona had legitimate state interests in such channeling.

The Channeling Argument

The Court accepted Clark’s characterization of the *Mott* rule as a rule of evidence rather than as a reworking of the elements of homicide, and conceded again that the evidence was relevant. As *Montana v. Egelhoff*⁷⁵ makes clear, states may of course preclude relevant defense evidence—even evidence that rebuts an element of the offense—if the state has a legitimate purpose in excluding the evidence. What were Arizona’s “legitimate purposes” in channeling some forms of expert mental state evidence into insanity and away from mens rea? Here is where the Court’s reasoning goes radically wrong.

The Court’s first identified “legitimate reason” is yet another example of its conflation of mens rea and legal insanity. In the part of the opinion addressing the insanity issue, the Court reaffirmed Arizona’s authority to define legal insanity as it wishes and to place the burden of persuasion for this defense on the defendant.⁷⁶ Consequently, the Court reasoned that if Arizona is to have this authority in practice as well as in theory, it “must be able to deny a defendant the opportunity to displace the presumption of sanity more easily when addressing a different issue”⁷⁷ This is a non-sequitur that is based on confusions about the presumption of sanity and “capacity” that we addressed above.⁷⁸ There is no presumption of sanity applicable to the mens rea elements. Even if there were, the presumption must be rebuttable and a crucial method of rebuttal in a case involving severe mental disorder would be by the introduction of expert testimony.

⁷³ *Id.* at 2731 n.38

⁷⁴ *Id.*

⁷⁵ See *supra* text accompanying notes **Error! Bookmark not defined.-Error! Bookmark not defined..**

⁷⁶ *Clark*, 126 S. Ct. at 2732-33.

⁷⁷ *Id.* at 2732.

⁷⁸ See *supra* text accompanying notes 72-56.

Mental disorder and capacity evidence used to negate mens rea could result in acquittal simply by presenting a reasonable doubt about mens rea, whereas the same type of evidence used to prove legal insanity would succeed in Arizona only if the defendant convinced the finder of fact of his insanity by a preponderance of the evidence. Such a difference is not a “displacement” of the presumption of sanity, however. It is simply a logical consequence of the interaction between the structure of criminal culpability and *Winship*’s constitutional requirements. Criminal responsibility can be avoided either by negating an element of the crime charged or by establishing an affirmative defense. Permitting an affirmative defense, as the Court recognizes, does not remove the State’s obligation to prove the elements beyond a reasonable doubt. Clark’s claim does not undermine Arizona’s right to define legal insanity as it wishes.

The Court’s second “legitimate purpose” is Arizona’s desire “to avoid a second avenue for exploring capacity, less stringent for a defendant.”⁷⁹ Mens rea is not about “capacity” or about “criminal responsibility” more broadly; it is about whether a mental state required by the definition of the crime charged was in fact formed. Contrary to the Court’s assertion, permitting a jury to use mental-disease or capacity evidence to decide if there is a reasonable doubt about mens rea is *not* “in functional terms . . . analogous to allowing jurors to decide upon some degree of diminished capacity to obey the law . . . that would prevail as a standalone defense.”⁸⁰ Negation of mens rea is not an independent “defense” of “diminished capacity.” Mens rea negation is simply a straightforward denial of the prima facie case that needs no special name. The question of “diminished capacity” is the province of the insanity defense or mitigating evidence at sentencing, not mens rea negation. The state undeniably has the authority to reject a partial responsibility mitigating doctrine, but this was not what Clark was requesting. The majority’s constant blurring of the mens rea and legal insanity issues simply perpetuates this confusion.

The Court then articulated a third legitimate purpose: Arizona has made a determination that mental state evidence is just too unreliable and risky to be the basis of a complete defense under which a defendant is entirely acquitted and unconditionally freed. Now, at last, we are at the heart of Arizona’s real concerns. In fact, the Court identified three separate such risks: (1) the controversial character of some categories of mental disease; (2) the potential of disease evidence to mislead; and (3) the danger of according greater certainty to capacity evidence than experts claim for it.⁸¹ The Court’s general conclusion about the second and third risks was that shifting the burden of persuasion to the defendant by channeling the evidence into the insanity issue would reduce the risk that misleading evidence would lead to incorrect verdicts.

The Court is certainly correct and has noted on many occasions that there is great debate about the concept of and criteria for mental disorders.⁸² As a result, caution is warranted “in treating psychological classifications as predicates for *excusing* otherwise criminal conduct.”⁸³ This is true

⁷⁹ *Clark*, 126 S. Ct. at 2733.

⁸⁰ *Id.*

⁸¹ *Id.* at 2734-36.

⁸² *See supra* Part II.

⁸³ *Clark*, 126 S. Ct. at 2734 (emphasis added). One of us is a long-time critic of the use of diagnostic information in civil or criminal cases. Stephen J. Morse, *Crazy Behavior, Morals and Science: An Analysis of Mental Health Law*, 51 S. CAL. L. REV. 527, 604-15 (1978) [hereinafter Morse, *Crazy Behavior*]; Stephen J. Morse, *Failed Explanations and Criminal Responsibility: Experts and the Unconscious*, 68 VA. L. REV. 971, 1055-70 (1982) [hereinafter Morse, *Failed Explanations*].

enough, but the same argument applies to any use of diagnostic information, which is routinely admitted in a wide array of civil and criminal law contexts, including the insanity defense. The Court's argument proves too much.

Moreover, the defendant is not seeking to excuse his conduct. This might be a valid reason for channeling in the other direction: allowing mental disorder evidence for mens rea and disallowing it for insanity. It is not a justification for channeling in the *Mott* direction, which *allows* excusing evidence but *disallows* evidence that would exonerate because it defeats the prima facie case. If Eric Clark genuinely thought that he was killing an alien impersonating an officer, he is simply not guilty, full stop, of homicide. He killed a person, but he did not commit the crime of homicide of a police officer, which requires that he intentionally kill a person with knowledge that the victim was a police officer.

The second risk—that mental-disease, i.e., general diagnostic, evidence may lead to the incorrect conclusion that the defendant lacks capacity to form mens rea when in fact he possessed mens rea—is true to a degree. Avoiding such a risk is no doubt a legitimate state interest. There is great heterogeneity in psychiatric diagnostic categories and imperfect fit, as the American Psychiatric Association recognizes, between those categories and legal questions.⁸⁴ Testifying experts, alas, do not always confine themselves to providing rigorously confirmed evidence about the characteristics of people like the defendant, and instead fall into the trap of using diagnostic terms that do not inform the legal issues. It is all too easy for professionals and lay people alike to make the mistake of begging legal questions based on a psychiatric or psychological diagnosis.

Indeed, the Court pointed out that the testifying experts in *Clark* made this error themselves while testifying about legal insanity: they agreed on the diagnosis of schizophrenia but disagreed about Clark's cognitive and moral capacity.⁸⁵ Given the dangers of mental-disease evidence to mislead, the Court concluded, and we partially agree, that it is reasonable for a state to decide to channel the evidence to the insanity defense on which the defendant can be assigned the burden of persuasion.

This justification for complete channeling of expert evidence to the insanity issue nevertheless again proves too much. As with some of the Court's other justifications, there is no reason a state's skepticism about psychiatric evidence should begin and end with the criminal law. States use diagnostic information in a host of other legal contexts.⁸⁶ If anything, the law should be more forgiving when criminal blame and punishment are at stake. Moreover, the problem arises less from the inherent tendency of diagnostic information to mislead than from confusion about the nature of the relation of such categories to a legal conclusion. Too often, as we noted above,⁸⁷ people wrongly believe that if a mental disorder played a causal role, the behavior is akin to a mechanism and the defendant is therefore not responsible. Further problems arise from the failures of the trial process adequately to cabin the experts.

⁸⁴ DSM-IV-TR, *supra* note **Error! Bookmark not defined.**, at xxxii-xxxiii, xxxvii.

⁸⁵ *Clark*, 126 S. Ct. at 2717-18.

⁸⁶ *See supra* text accompanying note 84.

⁸⁷ *See supra* text accompanying note **Error! Bookmark not defined.**

Note that the experts in *Clark* disagreed about his cognitive and moral capacities—that is, about whether he was legally insane. But insanity is a legal question to be resolved by a lay jury or judge. Why should we expect mental health experts to agree about this ultimate legal question about which they have no more expertise than lay jurors or judges? Indeed, it is for precisely this reason that expert witnesses in federal criminal trials are not permitted to offer ultimate legal conclusions about whether a defendant was legally insane.⁸⁸ Simple and sensible evidentiary rules like this would diminish the misleading tendencies of diagnostic information without the need for draconian rules that prevent defendants from defending themselves with relevant and reliable evidence.

