



1-1-2014

## Pulling on the Thread of the Insanity Defense

R. George Wright

Follow this and additional works at: <https://digitalcommons.law.villanova.edu/vlr>



Part of the [Criminal Law Commons](#), and the [Law and Psychology Commons](#)

---

### Recommended Citation

R. G. Wright, *Pulling on the Thread of the Insanity Defense*, 59 Vill. L. Rev. 221 (2014).

Available at: <https://digitalcommons.law.villanova.edu/vlr/vol59/iss1/5>

This Article is brought to you for free and open access by Villanova University Charles Widger School of Law Digital Repository. It has been accepted for inclusion in Villanova Law Review by an authorized editor of Villanova University Charles Widger School of Law Digital Repository.

2014]

PULLING ON THE THREAD OF THE INSANITY DEFENSE

R. GEORGE WRIGHT\*

“The general delusion about free will [is] obvious. . . . One must view a [wicked] man, like a sickly one”<sup>1</sup>

I. INTRODUCTION

RECENTLY, the Supreme Court declined, over the dissent of three of its members,<sup>2</sup> to decide whether the Constitution requires an insanity defense to a criminal charge,<sup>3</sup> as distinct from a prosecutor’s proving any intent element of the offense itself.<sup>4</sup> Whether the Constitution requires an insanity defense in an otherwise appropriate case certainly seems worthy of definitive resolution. As for when to resolve this important issue, we should all sympathize both with those seeking immediate judicial guidance, and with those preferring broader discussion before a definitive judicial resolution.

With regard to the nature and scope of the insanity defense, however, matters are actually much more complicated than we might imagine. It is tempting to believe that if the Court carefully approaches the question of the possible constitutional requiredness of the insanity defense with an informed judicial temperament, all can be tidily resolved. Either some version of an insanity defense will be required under appropriate circumstances, or it will not. Closure on the insanity defense would then readily be reached.

This Article suggests, however, that at this point in our cultural history, no such tidy resolution is at all likely. Real closure on the insanity defense and on related broader questions cannot be expected. Given our current collective beliefs, as discussed in this Article, judicially pulling at the apparently loose thread of the insanity defense is likely to be at best futile. This is true regardless of whether we seek to remove the supposedly loose thread, or to tie that thread more securely into our system of criminal adjudication. Pulling on the supposedly loose thread of the insanity

---

\* Lawrence A. Jegen Professor of Law, Indiana University Robert H. McKinney School of Law.

1. CHARLES DARWIN, CHARLES DARWIN’S NOTEBOOKS, 1836–1844, at 608 (Paul H. Barrett et al. eds., 2009) (entry of September 6, 1838). For a contrasting perspective, see JOHN GRAY, STRAW DOGS: THOUGHTS ON HUMANS AND OTHER ANIMALS, at xii (2007) (“A truly naturalistic view of the world leaves no room for secular hope.”).

2. *Delling v. Idaho*, 133 S. Ct. 504 (2012) (denying certiorari in *State v. Delling*, 267 P.3d 709 (Idaho 2011)).

3. *See id.* at 504–05 (Breyer, J., dissenting). Justices Ginsburg and Sotomayor also dissented from the denial of certiorari. *See id.*

4. *See id.* at 505.

defense is instead likely to eventually result, perhaps inadvertently, in the unraveling of much of our closely-knitted jurisprudence of criminal responsibility.

The aim of this Article is to illustrate the unraveling of our current system of criminal responsibility and adjudication that is likely to result from any serious judicial attention to the insanity defense and its underlying assumptions. This Article, for the sake of simplicity, refers to the unraveling “thread” of the insanity defense, in the singular. But there are actually numerous “threads,” in the plural, involved in any unraveling of the insanity defense’s logic, and of the broader logic of criminal responsibility.

## II. *DELLING* V. IDAHO: A THREAD-PULLING OPPORTUNITY DECLINED

On the basis of a 6–3 vote, the Supreme Court in *Delling* declined to consider the judgment of the Idaho Supreme Court<sup>5</sup> that the Constitution does not require a separate insanity defense, at least as long as the prosecution must prove, beyond a reasonable doubt, any necessary intent or *mens rea* elements of the case.<sup>6</sup> The particularities of *Delling* itself are, in the end, of no great consequence. Tracing some of those particularities will, however, allow us to follow one possible track of the unraveling of the insanity defense, and of the unraveling of the standard legal logic of criminal responsibility in general.

Justice Breyer wrote for those dissenting from the denial of certiorari in *Delling*.<sup>7</sup> Justice Breyer begins his brief dissent by focusing on one particular form, among others, of an insanity defense.<sup>8</sup> He observes that

---

5. See generally *State v. Delling*, 267 P.3d 709 (Idaho 2011).

6. See *id.* at 712–13; see also *State v. Bethel*, 66 P.3d 840, 851 (Kan. 2003) (“Due process does not mandate that a State adopt a particular insanity test.”).

7. See *Delling*, 133 S. Ct. at 504.

8. Often, commentators recognize five distinguishable tests for insanity as a criminal defense. First, and most familiar, the *M’Naghten* test holds a defendant insane if, at the time of the offense, the defendant’s mental disease or deficiency resulted in a defect of reason such that the defendant either did not know the nature and quality of the act, or that the act was in a relevant sense wrong. See JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 25.04, at 350 (5th ed. 2009) (citing *M’Naghten’s Case*, (1843) 8 Eng. Rep. 718 (H.L.)); WAYNE R. LAFAVE, CRIMINAL LAW § 7.2, at 397–98 (5th ed. 2010). The second test adds to the *M’Naghten* test a further exculpatory possibility, bearing upon matters of control or volition, rather than of knowledge or understanding. Insanity under this “irresistible impulse” test may take the form of an involuntary inability to “choose” but to perform the prohibited act in question. See DRESSLER, *supra*, at 353; LAFAVE, *supra*, at 411–12. The third test, the Model Penal Code test, as developed by the American Law Institute, revises elements of the *M’Naghten* and so-called “irresistible impulse” tests described above. This test finds insanity if as a result of mental disease or defect, the defendant lacked substantial capacity to appreciate the criminality or, as a drafting alternative, the moral wrongfulness, of an act, or else lacked the substantial capacity to conform to the relevant legal requirements. See DRESSLER, *supra*, at 354 (citing MODEL PENAL CODE § 4.01(1)); LAFAVE, *supra*, at 420–21. The fourth test, formulated by the distinguished Judge David Bazelon in the case of

“[t]he law has long recognized that criminal punishment is not appropriate for those who, by reason of insanity, cannot tell right from wrong.”<sup>9</sup> This particular formulation, whether intended as an insanity defense formulation or not, certainly does not itself tell us much about persons who could be thought of as insane, but who nevertheless can tell right from wrong at the relevant time.<sup>10</sup> Nor does it convey much about persons who cannot tell right from wrong at the relevant time, but whose inability to do so did not causally stem from any form of insanity.<sup>11</sup> Nor, finally, does this brief formulation begin to clarify whether the terms “right” and “wrong” are used here in an objective or a subjective sense; or in a legal, moral, or conventional sense; or some combination thereof.<sup>12</sup> Nor, understandably, does this brief formula hint at whether the moral wrongness of an act is something we can really know or tell.

In any event, Justice Breyer then notes that under Idaho law, “prosecutors are ‘still required to prove beyond a reasonable doubt that a defendant had the mental capacity to form the necessary intent.’”<sup>13</sup> There may be differences, however, between intent elements and whatever we take insanity to involve. How reassuring this intent element requirement by itself should be is unclear, as Justice Breyer then goes on to discuss.<sup>14</sup>

But even before attending to Justice Breyer’s concern over proving *mens rea* elements as a substitute for an insanity defense, there are a number of inescapable, more basic questions. For one, what must be proved is

---

*Durham v. United States* asks whether the criminal defendant’s act was “the product of mental disease or defect.” See *Durham v. United States*, 214 F.2d 862 (D.C. Cir. 1954); see also DRESSLER, *supra*, at 355; LAFAVE, *supra*, at 414–15. The D.C. Circuit itself abandoned this test in *United States v. Brawner*. See *United States v. Brawner*, 471 F.2d 969 (D.C. Cir. 1972). Fifth and finally, Congress adopted in 1984 a general federal statutory test for insanity as a criminal defense. Under this *M’Naghten*-influenced test, the defendant bears the burden of proving, by clear and convincing evidence, that because of a severe mental disease or defect, the defendant was unable to appreciate either the nature and quality or the wrongfulness of the conduct. See 18 U.S.C. § 17(a) (2012); see also DRESSLER, *supra*, at 356 (citing 18 U.S.C. § 17(a) (2012)); LAFAVE, *supra*, at 414–15, 415 n.1.

9. *Delling*, 133 S. Ct. at 504 (Breyer, J., dissenting) (citing traditional authority).

10. Compare the alternatives listed *supra* note 8, with the critical commentaries cited therein.

11. Whether the accused was coincidentally insane at the time or not.

12. For a sense of some of the possibilities, see, for example, *United States v. Ewing*, 494 F.3d 607, 616 (7th Cir. 2007) (discussing legal wrongfulness, public moral wrongfulness, and personal or subjective moral wrongfulness); *United States v. Polizzi*, 545 F. Supp. 2d 270, 274 (E.D.N.Y. 2008); *State v. Singleton*, 48 A.3d 285, 295–96 (N.J. 2012) (surveying state authorities); *In re Standard Jury Instructions in Criminal Cases* (No. 2005-5), 939 So. 2d 1052, 1054 (Fla. 2006) (per curiam) (discussing legal or societal moral standards); Walter Sinnott-Armstrong & Ken Levy, *Insanity Defenses*, in *THE OXFORD HANDBOOK OF PHILOSOPHY OF CRIMINAL LAW* 299, 303–04 (John Deigh & David Dolinko eds., 2011).

13. *Delling*, 133 S. Ct. at 505 (Breyer, J., dissenting) (quoting *State v. Delling*, 267 P.3d 709, 712 (Idaho 2011) (further quotation omitted)).

14. See *id.*

actually the “mental capacity” to form a particular intent, presumably at the relevant time.<sup>15</sup> It would hardly do, one might imagine, to prove merely that the defendant had that mental capacity at other times, or only in some “background” sense. Yet in some cases, this more general, or background, mental capacity might be all that a prosecutor could reasonably be expected to show. Mental capacity to form any necessary intent at the precise time of the offense might unavoidably be inferred by a jury from evidence of the defendant’s background capacities. But inferring a defendant’s capacities at the relevant time from the defendant’s capacities at other times may often be questionable. A defendant’s general capacity, by analogy, to recite the alphabet backwards will often provide only modest evidence, if any evidence at all, of the defendant’s capacity to perform this feat at the time of the offense.