The Court’s third and last argument justifying Arizona’s channeling rule considered the dangers that capacity evidence allegedly presents. The Court pointed out that opinions about the capacity for moral cognition or to form mens rea are inferential judgments “fraught with multiple perils,” including accurately determining the defendant’s mental state at the time of the crime and properly understanding the differences between psychological and legal judgments about capacity. Moreover, testimony about the defendant’s capacity to form mens rea is essentially “ultimate legal issue” testimony about which mental health experts have no special expertise. According to the Court, there is a real risk that the expert’s judgment about capacity will have an apparent authority that honest mental health professionals do not claim to have. States may reasonably address these dangers by channeling capacity evidence to the insanity issue and placing the burden of persuasion on the defendant.

The Court’s basic critique of capacity evidence is sound. As discussed previously,⁸⁹ evidence about a defendant’s capacity to form mens rea is extremely problematic and often lacks a solid clinical or scientific foundation. The criminal law would be better off if capacity evidence were strictly limited or even prohibited. Again, however, the Court’s argument proves too much, since capacity evidence, despite its limitations, is almost everywhere admissible to address every other question in criminal and civil law to which it may be relevant.⁹⁰ Why permit its limitation here, when so much is at stake for the criminal defendant? Furthermore, permitting an expert to give an opinion about legal insanity—which experts are permitted to do almost everywhere—is *a fortiori* ultimate issue testimony, which the Court rejects concerning the mens rea issue.

Moreover, the Court’s capacity argument was marred by providing all its examples from the context of legal insanity, yet again blurring the two doctrines. The ultimate issue of legal insanity is considerably less factual than the ultimate issue of mens rea. The former does depend on a factual understanding of the defendant’s mental state at the time of the crime, but the finder of fact ultimately must make a normative moral judgment that is not straightforwardly factual because the borders of all insanity tests are fuzzy and open to interpretation. In contrast, whether a defendant formed a requisite mens rea is, with few exceptions,⁹¹ a purely factual question. As Justice Kennedy said about Clark in his dissent, “Either Clark knew he was killing a police officer or he

⁸⁸ FED. R. EVID. 704(b).

⁸⁹ See *supra* Part III.C.

⁹⁰ California prohibits experts in criminal trials from offering an opinion about whether the defendant had the capacity to form mens rea. CAL. PENAL CODE § 22(b) (1998).

⁹¹ For example, the “premeditation” standard that in many jurisdictions makes an intentional killing a first degree murder often involves some degree of normative evaluation. Hate motivation for hate crimes is another example.

did not.”⁹² The question is not about moral responsibility but about empirical fact; therefore, the expert is less likely to mislead about mens rea than about legal insanity. Again, this would argue for channeling the evidence in the opposite direction.

The central problem with the majority’s channeling argument is one of basic fairness. When a citizen is being threatened with the most awesome exercise of state power—criminal blame and punishment—it seems that we should be most permissive in allowing that citizen to defend himself with the same relevant and reliable evidence allowed without limitation in all other legal contexts. What is it about the criminal trial that drives the Court to tolerate defense handicaps it would not tolerate in any other arena?

It is instructive to compare the Court’s approach to psychiatric evidence in another case, *Barefoot v. Estelle*,⁹³ which involved a due process challenge to the admission of a prosecution expert’s psychiatric opinion about a defendant’s future dangerousness. The opinion was elicited in the death penalty phase of a capital trial and was based entirely on the psychiatrist’s response to the prosecutor’s hypothetical questions. *Barefoot* claimed, with the support of all the relevant mental health organizations as amici, that clinical predictions by mental health professionals of a defendant’s future dangerous conduct were so inaccurate that they would inevitably lead to erroneous sentences.⁹⁴ Although capital punishment was at stake, the Court upheld the admission of such predictions and ruled that the deficiencies of the testimony went simply to its weight and could be addressed by cross-examination.⁹⁵

All the same arguments the Court mounts in *Clark* against mental-disease and capacity evidence apply *a fortiori* to predictions of dangerousness, and there is no reason the same remedy that saved the potentially misleading evidence in *Barefoot*—vigorous cross-examination to expose its defects—could not apply equally in *Clark*. How can it be fair to let the state present problematic mental health evidence to support imposition of capital punishment but deny the defendant the right to use similar evidence to defend himself against a charge that he even committed a crime?

The Court in *Clark* never satisfactorily addresses this basic issue of fairness, or the extraordinary degree to which these kinds of channeling rules compromise the right of citizens to demonstrate their innocence. *Clark* may or may not have believed that Officer Moritz was an alien—indeed, the trial judge concluded that he did not. If the trial judge really did not consider the relevant and reliable expert evidence that *Mott* excluded, *Clark* did not have a fair chance to cast reasonable doubt on the mens rea necessary to convict him of first degree murder.

Exclusion of Mental Abnormality Evidence to Negate Mens Rea is Wrong

I will not linger over whether states may constitutionally abolish mens rea. This has been an unresolved question of constitutional law ever since the emergence of the regulatory state and the

⁹² *Clark v. Arizona*, 126 S. Ct. 2709, 2743 (2006) (Kennedy, J., dissenting).

⁹³ 463 U.S. 880 (1983).

⁹⁴ *Id.* at 898-901.

⁹⁵ *Id.* at 903.

concomitant growth and acceptability of strict liability crimes.⁹⁶ Nonetheless, it has been clear at least since *Morrisette v. United States*⁹⁷ that mens rea is essential to criminal law, and that states and Congress bear a heavy burden of justifying departures from it. Nevertheless, the Supreme Court has never crossed the line from “essential” to “constitutionally required,” even for common law crimes at the core of criminal law whose mens rea elements predated the Constitution. Part of the reason the Supreme Court has never done this is that no legislature has ever been so bold as to purport to convert a serious common law crime into a strict liability crime. Consequently, this difficult constitutional question has lingered at the edges of justiciability, drawing the considerable attention of academics but not courts.⁹⁸ Nevertheless, in an era in which criminal justice reform aims at making mens rea requirements more robust, the opposite trend in the mental health context is anomalous and wrong.

That mens rea must be part of any constitutionally legitimate criminal system is not so much an historical conclusion as a moral one. It is simply immoral for the state to punish people whose mental disorder prevented them from forming mens rea as if they had in fact formed it. For those with and without mental disorder, mens rea expresses the defendant’s attitudes towards the rights and interests of fellow citizens.⁹⁹ Different culpable mental states justify different levels of punishment consistent with different degrees of desert. Without mens rea, no blame and punishment would be justified at all, at least for core crimes. The causal roots of that immorality may be in part biological,¹⁰⁰ but however our concern with mens rea began, it is now a fundamental feature of our moral practices. One need not be a neo-Darwinist to understand that if “fundamental rights” or “self-evident principles” mean anything, they must include guarantees that the state should satisfy its duties under the consent-to-be-governed bargain.

Mens rea is so crucial to fair ascriptions of blame and imposition of punishment that we should be wary of attempts to dilute it in the mental health context of any other by redefinitions of elements and affirmative offenses or by evidentiary rules. Although the criminal law, as a human institution, can never guarantee perfect justice and must often balance competing moral and practical considerations, the risk of error should seldom be shifted to the defendant, thus risking wrongful conviction or wrongful conviction of a more serious crime, unless there are supremely good reasons for doing so. The presumption should be against redefinitions and evidentiary rules that undermine the values *Winship* protects.

⁹⁶ See, e.g., *United States v. Dotterweich*, 320 U.S. 277 (1943) (upholding strict liability and punishment for up to a year for shipping misbranded pharmaceuticals); *United States v. Balint*, 258 U.S. 250 (1922) (upholding strict liability and punishment for up to five years for selling controlled substances without the order form required by the Narcotic Act of 1914).

⁹⁷ 342 U.S. 246 (1952).

⁹⁸ Compare Herbert L. Packer, *Mens Rea and the Supreme Court*, 1962 SUP. CT. REV. 107 (suggesting constitutional limits to the legislative abolition of mens rea), with Louis Bilionis, *Process, the Constitution, and Substantive Criminal Law*, 96 MICH. L. REV. 1269, 1278-79 (1998) (calling the notion that individual blameworthiness is central to criminal law a “myth”). Complicating the debate is the impact of punishment theory. Some commentators have argued that a retributionist might, in some circumstances, care less about mens rea than a rehabilitationist. See, e.g., Kenneth W. Simons, *When Is Strict Criminal Liability Just?*, 87 J. CRIM. L. & CRIMINOLOGY 1075 (1997). Though this detour is beyond the scope of this article, we must mention here that neither of us, who count ourselves as essentially in the retributionist camp, shares the idea that mens rea is separable from blameworthiness. On the contrary, as we discuss *infra* in the text accompanying notes **Error! Bookmark not defined.** to **Error! Bookmark not defined.**, the very idea of blameworthiness likely has its roots in our evolutionary ability to recognize intentionality.

⁹⁹ See *infra* Part III.C.

¹⁰⁰ We address this speculation in Part VI.B.