The idea of a “mental” capacity to form a particular intent is similarly unclear.<sup>16</sup> Are there capacities that are distinctively mental? Is a mental capacity the only kind of capacity necessary for the formation of the relevant intent? Could there also be something like a physical, or a physiological, capacity to form an intent? Is the mental capacity, and not the physical or the physiological capacity, alone of any legal relevance? Why, if both forms of capacities, mental and physical, might be necessary to form the intent? Perhaps the specific reference to a mental capacity to form an intent is thus misleading.

More fundamentally yet, we might wonder why the law insists on focusing on whether the defendant had, at whatever time, the mere capacity—something like an underlying ability or power—to form a particular intent.<sup>17</sup> It may sound as though having a capacity, at the relevant time, should always be crucial. But isn’t it possible to have the mere capacity to appreciate the wrongness of an act, at the time of the act, and yet to be somehow blocked or prevented from *actually* appreciating the act’s wrongness, as by some outside interfering factor?<sup>18</sup> What if an outsider, say, blocks affirmative exercise of our underlying capacity?

All of these preliminary problems may ultimately have satisfactory solutions. At the moment, though, jurisprudence and philosophy have no such consensually satisfactory set of solutions. And even more importantly, current trends in science and philosophy make any such broadly attractive resolution less likely. The jurisprudence of criminal insanity is, even on these preliminary matters, unraveling.

For the present, though, and much more specifically, Justice Breyer in the *Delling* case points to an example of one apparently loose thread in the much broader unraveling skein. Justice Breyer notes that “the difference

---

15. *See id.*

16. *See id.*

17. *See id.*

18. We might say that someone had the capacity to do something, or to appreciate something, but was momentarily distracted, and thus in a real, causal sense, unable to do that action at the relevant time.

between the traditional insanity defense and Idaho's standard is that the latter permits the conviction of an individual who knew *what* he was doing, but had no capacity to understand that it was wrong."<sup>19</sup> Justice Breyer then illustrates this apparently abstract distinction with a vivid hypothetical example, contrasting Idaho's dissimilar treatment of two arguably relevantly similar cases.

Justice Breyer thus imagines two cases in which the criminal defendant seems crucially motivated by a clearly insane delusion. In the first case, the defendant believes that his human victim is in fact a wolf.<sup>20</sup> In the second case, the defendant believes that he has been ordered to kill a particular human by a supernaturally empowered wolf.<sup>21</sup> The idea is that under Idaho law, the defendant in the first case, but not the defendant in the second case, could be appropriately accommodated.<sup>22</sup> The defendant in the first case did not know that he was killing a human being. Under Idaho law, this negates an element of the offense itself.<sup>23</sup> The defendant in the second case, however, intended to kill a human and knew that he did so, however insanely motivated or grossly misapprehending of the circumstances, thus fulfilling the relevant *mens rea* element of the offense.<sup>24</sup>

As Justice Breyer sensibly suggests, it is unclear why we should want to exculpate the defendant in the first case, but not in the second.<sup>25</sup> There seem to be only inessential differences in the roles played by the two sorts of insane belief in the two cases.<sup>26</sup> In light of the narrowed versions, or abolition, of the insanity defense,<sup>27</sup> and the arguable tension between

19. *Delling*, 133 S. Ct. at 505 (Breyer, J., dissenting).

20. *See id.*

21. Note one complication for broader, mainstream conceptions of an insanity defense: a defendant in the second case might know, or somehow have the capacity to know, that law or convention typically considers killing human beings to be wrong. But a defendant who believes that he has been ordered to kill not simply by a wolf, but by a wolf with the right sort of supernatural qualities, might well also sincerely believe that in such case, the killing is morally right even if legally or conventionally wrong. This may again presume, though, that the moral wrongness of an act is something that sane persons can typically "tell" or "know." And this is increasingly controversial. Metaethics itself is currently in a rather frayed condition. Many and various challengers to moral realism have arisen. *See, e.g.*, Geoff Sayre-McCord, *Metaethics*, STAN. ENCYCLOPEDIA PHIL. (Edward N. Zalta ed., 2012), <http://plato.stanford.edu/entries/metaethics> (last visited Mar. 11, 2013) (referring to proliferating and increasingly exotic options). For merely one interesting recent contribution, see Sharon Street, *A Darwinian Dilemma for Realist Theories of Value*, 127 PHIL. STUD. 109 (2006).

22. *See Delling*, 133 S. Ct. at 505 (Breyer, J., dissenting).

23. *See id.*

24. *See id.*

25. *See id.*

26. *See id.*

27. *See id.* (citing *State v. Bethel*, 66 P.3d 840, 843 (Kan. 2003)). The *Bethel* court concluded that "the affirmative insanity defense is a creature of the 19th century and is not ingrained in our legal system as to constitute a fundamental principle of law." *Bethel*, 66 P.3d at 851.

such provisions and the due process clause,<sup>28</sup> Justice Breyer would have granted certiorari in *Delling*.<sup>29</sup> Had the Court done so, however, it would have been tempted by a variety of apparently loose threads, a number of which are inseparable from current understandings of the logic of the insanity defense and of criminal responsibility more broadly.

### III. THE UNDERLYING LOGIC OF THE INSANITY DEFENSE AND THE PROCESS OF DOCTRINAL UNRAVELING

#### A. *An Initial Sense of the Unraveling of the Logic of Sanity and Insanity*

However controversial the idea of freedom of the will may be,<sup>30</sup> the law of criminal responsibility commonly refers to, and is commonly thought to depend upon, such an idea. Consider, merely for example, the claim by Justice Jackson in the well-known case *Morissette v. United States*<sup>31</sup> that:

The contention that an injury can amount to a crime only when inflicted by intention is no provincial or transient notion. It is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil.<sup>32</sup>

Justice Jackson relies on standard English and American authorities in support of this no doubt imprecise understanding of the nature and role of freedom of the will.<sup>33</sup> In particular, Justice Jackson cites Dean Roscoe Pound for the claim that “[h]istorically, our substantive criminal law is

28. *See Delling*, 133 S. Ct. at 505 (Breyer, J., dissenting).

29. *See id.* For other cases beyond *Delling* and *Bethel* that raise generally similar constitutional issues regarding the availability of an insanity defense, see, for example, *State v. Herrera*, 895 P.2d 359, 366 (Utah 1995) (“[T]he common law and our basic principles of ordered liberty are not offended by the mens rea model.”); *State v. Searcy*, 798 P.2d 914 (Idaho 1990) (stating that common law does not conflict with mens rea model); *State v. Korell*, 690 P.2d 992 (Mont. 1984) (same). *But see* *Finger v. State*, 575, 27 P.3d 66, 84 (Nev. 2001) (en banc) (“We conclude that legal insanity is a well-established and fundamental principle of the law of the United States. . . . [P]rotected by the Due Process Clauses of both the United States and Nevada Constitutions.”). For one permissible revision of more robust insanity defenses, see *Clark v. Arizona*, 548 U.S. 735, 756 (2006) (“We are satisfied that neither in theory nor in practice did Arizona’s 1993 abridgment of the insanity formulation deprive Clark of due process.”). The *Clark* court was referring, in particular, to the narrowed insanity test adopted in *State v. Mott*. *See* *State v. Mott*, 931 P.2d 1046 (Ariz. 1997); *see also Clark*, 548 U.S. at 753 (“There being such fodder for reasonable debate about what the cognate legal and medical tests should be, due process imposes no single canonical formulation of legal insanity.”).

30. *See, e.g.,* JOHN MARTIN FISCHER ET AL., *FOUR VIEWS ON FREE WILL* (2007) (featuring contributions by and debates among Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas).

31. 342 U.S. 246 (1952).

32. *Id.* at 250.

33. *See id.* at 250 n.4.

based upon a theory of punishing the vicious will. It postulates a free agent confronted with a choice between doing right and doing wrong and choosing freely to do wrong.”<sup>34</sup> A number of courts have since referenced Justice Jackson’s language in this regard.<sup>35</sup> Scholars have often written of free will in similar terms.<sup>36</sup>

Such mere references to free will do not establish the nature or meaning of free will in any legal context. Neither do they establish the extent, if any, to which criminal jurisprudence actually does or should rely on some conception of free will. We must look elsewhere if we are to develop any more nuanced sense of the contested idea of free will and its actual or appropriate role in establishing criminal responsibility.

If we look to the best of what is being thought and said by contemporary philosophical specialists in this area, we find no consensus and no analytical short cuts. We might imagine that we need only identify basic opposing “camps” of thinkers on free will, or the few broad “schools of thought” on such matters. Unfortunately, references to schools of thought on free will and responsibility inevitably involve serious oversimplifications. Schools of thought on free will and criminal responsibility have multiplied and proliferated at a remarkable rate. It would not be much of an exaggeration to say that at this point, nearly every serious academic philosopher in the field is developing, at a certain level of detail, his or her own slightly distinctive approach to free will and criminal responsibility, or the lack thereof. At this point, nearly every imaginable variation

---

34. *Id.* (quoting ROSCOE POUND, INTRODUCTION TO SAYRE, CASES ON CRIMINAL LAW (1927)).

35. *See, e.g.*, *Tison v. Arizona*, 481 U.S. 137, 159, 171 (1987) (Brennan, J., dissenting); *United States v. Burwell*, 690 F.3d 500, 527, 532 (D.C. Cir. 2012) (Kavanaugh, J., dissenting); *United States v. Manatau*, 647 F.3d 1048, 1051 (10th Cir. 2011); *United States v. Apollo Energies, Inc.*, 611 F.3d 679, 687 (10th Cir. 2010) (citing *Staples v. United States*, 511 U.S. 600 (1994)) (discussing logic and limited scope of “strict liability” crimes); *United States v. Parks*, 583 F.3d 923, 925 (6th Cir. 2009).