The Affirmative Defense of Legal Insanity

In the course of upholding the constitutionality of Arizona's extremely narrow legal insanity test in *Clark*, the Supreme Court noted that it had never ruled whether an affirmative defense of legal insanity is required by due process.¹⁰¹ A few older state court opinions held abolition of the defense unconstitutional as a violation of fundamental fairness and the right to a jury trial.¹⁰² In the later part of the 20th Century, five states have abolished the insanity defense, and in four of the states the state supreme court upheld the abolition.¹⁰³ In the fifth, the Nevada Supreme Court held that abolition violated due process under both the state and federal constitutions.¹⁰⁴ In 2018, however, Alaska became the first state in the 21st Century to abolish the defense,¹⁰⁵ so now there are five again.

Finally, in *Kahler v. Kansas*,¹⁰⁶ the Supreme Court held that the Constitution does not require a state to provide an independent, affirmative defense of insanity. It was sufficient, the majority held, if the jurisdiction permitted the defendant to introduce evidence of mental abnormality to negate the mens rea for the crime charged, a practice known as the "mens rea" alternative to the insanity defense. The Court opined that this rule in effect introduced a cognitive test defense of legal insanity to defeat the prosecution's prima facie case. In so doing, the Court implicitly conceded that mental disorder was sometimes relevant to fair blame and punishment. I have long thought that the Constitution did not require an independent affirmative defense of legal insanity if an alternative that would do equal justice could be found. The Court thought it had succeeded in finding one. The difficulty, however, is that this alternative is a practical failure that denies justice.

The moral and legal necessity of some form of an insanity defense or a reasonable alternative is easily explained. The capacity for rationality and, more controversially, for self-control are the touchstones of responsibility and their lack explains almost all the partial and complete affirmative defenses.¹⁰⁷ In some severe cases, mental abnormality markedly diminishes those capacities and thus the defendant is not a morally responsible agent who deserves blame and punishment. It is for similar reasons based on developmental immaturity and intellectual disability (itself a form of mental abnormality) that some juveniles and some defendants with intellectual disability do not deserve the most severe punishments even if they retain some degree of criminal responsibility. There was an unedifying and inconclusive dispute between the *Kahler* majority and dissent about whether the history supported the existence for centuries of some form of an independent insanity

¹⁰¹ *Clark v. Arizona*, 126 S. Ct. 2721 n.20 (2006).

¹⁰² *Sinclair v. State*, 132 So. 581 (Miss. 1931); *State v. Strasburg*, 110 P. 1020 (Wash. 1910).

¹⁰³ *State v. Bethel*, 66 P.3d 840 (Kan. 2003); *State v. Herrera*, 895 P.2d 359 (Utah 1995); *State v. Searcy*, 798 P.2d 914 (Idaho 1990); *State v. Korell*, 690 P.2d 992 (Mont. 1984).

¹⁰⁴ *Finger v. State*, 27 P.3d 66 (Nev. 2001).

¹⁰⁵ *Alaska Stat. §§12.47.010(a), 12.47.020* (2018);

¹⁰⁶ 140 S. Ct. 1021, 589 U.S. __ (2020).

¹⁰⁷ Duress is the obvious counter-example. There may be nothing wrong with the defendant's capacities for rationality and self-control, but we simply cannot expect people to do the right thing in these cases. It is too much to ask of reasonable, law-abiding people when they are sufficiently threatened. It is such reasoning that leads some to think that duress is best classified as a justification. See, R. Jay Wallace, *RESPONSIBILITY AND THE MORAL SENTIMENTS* (1994). Westen & Mangiafaco (2003).

defense. It was unedifying and inconclusive because until well into the 20th Century, courts, legal commentators and treatise writers often confused mental disorder as negating the elements with mental disorder as the basis of an affirmative defense. That confusion was apparent in both the *Clark* and *Kahler* majorities. Thus, no clear answer could be expected from the historical sources. What is clear, however, is that in one form or another, mental disorder has had exculpatory force for since the late middle ages in English law and since the founding in U.S. law.¹⁰⁸

A similar baseline principle explains the many competence doctrines employed in the criminal justice process. This Court has long recognized that at every stage justice demands that some people with severe mental abnormalities must be treated differently from those without substantial mental impairment because some impaired defendants are incapable of reason and understanding in a specific context. Competence to stand trial,¹⁰⁹ competence to plead guilty and to waive counsel,¹¹⁰ competence to represent oneself,¹¹¹ and competence to be executed,¹¹² are all examples in which the Constitution requires such special treatment. It is unfair and offensive to the dignity of criminal justice to treat people without understanding as if their understanding was unimpaired. Evidence of mental disorder is routinely introduced in all these contexts to determine if the defendant must be accorded special treatment.

As noted, the states that have adopted this *mens rea* alternative have implicitly accepted this baseline principle. In what follows, I first address the “*mens rea* alternative” and then consider another alternative, taking mental disorder into account at sentencing. Neither works, however, and no jurisdiction should accept the Supreme Court’s permission to use one of them instead of an affirmative defense of legal insanity.

The negation of *mens rea* and the affirmative defense of legal insanity are different claims that preclude criminal liability by different means. The former denies the *prima facie* case of the particular crime charged; the latter is an affirmative defense that precludes liability in those cases in which the *prima facie* case is established. The post-verdict consequences are also different. The former leads to outright acquittal; the latter results in some form of involuntary civil commitment. Although in some cases the same mental disorder evidence may be used to prove the two different claims, they are not equivalent.

The primary reason that permitting a defendant to introduce evidence of mental disorder to negate *mens rea* cannot justly replace the affirmative defense of legal insanity is that the *mens rea* alternative is based on a mistaken view of how severe mental disorder affects human behavior. In virtually all cases, mental disorder, even severe disorders marked by psychotic symptoms such as delusions and hallucinations, does not negate *mens rea*.¹¹³ It is difficult to prove a negative, but cases, especially those involving serious crime, in which most or all *mens rea* is negated are rare to the vanishing point. Rather, mental disorder affects a person’s motivations or reasons for

¹⁰⁸ Brief for 290 ...

¹⁰⁹ See *Drope v. Missouri*, 420 U.S. 162 (1975).

¹¹⁰ See *Godinez v. Moran*, 509 U.S. 389 (1993).

¹¹¹ See *Indiana v. Edwards*, 534 U.S. 164 (2008).

¹¹² See *Ford v. Wainwright*, 477 U.S. 399 (1986), *Panetti v. Quarterman*, 551 U.S. 930 (2007), *Madison v. Alabama*, 139 S. Ct. 718 (2019).

¹¹³ Stephen J. Morse, *Undiminished Confusion in Diminished Capacity*, 75 J. CRIM. L. & CRIMINOLOGY 1, 16 (1984); Morse, *Mental*, *supra*, at 933.

committing the criminal acts. A mentally disordered defendant's irrationally distorted beliefs, perceptions or desires typically and paradoxically give him the motivation to form the *mens rea* required by the charged offense. Mental disorder rarely interferes with the ability to perform the necessary actions to achieve irrationally motivated aims. In cases of self-regulation problems, the defendant does form the *mens rea* but lacks substantial capacity to conform his conduct to the law.

Consider the following, typical examples, beginning with Daniel M'Naghten himself.¹¹⁴ M'Naghten delusionally believed that the ruling Tory party was persecuting and intended to kill him.¹¹⁵ As a result, he formed the belief that he needed to assassinate the Prime Minister, Peel, in order to end the threat. He therefore formed the intention to kill Peel. Thus, M'Naghten would have been convicted of murder if a defense of legal insanity was not available. Indeed, his case has come to stand for one of the "rules" enunciated by the House of Lords – that a defendant should be acquitted on grounds of insanity if he "was laboring under such a defect of reason, from a disease of the mind, as not to know the nature and quality of the act he was doing, or if he did know it, that he did not know he was doing what was wrong."¹¹⁶

For a more contemporary example, consider the case of Ms. Andrea Yates, the Texas woman who drowned her five children in a bathtub. She delusionally believed that she was corrupting her children and that unless she killed them, they would be tortured in Hell for all eternity.¹¹⁷ She therefore formed the intention to kill them. Indeed, she planned the homicides carefully. Ms. Yates was nonetheless acquitted by reason of insanity because she did not know that what she was doing was wrong. Even if she cognitively recognized that her conduct violated the law of Texas and that she would be arrested, she was deeply convinced that the homicides were necessary to assure the eternal well-being of her children under the circumstances. In her psychotic thinking, everyone else would approve of her conduct as justified if they knew what she knew.

For a final example, suppose an offender hallucinates that he is hearing God's voice or delusionally believes that God is communicating with him and that God is commanding him to kill.¹¹⁸ If the offender kills in response to this "command hallucination" or delusion, he surely forms the intent to kill to obey the divine decree. Nonetheless, it would be unjust to punish this defendant because he, too, does not know right from wrong given his beliefs for which he is not responsible.