36. *See, e.g.*, 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 27 (Univ. of Chi. 1979) (1769); DRESSLER, *supra* note 8, at 345 (quoting HERBERT L. PACKER, THE LIMITS OF THE CRIMINAL SANCTION 132 (1968)) (“[W]e . . . put up with the bother of the insanity defense because to exclude it is to deprive the criminal law of its chief paradigm of free will.”); Julie E. Grachek, Note, *The Insanity Defense in the Twenty-First Century: How Recent United States Supreme Court Case Law Can Improve the System*, 81 IND. L.J. 1479, 1481 (2006) (quoting Peter Arenella, *Reflections on Current Proposals to Abolish or Reform the Insanity Defense*, 8 AM. J.L. & MED. 271, 273 (1982) (assuming “a criminal justice system based on free will”); John F. Stinneford, *Punishment Without Culpability*, 102 J. CRIM. L. & CRIMINOLOGY 653, 659 (2012) (quoting Blackstone). For a philosophical defense of the crucial importance of a robust free will for moral responsibility and other valued qualities, see ROBERT KANE, THE SIGNIFICANCE OF FREE WILL 80 (1996). For a neuroscience-based skeptical account of free will that nevertheless recognizes the criminal law of insanity as presuming the possibility of free will, see Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Everything and Nothing*, 359 PHIL. TRANS. R. SOC. LOND. 1775, 1776 (2004), available at <http://www.wjh.harvard.edu/~jgreene/GreeneWJH/GreeneCohenPhilTrans-04.pdf>.



on the basic themes of free will, responsibility, and their meaning and roles has been articulated by at least one philosopher. If an approach can be imagined, it has likely been advocated.

Of course, it would be tedious and impractical to try to prove this latter claim, so we leave it to be casually inferred, if anyone is so inclined, from the numerous academic sources referred to throughout. But before we can glance at even a fragment of the contemporary complications, we must start out with some simplified, mainstream set of distinctions, however contestable any such framework may be. Merely for the sake of convenience, then, let us adopt for the moment a few basic distinctions as articulated by the well-respected contemporary philosopher Alfred Mele.<sup>37</sup>

B. *Some Basic Distinctions Among Approaches to Free Will and Responsibility*

To begin with, then, some people believe—rightly or wrongly—that whether we can ever choose, will, or act with the relevant sort of freedom depends on whether the world operates “deterministically.” Determinism is the controversial view that a complete description of the universe, at any point in time, along with a complete account of the laws of nature, would allow the prediction, no matter how far into the future, of every human decision, action, and choice.<sup>38</sup>

It might initially seem clear that determinism would rule out the relevant sort of freedom. But many philosophers think that determinism would be compatible with what we could fairly and meaningfully call free will and freedom of action. Conveniently, the belief that determinism is somehow compatible with some relevant and meaningful sense of free will is known as “compatibilism.”<sup>39</sup> And in parallel, the belief that determinism is incompatible with the exercise of such free will is known as “incompatibilism.”<sup>40</sup>

Incompatibilism faces in two opposing directions, and divides into two large opposing camps. This should not be surprising, because incompatibilities in general between any two options can be resolved in more than one way. Some incompatibilists deny that the world is deterministic and, complications aside, also embrace the possibility of at least the occasional exercise of a meaningful freedom of will. Somehow, the world “leaves room” for free will. Such incompatibilists are known, in a non-political sense, as “libertarians.”<sup>41</sup> And some incompatibilists, in contrast, accept that the world is deterministic, and at least partly on that basis,

---

37. See ALFRED R. MELE, *EFFECTIVE INTENTIONS: THE POWER OF CONSCIOUS WILL* 150 (2009).

38. *See id.*

39. *See id.*

40. *See id.* Again, we are setting aside all sorts of refinements and complications.

41. *See id.*

deny the possibility of any genuine freedom of the will.<sup>42</sup> One might think here of a purely mechanical universe. These distinctions at least give us a starting point on some of the relevant terminology.

C. *The Basic Distinctions Lose Their Hold and Begin  
the Jurisprudential Unraveling*

This exercise in specifying some basic terminology will at least allow the argument to move forward. The basic terms, however, today serve as much as a launching point for proliferating complications as for steering or directing a readily summarizable debate. Today, in a phrase, nearly everything is up for grabs. There is very little genuinely common ground.

Merely for example, it has been argued with great sophistication both that “free will . . . is not foundational for criminal responsibility”<sup>43</sup> and that we can have some version of free will while rejecting the idea of moral responsibility in criminal contexts.<sup>44</sup> Already, the complications are thus underway. But we need not think of free will as a binary, entirely present or entirely absent, capacity that one either possesses, at the moment, or that one does not. Perhaps free will comes in degrees, depending upon various circumstances.<sup>45</sup> Or perhaps we should reject the apparently common sense idea of symmetry between the free will requirements for praising and for blaming.<sup>46</sup> Perhaps, on such a theory, we can thus be confident enough in our belief in free will and responsibility for purposes of genuinely morally praising someone, while lacking sufficient confidence to justify blaming, or imposing retributive punishment on, criminal defendants.<sup>47</sup>

Or, to take a somewhat different tack, perhaps praiseworthiness does not require the ability to have done otherwise than one did, whereas blameworthiness, or punishability in that sense, does, in contrast, require

---

42. See *id.* at 150–51. For slightly more complex taxonomies of some basic distinctions, see, e.g., Stefaan E. Cuypers, *Moral Shallowness, Metaphysical Megalomania, and Compatibilist Fatalism*, 16 *ETHICAL THEORY & MORAL PRAC.* 173 (2013), available at <http://link.springer.com/content/pdf/10.1007%2Fs10677-011-9318-3.pdf>; Peter van Inwagen, *How to Think About the Problem of Free Will*, 12 *J. ETHICS* 327, 330 (2008) (defining various relevant terms).