In all three cases, one could also claim that the defendant did not know what he or she was doing in a fundamental sense because the most material reason for action, what motivated them to form *mens rea*, was based on a delusion or hallucination that was the irrational product of a disordered mind. Finally, in all three cases the defendant's instrumental rationality, the ability rationally to achieve one's ends, was intact despite their severe disorders. They were able effectively to carry out their disordered plans.

¹¹⁴ M'Naghten's Case, 8 Eng. Rep. 718 (1843).

¹¹⁵ Richard Moran, *Knowing Right From Wrong: The Insanity Defense Of Daniel McNaughtan* 10 (2000).

¹¹⁶ 8 Eng. Rep. at 722.

¹¹⁷ Deborah W. Denno, *Who is Andrea Yates? A Short Story About Insanity*, 10 DUKE J. GENDER L. & POL'Y 1 (2003).

¹¹⁸ See, e.g., *People v. Serravo*, 823 P. 2d 828, 830 (Colo. 1992) (en banc).

There are very few contemporary data about the operation of the insanity defense and virtually none about the operation of the *mens rea* alternative. Montana is the only state for which there is a systematic study of mental disorder claims pre- and post-abolition of the insanity defense. The picture is complicated, but in brief, the number of cases, the types of defendants and the types of crimes did not change. There were two major effects, however. Under the *mens rea* alternative, more defendants were convicted and the number of defendants found incompetent to stand trial increased markedly.¹¹⁹ The increase in convictions in Montana demonstrates that abolition of the insanity defense does, in fact, expose severely mentally ill offenders to unjust punishment. Moreover, the rise in the number of defendants found incompetent to stand trial who would previously have been found competent and acquitted suggests that an incompetency finding is being used as a tool for diversion in cases involving less serious charges that likely would have led to stipulated insanity acquittals under the pre-abolition statute. This is also objectionable.

Although Kahler seemed to have had the *mens rea* required for the charged offense of capital murder, his expert evidence established that he suffered from a number of mental disorders, including a major mental disorder, severe depression. As a result, the expert opined, Kahler's perception and judgment were so distorted that he may have become dissociated from reality at the time of the crime. The expert also testified that Kahler could not refrain from his conduct. Because an insanity defense was not available and Kahler's conduct met the criteria for capital murder, his conviction for the most serious crime in the criminal law was improperly a foregone conclusion. Justice Breyer's dissent in *Kahler* fully understood the failure of the *mens rea* alternative as a result of how mental disorder affects *mens rea*, but to no avail.¹²⁰

To further understand the injustice of the *mens rea* alternative, consider a case in which *mens rea* may plausibly be negated. Suppose a defendant charged with murder claims that he delusionally believed that his obviously human victim of a shooting was in reality the devil and not a human being.¹²¹ If his beliefs were genuine, the defendant did not intentionally kill a human being. Indeed, in a *mens rea* alternative jurisdiction, he could not be convicted of purposely, knowingly or recklessly killing a human being because his delusional beliefs negated all three mental states. After all, he fully believed that he was shooting at the devil, not a human being. The defendant would be convicted of negligent homicide, however, because the standard for negligence is objective reasonableness and the motivating belief was patently unreasonable.

Of course, convicting the severely disordered defendant of a crime based on a negligence standard is fundamentally unjust, as even Mr. Justice Holmes recognized in his rightly famous essays on the common law.¹²² The defendant's unreasonable mistake was not an ordinary mistake

¹¹⁹ Lisa Callahan et al., *The Hidden Effects of Montana's "Abolition" of the Insanity Defense*, 66 PSYCHIATRIC Q. 103 (1995). Neither change is desirable. For the reasons given in Part II, *supra* note ___, conviction is unjust in any case in which the defendant should have been acquitted by reason of insanity.

¹²⁰ The facts in *Kahler* were not nearly as favorable to an insanity defense as those in an earlier case in which the Supreme Court denied certiorari, *Delling v. Idaho*, 568 US 1038 (2012). *Delling* was clearly delusional and clearly formulated his plan to kill intentionally based on those delusions. Although *Delling* possessed the *mens rea* for murder and was found guilty, the trial judge opined that *Delling* did not know right from wrong. It was the perfect case to illustrate the failure of the *mens rea* alternative, but over a written dissent from Justice Breyer, the Court ducked the issue. Justice Breyer used the same argument he adopted in *Kahler*.

¹²¹ See Richard J. Bonnie, *The Moral Basis of the Insanity Defense*, 69 A.B.A. J. 194, 195-96 (1983).

¹²² Oliver Wendell Holmes, Jr., *The Common Law* 50-51 (Little, Brown & Co. 1945) (1881).

caused by inattention, carelessness or the like. Defendants are responsible for the latter because we believe that they had the capacity to behave more reasonably by being more careful or attentive. In contrast, the hypothetical defendant's delusional "mistake" was the product of a disordered mind and thus he had no insight and no ability to recognize the gross distortion of reality. He was a victim of his disorder, not someone who deserves blame and punishment as a careless perpetrator of involuntary manslaughter. He does not deserve any blame and punishment, and only the defense of legal insanity could achieve this appropriate result. Paradoxically, such a defendant's potential future dangerousness if he remains deluded would be better addressed by an insanity acquittal and indefinite involuntary commitment, a practice this Court has approved,¹²³ than by the comparatively short, determinate sentences for involuntary manslaughter.

Thus, the *mens rea* alternative is not an acceptable replacement or substitute for the insanity defense. Only in the exceedingly rare case in which mental disorder negates all *mens rea* would the equivalent justice of a full acquittal be achieved, albeit for a different reason. But again, this is the rarest of cases. Most legally insane offenders form the *mens rea* required by the definition of the charged offense and only the defense of legal insanity can respond justly to their lack of blameworthiness.

Consideration of mental disorder for purposes of assessing both mitigation and aggravation is a staple of both capital and non-capital sentencing, but it is no substitute for the affirmative defense of legal insanity. On moral grounds, it is unfair to blame and punish a defendant who deserves no blame and punishment at all, even if the offender's sentence is reduced. Blaming and punishing in such cases is unjust, full stop. Sentencing judges might also use mental disorder as an aggravating consideration because it might suggest that the defendant is especially dangerous as a result. Thus, sentences of severely mentally ill offenders might be enhanced. Again, injustice would result, and public safety would not be protected as well as an indeterminate post-acquittal commitment would achieve. Third, unless a sentencing judge is required by law to consider mental disorder at sentencing, whether the judge does so will be entirely discretionary. Again, this is a potent, potential source of injustice. Finally, in some jurisdictions the sentencing judge in cases of serious crime has little if any sentencing discretion, so even the most sensitive, sympathetic judge would be handcuffed doing justice. In short, only a required insanity defense would insure that arguably blameless mentally disordered offenders have an opportunity to establish that state blame and punishment are not justified.

II. THE EXTERNAL CHALLENGES TO RESPONSIBILITY FROM BEHAVIORAL NEUROSCIENCE & GENETICS

As millennia of philosophizing attest, there are challenging questions about the existence, source and content of meaning, morals and purpose in human life, but present and foreseeable neuroscience will neither obliterate nor resolve them. Neuroscience, for all its astonishing recent discoveries, raises no new challenges in these domains. It poses no unique threat to our life hopes or to our ability to decide how to live and how to live together. The supposed challenges were best summed up by an editorial warning in *The Economist*:

¹²³ Jones v. U.S., 463 U.S. 354 (1983)

“Genetics may yet threaten privacy, kill autonomy, make society homogeneous and gut the concept of human nature. But neuroscience could do all of these things first.”¹²⁴

The primary quarries of those who think that neuroscience poses a challenge to meaning, morals and purpose are the related concepts of responsibility and desert, especially as they play a role in criminal law. After all, responsibility and desert are intrinsic features of present moral and criminal legal concepts, practices and institutions, including the imposition of punishment. Most of those who challenge responsibility and desert think that these concepts are philosophically questionable and lead to primitive, pre-scientific practices, such as overly harsh punishments. Responsibility and desert are also intrinsic to civil law, but virtually none of these challengers considers how their views would affect desert theories in contracts, torts and property law, for example. Although critics have a duty to embed their criticisms of criminal law in a wider understanding of the implications of the criticisms, this chapter will nonetheless engage with the dominant critique by limiting itself to the potential effect of neuroscience on criminal law and moral responsibility more generally.

As is well known, the primary challenges neuroscience allegedly present to responsibility, desert, and retributive justifications of punishment are the threat from determinism and the specter of the person as simply a “victim of neuronal circumstances” (VNC)¹²⁵ or “just a pack of neurons” (PON).¹²⁶ Allegedly, no one is responsible for any of his behavior and no one deserves a proportionate response to his behavior either because determinism is true and inconsistent with responsibility or because mental states are epiphenomenal and we are therefore not the sort of creatures that can be guided by reason. I have argued repeatedly that no such soul-bleaching outcome as *The Economist* dreads is remotely justified by neuroscience at present,¹²⁷ or by any other science for that matter. This chapter re-iterates some of these arguments and responds to the challenge of the new determinism, termed “hard incompatibilism” (HI) by its proponents.