43. See Stephen J. Morse, *Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience*, 9 *MINN. J.L. SCI. & TECH.* 1, 3 (2008).

44. See BRUCE N. WALLER, *AGAINST MORAL RESPONSIBILITY* 20, 44–45, 257–61 (2011).

45. See 1 ALAN RYAN, *ON POLITICS: A HISTORY OF POLITICAL THOUGHT: FROM HERODOTUS TO THE PRESENT* 176 (2012).

46. See, e.g., Benjamin Vilhauer, *Persons, Punishment, and Free Will Skepticism*, 162 *PHIL. STUD.* 143 (2013).

47. See *id.* For an attempt to separate the capacity for (present) responsible agency from (absent) legal blameworthiness, see Nicola Lacey & Hanna Pickard, *From the Consulting Room to the Court Room? Taking the Clinical Model of Responsibility Without Blame into the Legal Realm*, 33 *OXFORD J. LEGAL STUD.* 1 (2013), available at <http://ojls.oxfordjournals.org/content/33/1/1.full.pdf+html>.

that the defendant have had the capacity to have acted otherwise.<sup>48</sup> Or perhaps, on an entirely different path, responsibility should be considered not as a matter of free will, but as a matter of one's rational powers, and as a matter of degree of control.<sup>49</sup> Or perhaps the purportedly basic distinction between the rich varieties of compatibilist<sup>50</sup> and incompatibilist<sup>51</sup> approaches to free will may turn out not to be basic after all.<sup>52</sup>

In any event, these few examples of a much wider range of emerging complications, or fraying threads, illustrate how little philosophical consensus on any issue underlies the jurisprudence of the insanity defense in particular and criminal responsibility in general.

#### D. *Some Contemporary Doubts About Traditional Free Will: Another Perspective*

Another way to illustrate the unraveling of our basic understandings of free will, insanity, and criminal responsibility could start with the traditional legal theory, and then note the rapidly developing fraying. It has been observed, for example, that "[t]he law has long recognized that those who do not qualify as moral agents should not be punished."<sup>53</sup> To not qualify as a genuine moral agent is to be ineligible for deserved punishment. Traditionally, it has been thought that only a relative few criminal defendants could not justly be blamed for their actions,<sup>54</sup> and thus did not deserve to be punished.<sup>55</sup> The rest of us, most of the time, were thought of as moral agents.

This traditional understanding has, of late, been unraveling. Increasingly, thoughtful observers have come to distrust traditional ideas of genu-

---

48. See, e.g., DANA KAY NELKIN, MAKING SENSE OF FREEDOM AND RESPONSIBILITY 3 (2011); SUSAN WOLF, FREEDOM WITHIN REASON 79 (1993); Jules Holroyd, *Review of: Making Sense of Freedom and Responsibility*, 73 ANALYSIS REVS. 198–202 (2013) (reviewing DANA KAY NELKIN, MAKING SENSE OF FREEDOM AND RESPONSIBILITY (2011)), available at <http://analysis.oxfordjournals.org/content/73/1/198.full.pdf+html>.

49. See R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 7–8 (1994). But see Randolph Clarke, *Free Will and the Conditions of Moral Responsibility*, 66 PHIL. STUD. 53, 55 (“[T]he conviction that free will is a necessary condition of being morally responsible for what one does.”).

50. See *supra* note 39 and accompanying text.

51. See *supra* note 40 and accompanying text.

52. See, e.g., Bruce N. Waller, *A Metacompatibilist Account of Free Will: Making Compatibilists and Incompatibilists More Compatible*, 112 PHIL. STUD. 209, 211 (2002).

53. Melinda Carrido, Comment, *Revisiting the Insanity Defense: A Case for Resurrecting the Volitional Prong of the Insanity Defense in Light of Neuroscientific Advances*, 41 SW. L. REV. 309, 310 (2012). The Supreme Court has concluded that legal, if not medical accounts of insanity, must take issues of individual responsibility into consideration. See *Kansas v. Hendricks*, 521 U.S. 346, 359 (1997); Ralph Slovenko, *The DSM in Litigation and Legislation*, 39 J. AM. ACAD. PSYCHIATRY L. 6, 7 (2011) (quoting *Hendricks*).

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

55. See *id.*

ine agency and of free will and responsibility,<sup>56</sup> with the idea of free will in particular increasingly being thought of as a persistent illusion.<sup>57</sup> The idea of freedom of the will is seen increasingly not as drawn from scientific investigation, but as derived instead from religion,<sup>58</sup> or as otherwise suspect. Perhaps the illusion of free will is merely naturally “selected” for its evolutionary survival value. While neuroscientists have recently been conspicuous in this regard,<sup>59</sup> substantial numbers of scientists in general<sup>60</sup> and of philosophers<sup>61</sup> have expressed their skepticism regarding a robust free will, or of genuine agency.

Along these lines, merely for example, the distinguished scientist Francis Crick has concluded that “‘You,’ your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. . . . ‘You’re nothing but a pack of neurons.’”<sup>62</sup> Are there, in effect, sane machines and insane machines, in the sense traditionally intended by the criminal law?