A criticism based on the claim that no one is genuinely responsible for wrong-doing is an external criticism because it assumes that a partial or wholly responsibility-based system is incoherent because it rests on a fundamental mistake. The determinist and VNC/PON challenges are external.

As proponents of these challenges fully recognize, they provide no basis whatsoever for internal reform of a responsibility-based criminal justice system. Determinism is not selective or partial. If it grounds a moral or legal practice, it applies to all who come within the practice and it cannot make the distinctions concerning guilt and desert that are at the heart of criminal justice. Relatedly, determinism is not the equivalent of a folk-psychological compulsion excuse. If it were, then everyone would be compelled and all would be excused. This, too, would fail to make the distinctions our criminal justice system makes. The same is true of the VNC/PON challenge. It denies the possibility of responsibility, applies to all and would entail abandoning the moral responsibility distinctions our system now makes.

¹²⁴ *The Economist* (2002)

¹²⁵ Greene & Cohen 2006

¹²⁶ Crick 1994

¹²⁷ *See e.g.*, Morse 2011a.

As is apparent, the consequences of accepting these critiques would be nothing short of radical and completely unmoored from standard views of responsibility. This is no reason not to adopt such changes if they are justified, but it seems clear that the burden of proof should clearly be placed on the proponents of radical change. They seek to abandon a system that has evolved for centuries, that is in accord with commonsense and with moral, political and legal theories that are widely endorsed, and that seems to work, albeit imperfectly.¹²⁸ These challenges cannot borrow internal reformist changes because they are total critiques. If no one is really morally responsible and, consequently, no one deserves any blame or punishment, then only a radically new system of social control is justified. What reason would there be to replace the time-tested system based on an unresolvable metaphysical argument (incompatibilist) or an unproven scientific (VCN/PON) claim?

Proponents of the challenges often claim that responsibility and desert theorists should bear the burden of persuasion because they are justifying harsh treatment. But this counterclaim begs the question. It assumes that all punishment is unjustifiably harsh because no one deserves any punishment at all, but that is precisely what desert theorists deny. More important, until the brave new world the radical challenges propose is fully described, it is not clear that it is either workable or that it will not be even more inconsistent with human flourishing than the current system based on responsibility and desert.

The Determinist Challenge

In one form or another, the challenge from determinism to “free will” and responsibility has been mounted for millennia. Neuroscience poses no new challenge in this respect. No science can prove the truth of determinism (although some seem to push intuitions harder in that direction) and the answers to the determinist threat are the same to neuroscience as they have been to any of its also deterministic predecessors such as psychodynamic psychology or genetics. Neuroscience poses no new assault on meaning, morals and purpose if these aspects of life are at risk from the truth of determinism.

The primary reason people care about the issue is because it allegedly underwrites conceptions of responsibility, agency, and dignity that are crucial to our image of ourselves and to our political and legal practices and institutions. If responsibility is undermined, meaning, morals and purpose may indeed be undermined. The alleged incompatibility of determinism and free will and responsibility is therefore a foundational metaphysical and moral issue. Determinism is not a continuum concept that applies to various individuals in various degrees. To the best of our knowledge, there is no partial or selective determinism. If the universe is deterministic or something quite like it, responsibility is possible or it is not. If human beings are fully subject to the causal laws of the universe, as a thoroughly physicalist, naturalist worldview holds, then many philosophers claim that “ultimate” responsibility is impossible.¹²⁹ On the other hand, plausible “compatibilist” theories suggest that for many different reasons, responsibility is possible in a deterministic universe even if no human being has the god-like contra-causal, ultimate freedom

¹²⁸ See MacIntyre, ch. 13 (2007) for a defense of this stance).

¹²⁹ See e.g., Pereboom 2001; Strawson 1989.

that incompatibilists require.¹³⁰ Compatibilists hold that human beings possess whatever degree of freedom or control is necessary for responsibility.

Compatibilism is the dominant view among philosophers of responsibility and it most accords with common sense. Much about who we are and what we do is not a product of our rational action, including our genetic endowment, early environment and the opportunities that present themselves to us. Luck clearly plays an immense role in human life.¹³¹ Nonetheless, the compatibilist claims that we retain sufficient capacity to be guided by reason and to choose otherwise when we act or omit.¹³² This is our ordinary view of ourselves and our agency. When any theoretical notion contradicts common sense, the burden of persuasion to refute common sense must be very high and no metaphysics that denies the possibility of responsibility exceeds that threshold.

There seems no resolution to the incompatibilism/compatibilism debate in sight, but our moral and legal practices do not treat everyone or no one as responsible. Determinism cannot be guiding our practices. If one wants to excuse people because they are genetically and neurally determined or determined for any other reason, one is committed to negating the possibility of responsibility for everyone.

Our criminal responsibility criteria and practices have nothing to do with determinism or with the necessity of having so-called “free will.”¹³³ The metaphysical libertarian capacity to cause one’s own behavior uncaused by anything other than oneself, the strongest conception of free will, is neither a criterion for any criminal law doctrine nor foundational for criminal responsibility. Criminal responsibility involves evaluation of intentional, conscious, and potentially rational human action. And few participants in the debate about determinism and free will and responsibility argue that we are not conscious, intentional, potentially rational creatures when we act. The truth of determinism does not entail that actions and non-actions are indistinguishable and that there is no distinction between rational and non-rational actions or compelled and uncompelled actions. Our current responsibility concepts and practices use criteria consistent with and independent of the truth of determinism.

In short, the hard determinist incompatibilists (including the hard incompatibilists) are proposing a radical revision of practices and institutions that have evolved for centuries on the basis of an unresolvable metaphysical claim and in the face of commonsense. The case simply is not proven. But, in my view, because this viewpoint makes no obvious internal mistakes, it must be taken seriously, at least in so far as proponents offer a different vision of social ordering rather than just a pure metaphysical argument. Section IV of this chapter returns to that vision. Rather than attempt to resolve the unresolvable, it assumes that a compatibilist also makes no internal mistakes and turns to a comparative analysis of the world as we know it and the world that hard incompatibilist is offering us. But first let us consider the most radical claim that neuroscience arguably presents: VNC/PON.

¹³⁰ Moore 2016; Vihvelin 2013; Wallace 1994.

¹³¹ Frank, 2016.

¹³² Moore, 2016.

¹³³ Morse 2007

The Truly Radical Challenge to Meaning, Morals and Purpose

This section addresses the hyper-reductive claim that the new sciences, and especially neuroscience, will cause a paradigm shift in our view of ourselves as agents who can direct our own lives and can be responsible by demonstrating that we are “merely victims of neuronal circumstances” or a “pack of neurons” (or some similar claim that denies human agency). This claim holds that we are not the kinds of intentional creatures we think we are. If our mental states play no role in our behavior and are simply epiphenomenal, then traditional notions of agency and responsibility based on mental states and on actions guided by mental states would be imperiled. More broadly, neurons, or even a big pack of them like the connectome, do not have meaning, morals and purpose. They are just biophysical mechanism. They are not agents. Meaning, morals and purpose are products of people, not mechanisms. But is the rich explanatory apparatus of agency and intentionality simply a post hoc rationalization that the brains of hapless homo sapiens construct to explain what their brains have already done? Will the criminal justice system as we know it wither away as an outmoded relic of a prescientific and cruel age? If so, criminal law is not the only area of law in peril. What will be the fate of contracts, for example, when a biological machine that was formerly called a person claims that it should not be bound because it did not make a contract? The contract is also simply the outcome of various “neuronal circumstances.”

Before continuing, we must understand three things. The compatibilist metaphysics discussed above does not save agency, responsibility, morals, meaning and purpose if the radical claim is true. If determinism is true, two states of the world concerning agency are possible: agency and all that it entails exists or it does not. Compatibilism assumes that agency is true because it holds that agents can be responsible in a determinist universe. It thus essentially begs the question against the radical claim. If the radical claim is true, then compatibilism is false because no responsibility is possible if we are not agents. It is an incoherent notion to have genuine responsibility without agency. Second, those forms of hard determinism that accept agency but not responsibility are equally false if the radical claim is true. Finally, this challenge is not new to neuroscience. This type of reductionist speculation long precedes the contemporary neuroscientific era fueled by non-invasive imaging. Neuroscience appears to at last provide the ability to prove the claim by discovering the underlying neural mechanisms that are doing all the work that mental states allegedly do, but the claim is not novel to it. The question remains, however: Is the radical claim true?