---

56. For introductory discussion, see, e.g., David Hodgson, *Criminal Responsibility, Free Will, and Neuroscience, in* DOWNWARD CAUSATION AND THE NEUROBIOLOGY OF FREE WILL 227–41 (Nancey Murphy, George F.R. Ellis & Timothy O’Connor eds., 2009) (Hodgson was both first-rate philosopher and Justice of Supreme Court of New South Wales); Tamler Sommers, *The Objective Attitude*, 57 PHIL. Q. 321 (2007). For a relatively optimistic view of likely future developments, see Stephen J. Morse, *Neuroscience and the Future of Personhood and Responsibility, in* 8 CONSTITUTION 3.0: FREEDOM AND TECHNOLOGICAL CHANGE 114 (Jeffrey Rosen & Benjamin Wittes eds., 2011).

57. See Hodgson, *supra* note 56, at 232.

58. See GRAY, *supra* note 1, at xii.

59. For a sophisticated overview, see generally Michael S. Moore, *Responsible Choices, Desert-Based Legal Institutions, and the Challenges of Contemporary Neuroscience*, 29 SOC. PHIL. & POL’Y 233 (2012). Among the neuroscientists, see Kelly Burns & Antoine Bechara, *Decision Making and Free Will: A Neuroscience Perspective*, 25 BEHAV. SCI. & L. 263, 283 (2007) (noting neuroscience developments as undermining legal logic of freedom of will, but with potential result of more humane and effective criminal justice system); Nigel Eastman & Colin Campbell, *Neuroscience and Legal Determination of Criminal Responsibility*, 7 NATURE REV. NEUROSCIENCE 311, 317 (2006), available at <http://lib.semi.ac.cn:8080/tsh/dzzy/wsqq/selected%20papers/Nature%20Reviews%20Neuroscience/7-311.pdf>; see also Greene & Cohen, *supra* note 36, at 1783–84; Dean Mobbs et al., *Law, Responsibility, and the Brain*, 5 PLOS BIOLOGY 693, 693 (2007) (“[M]any contemporary neuroscientists assume . . . . [T]hat violence and antisocial behaviour emanate from a mechanistically determined brain.”).

60. See *infra* notes 62–64.

61. For a particularly straightforward example, see RICHARD DOUBLE, *THE NON-REALITY OF FREE WILL* (1991) (noting claims to free will as not so much objectively false as incoherent or meaningless).

62. FRANCIS CRICK, *THE ASTONISHING HYPOTHESIS: THE SCIENTIFIC SEARCH FOR THE SOUL* 3 (1995); see also DANIEL C. DENNETT, *FREEDOM EVOLVES* 2–3 (2003) (“We are each *made of* mindless robots and nothing else, no non-physical, non-robotic ingredients at all. The differences among people are all due to the way their particular robotic teams are put together, over a lifetime of growth and experience.”); Anthony R. Cashmore, *The Lucretian Swerve: The Biological Basis of Human Behavior*

Also among the scientists and philosophers, Daniel Wegner, for example, has elaborated on the status of free will along these lines:

[I]t seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. Although it is sobering and ultimately accurate to call all this an illusion, it is a mistake to conclude that the illusory is trivial. On the contrary, the illusions piled atop apparent mental causation are the building blocks of human psychology and social life.<sup>63</sup>

Again, if we do not have selves, minds, or agency, in what remaining sense can some of us be sane and others insane?

The widely respected philosopher Saul Smilansky has also reached a remarkably sophisticated but equally untidy conclusion in this regard. Professor Smilansky argues that:

[O]ur priority should be to live with the assumption of libertarian free will although there is no basis for this other than our very need to live with this assumption; but as we cannot accept this way of seeing things, and confront dangers to our beliefs, illusion must play a central role in our lives.<sup>64</sup>

Each of these emerging approaches to freedom of the will and to criminal responsibility, and their implications for an insanity defense in particular, is itself controversial. Jointly, however, they serve well enough as illustrations of the current unraveling of traditional ideas of free will, moral responsibility, sanity and insanity, and criminal responsibility. Unavoidably, at some point, the law of insanity must somehow acknowledge, favorably or unfavorably, the increasing influence of materialism,<sup>65</sup> mech-

---

*and the Criminal Justice System*, 107 PROC. NAT'L ACAD. SCI. USA 4449, 4502, 4504 (2010), available at [www.pnas.org/cgi/doi/10.1073/pnas.0915161107](http://www.pnas.org/cgi/doi/10.1073/pnas.0915161107).

63. DANIEL M. WEGNER, *THE ILLUSION OF CONSCIOUS WILL* 341–42 (2002); see also ALEX ROSENBERG, *THE ATHEIST'S GUIDE TO REALITY: ENJOYING LIFE WITHOUT ILLUSIONS* 19 (2011) (“The love of stories comes to us in a package that . . . includes the illusion of free will. . .”).

64. Saul Smilansky, *Free Will: From Nature to Illusion*, 101 PROC. ARISTOTELIAN SOC'Y 71, 93 (2001). At greater length, see SAUL SMILANSKY, *FREE WILL AND ILLUSION* (2000) [hereinafter *FREE WILL AND ILLUSION*]. For a straightforward account of one possible grounding of Smilansky's skepticism about free will, see RICHARD RORTY, *PHILOSOPHY AND THE MIRROR OF NATURE* 387 (1979) (“Every . . . thought . . . will turn out to be completely predictable in purely naturalistic terms. Some atoms-and-the-void account of micro-processes within individual human beings will permit the prediction of every sound or inscription which will ever be uttered.”).