Given how little we know about the brain-mind and brain-mind-action connections—we do not know how the brain enables the mind and action¹³⁴ —to claim that we should radically change our conceptions of ourselves and our legal doctrines and practices based on neuroscience is a form of “neuroarrogance.” It flies in the face of common sense and ordinary experience to claim that our mental states play no explanatory role in human behavior and thus the burden of persuasion is firmly on the proponents of the radical view, who have an enormous hurdle to surmount. Although I predict that we will see far more numerous attempts to use the new sciences to challenge traditional legal and common sense concepts, I have elsewhere argued that for conceptual and scientific reasons, there is no reason at present to believe that we are not agents.¹³⁵

¹³⁴ Adolphs (2015)

¹³⁵ Morse 2011a: 543-54; 2008

In particular, I can report based on earlier and more recent research that the “Libet industry” appears to be bankrupt. This was a series of overclaims about the alleged moral and legal implications of neuroscientist Benjamin Libet’s findings, which were the primary empirical neuroscientific support for the radical claim. This work found that there is electrical activity (a readiness potential) in the supplemental motor area of the brain prior to the subject’s awareness of the urge to move his body and before movements occurred. This research and the findings of other similar investigations¹³⁶ led to the assertion that our brain mechanistically explains behavior and that mental states play no explanatory role. Recent conceptual and empirical work has exploded these claims.¹³⁷ Moreover, even hard incompatibilists (see Section IV below) agree that the conceptual and empirical criticisms of the implications of the Libetian program are decisive.¹³⁸ In short, I doubt that the Libet industry will emerge from whatever chapter of the bankruptcy code applies in such cases. It is possible that we are not agents, but the current science does not remotely demonstrate that this is true. Just for completeness, I should add that similar radical views from psychology, such as “the illusion of conscious will”¹³⁹ and the “automaticity juggernaut,”¹⁴⁰ suffer from the same defects.¹⁴¹ The burden of persuasion is still firmly on the proponents of the radical view.

Most important, contrary to its proponents’ claims, the radical view entails no positive agenda. If the truth of pure mechanism is a premise in deciding what to do, no particular moral, legal, or political conclusions follow from it. This includes the pure consequentialism that Greene and Cohen incorrectly think follows. The radical view provides no guide as to how one should live or how one should respond to the truth of reductive mechanism. Normativity depends on reason, and thus the radical view is normatively inert. Reasons are mental states. If reasons do not matter, then we have no reason to adopt any particular morals, politics, or legal rules or to do anything at all.

Suppose we are convinced by the mechanistic view that we are not intentional, rational agents after all. (Of course, what does it mean to be “convinced” if mental states are epiphenomenal? Convinced usually means being persuaded by evidence and argument, but a mechanism is not persuaded, it is simply physically transformed. But enough.) If it is really “true” that we do not have mental states or, slightly more plausibly, that our mental states are epiphenomenal and play no role in the causation of our actions, what should we do now? If it is true, we know that it is an illusion to think that our deliberations and intentions have any causal efficacy in the world. We also know, however, that we experience sensations—such as pleasure and pain—and we care about what happens to us and to the world. We cannot just sit quietly and wait for our brains to activate, for determinism to happen. We must, and will, deliberate and act. And if we do not act in accord with the “truth” that the radical view suggests, we cannot be blamed. Our brains made us do it.

Even if we still thought that the radical view was correct and standard notions of agency and all that it entails were therefore impossible, we might still believe that the law would not necessarily have to give up the concept of incentives. Indeed, Greene and Cohen concede that we would have

¹³⁶ See, e.g., Soon and others 2008

¹³⁷ Mele 2009, 2014; Moore 2011; Nachev and Hacker 2015; Schurger and others 2012; Schurger and Uithol 2015).

¹³⁸ (Pereboom and Caruso).

¹³⁹ Wegner, 2002

¹⁴⁰ Kihlstrom, 2008

¹⁴¹ Morse 2011a.

to keep punishing people for practical purposes.¹⁴² The word “punishment” in their account is a solecism, because in criminal justice it has a constitutive moral meaning associated with guilt and desert. Greene & Cohen would be better off talking about positive and negative reinforcers or the like. Such an account would be consistent with “black box” accounts of economic incentives that simply depend on the relation between inputs and outputs without considering the mind as a mediator between the two. For those who believe that a thoroughly naturalized account of human behavior entails complete consequentialism, this conclusion might be welcomed.

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

As the eminent philosopher of mind and action, Jerry Fodor, reassured us:

[W]e have . . . no decisive reason to doubt that very many commonsense belief/desire explanations are—literally—true.

Which is just as well, because if commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we’re that wrong about the mind, then that’s the wrongest we’ve ever been about anything. The collapse of the supernatural, for example, didn’t compare; theism never came close to being as intimately involved in our thought and our practice . . . as belief/desire explanation is. Nothing except, perhaps, our commonsense physics—our intuitive commitment to a world of observer-independent, middle-sized objects—comes as near our cognitive core as intentional explanation does. We’ll be in deep, deep trouble if we have to give it up.

I’m dubious . . . that we *can* give it up; that our intellects are so constituted that doing without it (. . . *really* doing without it; not just loose philosophical talk) is a biologically viable option. But be of good cheer; everything is going to be all right.¹⁴³

Everything is going to be all right. Given what we know and have reason to do, the allegedly disappearing person remains fully visible and necessarily continues to act for good reasons, including the reasons currently to reject the radical view. We may be a pack of neurons, but that’s not all we are. We are not Pinocchios, and our brains are not Geppettos pulling the strings. And this is a very good thing. Ultimately, I believe that the radical view’s vision of the person, of interpersonal relations, and of society bleaches the soul. In the concrete and practical world we live in, we must be guided by our values and a vision of the good life. I do not want to live in the radical’s world that is stripped of genuine agency, desert, autonomy, and dignity. In short, in a world that is stripped of meaning, morals and purpose. For all its imperfections, the law’s vision of the person, agency and responsibility is more respectful and humane.

¹⁴² (Greene and Cohen 2006).

¹⁴³ (Fodor 1987: xii).

The Hard Incompatibilist Vision and Its Discontents

In recent years, a group of theorists who adopt a position called “hard incompatibilism” (HI) have recognized that if scholars care about the practical implications of a philosophical position, then they have an obligation to propose the details of psychology, politics and law that would follow.¹⁴⁴ HI does not deny the causal theory of action and that human beings are capable of being reasons-responsive. It accepts the possibility of meaning, morals and purpose because we are agents, but it denies that anyone is responsible for any conduct and that backward-looking blame and punishment is never justified. Value is preserved, including the distinction between right and wrong, on at least axiological grounds. The reactive attitudes such as anger, indignation and resentment that P.F. Strawson famously thought were appropriate¹⁴⁵ when a responsible agent violated a justified prohibition are never justified according to HI. We may feel them, but we should not. Instead, HI proposes that we should feel emotions such as sorrow and regret for harms that have been caused by the harmdoer’s actions. We can try to guide such people to do right in the future by attempting to persuade them to see the good reasons not to behave in such injurious ways, but responsibility plays no appropriate role in the social ordering. The role of law and social ordering practices to address harmful behavior is entirely forward-looking. The moral goals of the HI approach are protection, moral formation and reconciliation, goals that justify forward-looking blame but never deserved punishment.

Even though HI claims that no one is responsible, it properly concedes that dangerous people who commit crimes exist and must be controlled for the good of society at large. It recognizes the common criticisms of a purely consequential social control system, such as the potential for intervening more harshly than an individual’s dangerous conduct might suggest in order to further general deterrence. To avoid these problems, it adopts a public health-like system of quarantine or of lesser but still intrusive measures to control dangerous agents that are justified by societal self-defense. HI also borrows the proportionality principle from individual self-defense. That is, unnecessary intervention to achieve social safety is unjustified because it would be akin to using disproportionate force when one has a right to defend oneself. Because no one is responsible and deserves to suffer, these interventions should never be painful unless inflicting pain is ultimately unavoidable. Even then, the pain should be minimized consistent with the preventive goal of imposing it. Thus, pure quarantine would have to be under conditions of relatively comfortable confinement. Moreover, when a dangerous agent is being controlled, society has a duty to try to engage the agent’s reasons or otherwise to offer him a “fix” so that he no longer needs to be controlled. Finally, society has a duty to create the social conditions that decrease the probability of harmdoing generally.

One possible response in a world without responsibility is to treat people as good and bad bacteria, a story that treats potential human harmdoers no differently from any other organic (or inorganic) harmdoer.¹⁴⁶ It is a pure prediction/prevention scheme. Human animals must simply be managed to increase beneficent action and to decrease harmful action and any means calculated to achieve this end would be justified because bacteria have no rights or even interests. HI attempts to avoid this dystopian vision, by arguing that we are reasons-responsive and thus agents even

¹⁴⁴ (Pereboom and Caruso 2017).

¹⁴⁵ See Strawson (1962).

¹⁴⁶ (Morse 1999).

though we are not responsible.

HI seems benevolent and humane, but what would be the costs? My major response is that HI tries to have it both ways, that is, to preserve most of the desirable attitudes and practices associated with responsible agency, such as a right to liberty, autonomy and dignity, but to do so without a robust conception of agency. It is agency on the cheap, much as some compatibilists, such as Dennett and Pinker, try to have compatibilism on the cheap. They suggest that there must be responsibility and punishment but without adopting a robust conception of responsibility.¹⁴⁷ I shall suggest that HI provides only a pale simulacrum of agency and does not fully comprehend the implications of its view. Although I am a mixed theorist about punishment who believes that we do have basic desert and that retribution is at least a necessary precondition for fair blame and punishment, the argument of this section does not claim that we have basic desert. Rather, I try to stay internal to HI and to explore its implications assuming that we lack basic desert.

Let us begin by noting that neuroscience plays no necessary role in the HI program and by exposing an error about retributivism that seems to in part motivate HI's criticism of retributivism. The HI proposal is perfectly defensible independently of any of the alleged truths of neuroscience (or psychology). Any successful prediction or behavioral change method, for example, would provide the type of practical assistance crucial to the fair workings of the HI scheme.

Pereboom and Caruso, for example, tie excessive and harsh punishments to retributivism,¹⁴⁸ but most of the excessive and harsh criminal justice practices they and I and many other decry have been justified on consequential grounds such as deterrence and incapacitation. It is the retributivists who have had the resources to criticize such programs. HI claims that it, too, has the resources to avoid harsh and intrusive interventions, but does it? Now let us turn to the central difficulties.

It is crucial to understand that nothing is "up to" the person according to HI because no one has sufficient "control" to justify what HI terms "basic desert." Whether a person acts well or poorly is not up to them. Whether they feel joy or sorrow in response to their beneficence or harmdoing is not up to them. They cannot be morally praised or blamed for what they have already done or felt (although the action or emotional response may be judged good or bad itself) and they deserve nothing thereby. Whether they respond to moral address and concern is not up to them. Whether they are moved by the right sorts of moral and personal reasons is not up to them. Whether an intervention that addresses or bypasses their reasons is effective is not up to them. They may be reasons-responsive in theory, but reasons in HI's vision are simply mechanisms for changing behavior and are no different in principle from any other mechanism for doing so. Those who believe in responsibility, in contrast, believe that the capacity for reasons-responsiveness or something like it occupies a privileged position because such a capacity indicates that much of our behavior is up to us and we deserve appropriate responses to our action. To treat reasons as simply a mechanism, as HI implicitly does, suggests that we are simply creatures to be manipulated in the right ways to do the right thing rather than being genuinely autonomous agents. We are suspiciously close to the good bacteria/bad bacteria story, although bacteria do not have access to

¹⁴⁷ (Moore 2012).

¹⁴⁸ Pereboom and Caruso (2017)

reasons as a mechanism (or so we think).

Although whether we respond to reasons and what reasons we respond to is not up to us according to HI, we nonetheless have a right to liberty, autonomy and dignity. But why? Let us compare the current legal regime concerning responsibility and preventive state action, remembering that prevention of danger is the crux of the radical HI proposal. Anglo-American law in this respect is governed by what I have termed desert-disease jurisprudence.¹⁴⁹ The state may clutch an individual to promote social safety only if the agent deserves to be clutched because he has committed a crime (including attempts, of course) or if the agent is dangerous but is not responsible for his dangerous conduct. Agents who are dangerous but responsible must be left alone until they commit a crime, no matter how dangerous they are and no matter how certain they are to offend if they are not controlled. The best explanation for this jurisprudence that limits infringements of liberty as described is the respect the law grants responsible agents because they are responsible agents. We can act preventively in the absence of harmful action in the case of younger children, some people with mental disorder or intellectual disability, and some people with dementia precisely because they are not capable of rationality and thus are not responsible for themselves. But we cannot intervene with responsible agents unless they commit crimes because we respect their capacity to act well even if we are quite sure they will not. It is responsibility that underpins the full panoply of rights that protect liberty, autonomy and dignity.

If nothing is up to us, what is the source of our liberty, dignity and autonomy? It is understandable that we would have a sense of good (not harmful; beneficent) and bad (harmful; not beneficent) in the HI society, but why, for example, do we have a right to be at liberty in the absence of actual harmdoing? What is the meaning of autonomy when nothing is up to us? Why does the fact that reasons are sometimes successful mechanisms confer dignity on us if those reasons themselves are not up to us in any important way? In assessing reason-bypassing interventions, especially of a more radical type such as deep brain stimulation (DBS, which a completely experimental, unproven intervention for behavioral abnormalities), Pereboom and Caruso suggest that to respect autonomy, decisions should be left up to the subject as far as possible.¹⁵⁰ But why respect a decision that is not really up to the subject? In short, the agency HI concedes is a simulacrum of what we usually mean by agency.

Consider a response of the agent who has acted harmfully and feels no sorrow or regret. He claims that doing it was not up to him and neither was lack of what would be the appropriate emotions. He is now addressed by the agents of social control with moral concern and guidance with the goal of future protection, reconciliation and moral formation. HI tells the agent that he is not being blamed for what he did, but he should be altered so as properly to be guided in the future to do what he ought to do. Is the HI proposal internally consistent in this regard? Mitchell Berman doubts it in the context of determinist concerns. Here is his formulation.

...determinism presents a challenge not only for any retributivist response to Jn(P)[the demand for the need to justify punishment] but—more fundamentally—even to Jn(P) itself. That is because, in my view, determinism threatens not only judgments of responsibility,

¹⁴⁹ (Morse 1999, 2002).

¹⁵⁰ Pereboom and Caruso

blameworthiness, or desert but all normative judgments on the ground that the forward-looking judgment “ought to” implies the cogency of the corresponding backward-looking judgment “ought to have.” If so, then to the extent that determinism threatens the latter, it threatens the former too.¹⁵¹

I believe that the same critique applies to HI. If the agent is not responsible for what he has done in the past, he is also certainly not responsible for what he will do in the future. What does it mean to say that he ought to do better? Whether he continues to behave badly or is manipulated into behaving well by reasons or by any other mechanism, it is not up to him and he deserves neither praise nor blame for his future conduct. We can say that it would be a good thing if he behaved well and a bad thing if he behaved badly, but whether he acts that way or not is not up to him. The “ought” in HI, like the concepts of agency, dignity and autonomy, is pallid. If in response it is claimed that the agent can be guided by reason, then we might ask whether he is guided because he directs himself to be or simply because he is. If the answer is the former, then this sounds like compatibilist agency and why is this not sufficient for desert? If it is the latter, the guidance is ephemeral and certainly not up to the person.

Before turning to the systemic and practical implications of HI, let us consider how our reactions to wrongdoing would be affected by the genuine, intellectually internalized adoption of HI. We know that we have the psychological capacity to behave consistently with the truth of desert and the justifiability of blame and punishment. That is the world we live in now. Over time, HI hopes that HI-appropriate emotions and responses will replace those fueled by unjustified acceptance of desert. I will assume that we could remake our psychology consistent with the prescription of HI, although I do think it is an open question whether this is possible. The question is whether one would want to remake ourselves psychologically in the way HI desires. Recall that I am assuming that HI makes no mistakes and that the underlying metaphysical dispute between it and a desert regime is unresolvable. The question, then, is which regime we lend our support.

Let us start with an everyday case: an act of interpersonal disloyalty and betrayal, such as the infidelity of one’s allegedly committed partner. Assume for the sake of simplicity that according to traditional, pre-HI standards that there is no mitigating or excusing condition present. According to HI, anger, resentment and indignation would be unwarranted and no blame would be deserved for the betrayer’s behavior. HI concedes that understandable but unwarranted emotional reactions such as anger will occur, but the betrayed partner cannot be blamed for them, even if they have fully embraced the truth of HI. This would be so even if they apparently tried but failed to suppress the unjustified emotions and their equally unjustified expression such as blaming the betrayer. If the betrayed partner morally addresses the betrayer and the latter seems to have seen the error of his ways, what should the betrayed partner do if the betrayer lapses again? Anger and blame are not warranted. Just more sorrow and regret? Is this possible? Even if it is, is it really a preferable way to live?

The next example is Bernard Madoff, whose extraordinary fraud wreaked havoc in the lives of countless innocent people and ruined many. The fraud was fully intentional and knowing. To the best of our knowledge, Madoff is not a psychopath and suffers from no major mental disorder and he had no other potential traditional mitigating or excusing condition. United States

¹⁵¹ Berman 2008: 271, n. 34

Federal District Court Judge Denny Chin imposed a sentence of 150 years on Madoff, whose crime he described as extraordinarily evil.¹⁵² As Judge Chin recognized, the ferocious length of the sentence was symbolic and unnecessary because Madoff's life expectancy at that time was 13 years and he presented no future danger to the community.¹⁵³ Nevertheless, Judge Chin emphasized that retribution, deterrence and justice for the victims demanded a harsh sentence.¹⁵⁴ Now, how would HI respond? As usual, sorrow and regret would be justified, but not anger and blame. Madoff was not responsible. But what should society do with Madoff? He fully understood that he had done grievous wrongs, he seemed genuinely regretful and he presents no further danger to society. On the public health quarantine model for social self-defense, nothing needs to be done. And general deterrence is barred in this model. If one tries to assimilate general deterrence to social self-defense, then all the familiar problems with consequentialism arise. Which world do you prefer to live in? Judge Chin's, that is, ours, or the world HI proposes.

Consider a more extreme, exceptional final example, Timothy McVeigh. McVeigh, a decorated and honorably discharged veteran who had seen combat service, was an extreme critic of the United States government who believed that it was systematically destroying our precious liberties. He was particularly enraged by the siege at Waco. After years of embracing and espousing such attitudes, he decided to act. On April 19, 1995, McVeigh methodically planned and executed the bombing of the Alfred P. Murrah Federal Building in Oklahoma City. The blast killed 168 people, including 19 children in a day care center in the building, and injured 684 others. Although on a few occasions, McVeigh said he regretted the loss of life, he also said that it was necessary. There was apparently no genuine moral regret; it was just the cost of doing business. And his last words about the attack were that he wished he had leveled the entire building, which would of course have caused far more loss of life. To the best of our knowledge, McVeigh was not a psychopath and had no major mental disorder.

HI would claim that the only appropriate emotions and responses to the carnage would be sorrow and regret and moral address with a view to altering McVeigh's reasons, but not anger, blame and punishment. Suppose McVeigh had a "conversion" not long after being apprehended. Even without any intervention, he came to understand the enormous monstrosity of his deeds. Of course, guilt, remorse and asking for forgiveness would be unwarranted because he was not responsible for his mass murder, but assume that his expressions of regret were genuine and we believed that he would not do it again. I assume that HI would demand that he be released, perhaps with some minimal supervision to ensure that his conversion was genuine and permanent. If no such conversion occurs, then he will presumably be confined in a most comfortable place that allows him to pursue his projects while he is subject to moral address and attempts to change his reasons. If he cannot, be changed or remains predictively dangerous for any reason, his comfortable confinement will continue. Even if such a set of reactions to McVeigh's outrage is possible, which I consider an open question, do you want to adopt such attitudes and all that they entail?

¹⁵² United States v. Madoff, (2009)

¹⁵³ *Id.*

¹⁵⁴ *Id.*

It is worth pausing to note that the shift in emotions and attitudes HI prescribes would be radical. Experimental work indicates that people are intrinsically retributivist.¹⁵⁵ This does not mean that retributive attitudes and practices either should be fostered or cannot be altered. Assuming it is possible to change ourselves as HI desires, however, we would be new people. The history of attempts to re-make human beings to achieve a radical new vision of society is one of failure and, indeed, horror. Furthermore, not only criminal justice would be affected. Deontological theories abound in many central legal doctrines, including torts, contracts and property.¹⁵⁶ Much of the legal system would have to be redesigned. HI cannot cherry pick criminal justice. Caution is warranted.

HI does not treat all cases of dangerous propensity as examples of a behavioral disorder (or any other kind of disorder), but neither the disordered dangerous agent nor the non-disordered dangerous agent is responsible and neither can be blamed. The only justification for differential treatment of those with and without disorder and generally for the treatment of any potentially dangerous agent would have to be purely consequential. Do the authorized decision-makers reasonably believe that the agent will be dangerous in the future and what needs to be and can be reasonably done to prevent the danger? Having already committed a dangerous action should simply be data to support the decision. Some people who have done great harm, such as Madoff, are unlikely to do it again, and some people who have not yet done a dangerous action might be highly predictively likely to do so without intervention. Any prediction will likely have false positives, especially if predicting low base-rate behavior, even if our predictive tools become ever more sensitive.

The questions for HI are what belief about the likelihood of an impending danger should be considered reasonable and whether harm should have to be imminent? Why not adopt a “screen and intervene” system that systematically screens all members of society and takes whatever interventions are necessary to prevent predicted danger? HI cannot of course borrow responsible agency from traditional self-defense doctrine as a limit on anticipatory intervention. Nor can it use responsible agency as a foundation for the right to liberty. It can use only consequential justifications to decide when anticipatory intervention is justifiable. Compare for example the HI criticisms of Saul Smilansky’s well-known, clever “funishment” argument, which says that those incarcerated for public safety reasons in a world without desert should be compensated for their undeserved deprivation of liberty.¹⁵⁷ The response is entirely consequential.¹⁵⁸ In effect, the use of self-defense to avoid the problems with consequentialism will not succeed because self-defense will inevitably collapse into consequentialism in the absence of responsible agency.

On consequential grounds, whether and to what degree anticipatory intervention is justifiable will be theoretically controversial. How many unnecessary anticipatory interventions would be justified to prevent one horrific crime? One cannot simply say that unacceptable false positive rates will limit anticipatory intervention because it begs the questions of how many false positives will be acceptable for what goals. In the involuntary civil commitment system, for example, we know the false positive rate is very high for low base-rate, serious harmful behavior

¹⁵⁵ See, e.g., Fehr & Gächter (2002),

¹⁵⁶ Moore, (2007).

¹⁵⁷ Smilansky, (2011)

¹⁵⁸ Pereboom and Caruso (2017)

because severely mentally disordered people do not commit serious crimes as a result of mental disorder alone.¹⁵⁹ The practice continues nonetheless because lack of responsible agency abridges traditional liberty interests. Notice, too, that as our prediction and prevention tools became more sensitive, the balance would shift towards ever more anticipatory intervention and thus more liberty intrusion. Even if there was social consensus on the precise interests to be weighed, doing that weighing is well-nigh impossible. HI offers no real limiting principle for the application of anticipatory intervention and thus for massive intrusions on liberty.

Finally, consider how the HI system will be administered. If all members of a society believe that no one is ever responsible for anything they do or will do in the future, what will be the effect on the balance of liberty versus safety? Will we value liberty as much? Will there be a shift towards a stronger preference for safety as the system proceeds? In the Anglo-American legal system there are constant attempts to “fill the gaps” in desert-disease jurisprudence either by responding harshly to responsible agents or by increasing the category of those considered not responsible,¹⁶⁰ Recidivist sentencing enhancement is an example of the former; involuntary civil commitment of so-called “mentally abnormal predators” is an example of the latter. All of these “gap-fillers” are justified consequentially on public safety grounds and virtually all have been condemned by informed commentators of every theoretical stripe as abusive of basic rights. Yet the political process produced them because the public demands them and the state agents that apply them are complicit. Desert-disease jurisprudence, which entirely depends on the distinction between responsible and non-responsible agency, at least sets imperfect limits on state overreaching. What similar constraints will be available in HI? It will be massive system if it is done properly, even if most interventions might be reasonably minimal. Will there be sufficient tools and experts? Unless those charged with administering the system have objective tools and no discretion, what will be done to prevent excesses and abuses, especially if the administrators do not think they are responsible for what they do? Even if there are sufficient experts and tools, there will be no limits on state intervention except a controversial, manipulable, and ultimately unknowable social welfare calculus. It will be a brave new world, indeed.

A few last points deserve mention. If our society were to increase equality and mental health services, a move compatible with both desert and HI regimes, there would still be a crime problem and differential rates of criminality among various demographic groups for many different reasons. The HI system would thus have disparate impact on some members of society who are already less advantaged, much as our current criminal justice system now does. Further, a desert-based regime has all the resources necessary to soften what seem to be disproportionately harsh and inutile criminal justice doctrines, procedures and practices and to adopt what HI considers more enlightened social welfare policies. I have recommended desert-based softening myself.¹⁶¹ Recall that virtually all the worst excesses in criminal justice were the product of consequential justifications and that desert had the resources to criticize them. We can have a fairer, kinder criminal justice system without abandoning desert.

Whether to adopt HI or to retain our regime of desert and responsibility is a political question. We are now ready to vote. Before you do, however, please recognize as we saw

¹⁵⁹ (Monahan et al 2001).

¹⁶⁰ (Morse 2011b, 2013a).

¹⁶¹ (e.g., Morse 2003)

previously that the new neuroscience plays no proper justificatory role in your decision because it raises no new issues in the debate between HI and desert. And be careful what you wish for.

Conclusion

We are in an era of culpability ferment. Although many mens rea reform efforts are meant to make the necessity for genuine culpability more robust by enhancing mens rea, but there are disquieting efforts internal to criminal law, especially in the context of mental disorder, and there are new, radical challenges both to mens rea and culpability generally arising from new scientific advances. But mental disorder should not be a carve-out from responsibility reform. And despite their admirable motivations and attempts to avoid the pitfalls of consequentialism, the radical challenges nonetheless offer a pallid conception of agency and a radical proposal for re-engineering social ordering that is more frightening than our current, imperfect regime. As C. L. Lewis recognized long ago, a system that treats people as responsible agents is ultimately more humane and respectful.¹⁶²

¹⁶² C. S. Lewis (1953)