65. See, e.g., *CONTEMPORARY MATERIALISM: A READER* (Paul K. Moser & J.D. Trout eds., 1995); PETER ATKINS, *ON BEING: A SCIENTIST'S EXPLORATION OF THE GREAT QUESTIONS OF EXISTENCE*, at xii n.1 (2011) (“[T]he whole of all there is can be accounted for in terms of matter and its interactions.”); Andrew M. Bailey, *Review of: The Waning of Materialism*, 120 MIND 534, 534 (2011) (“Materialism is widely taken as orthodoxy in the metaphysics and philosophy of mind.”) (reviewing, to

anism,<sup>66</sup> physicalism,<sup>67</sup> and some strong forms of naturalism,<sup>68</sup> as illustrated above.<sup>69</sup>

Judicially, threads will thus inevitably be pulled, and the largely unanticipated doctrinal unraveling will further proceed, for good or ill. Let us now further explore some complications raised, in their particular turn, by controversies over determinism and randomness, then by compatibilist approaches, and finally by the philosophical doctrines known as illusionism and fictionalism.

1. *Determinism, Indeterminism, Randomness, and Further Doctrinal Unraveling*

The strongest formulations of causal determinism<sup>70</sup> pose serious problems for much of criminal and sentencing law, and for the law of insanity in particular. One leading contemporary philosopher, referring to strong formulations of determinism, suggests that:

The basic idea behind . . . determinist positions of this kind is that determinism undermines moral distinctions between the guilty and non-guilty, and so we should only be concerned with the danger to society posed by certain people. If a person commits a crime, the central issue is not his bad intention and similar issues *with which the legal system is presently concerned* but rather social-medical questions about his future behaviour.<sup>71</sup>

If this view is right, trying to distinguish between legal sanity and insanity would be beside the point. Considerations of danger, however defined and measured, would displace our standard concerns with sanity and insanity.

---

contrary, THE WANING OF MATERIALISM (Robert C. Koons & George Bealer eds., 2010)).

66. See, e.g., Heath White, *Mattering and Mechanism: Must a Mechanistic Universe Be Depressing?*, 24 RATIO 326 (2011).

67. See, e.g., David Papineau, *Physicalism and the Human Sciences*, in PHILOSOPHY OF THE SOCIAL SCIENCES: PHILOSOPHICAL THEORY AND SCIENTIFIC PRACTICE (C. Mantzavinos ed., 2009) (“We are all physicalists now. It was not always so. . . . This is a profound intellectual shift.”).

68. See, e.g., DAVID PAPINEAU, PHILOSOPHICAL NATURALISM (1993); STEWART GOETZ & CHARLES TALIAFERRO, NATURALISM 16 (2008) (“Strict naturalism is incompatible with libertarian freedom. . . .”); James Lenman, *Naturalism Without Tears*, 22 RATIO 1 (2009). For a relatively brief but broad-ranging and sophisticated overview, see Brian Leiter, *Naturalism in Legal Philosophy*, STAN. ENCYCLOPEDIA PHIL. (July 31, 2012), <http://plato.stanford.edu/entries/lawphil-naturalism>.

69. See *supra* notes 56–64 and accompanying text.

70. See *supra* note 38 and accompanying text.

71. FREE WILL AND ILLUSION, *supra* note 64, at 85 (emphasis added); cf. Michael S. Moore, *Causation Revisited*, 42 RUTGERS L.J. 451, 489 (2011) (“Only the pure utilitarian [about punishment] believes that moral blameworthiness is irrelevant to punishability.”).

The less obvious point, however, is that concerns for the social dangers posed by a person—either before or after the commission of any crime—need not be reducible to benign, humane, and progressive forms of medicine,<sup>72</sup> rehabilitation, resource equalization, denied opportunities, or to structural economic and social reform. Even if we think in terms of medicine, a narrow concern for danger and a broad concern for maximizing overall utility<sup>73</sup> clearly need not lead to a progressive response to those we would otherwise have characterized as insane. Dangers may, frankly, be most cheaply reduced by non-progressive measures. Cost-effectiveness and overall utility maximization may well lead a legislature or a court either to humane, or to less than humane, screening based on genetic analyses, “quarantining,”<sup>74</sup> or even to non-consensual psycho-surgery.<sup>75</sup>

Thus abandoning our traditional concern for legal sanity or insanity in favor of a concern for overall danger, or other utilitarian considerations, could well lead our jurisprudence in unexpectedly unprogressive directions. Would it be possible, though, to minimize this risk by taking a more subtle approach to the idea of causal determinism?

It has been argued, in particular, that uncertainties about determinism, and about whether anyone is morally responsible, should affect our policy thinking.<sup>76</sup> This approach does not argue that if we are uncertain about determinism and moral responsibility, we should find everyone not responsible.<sup>77</sup> Rather, the conclusion is one of agnosticism.<sup>78</sup> Thus in the absence of additional evidence, the moral and legal systems should on this theory seek to avoid, suspend, or withhold judgments as to responsibility, either way. Whatever the criminal justice system’s response to crime may be, such a response should not on this approach be based on attributing moral responsibility or non-responsibility to the defendant.

72. See FREE WILL AND ILLUSION, *supra* note 64, at 85.

73. See Moore, *supra* note 71, at 489; see also R.B. Brandt, *The Insanity Defense and the Theory of Motivation*, 7 L. & PHIL. 123, 124 (1988) (“More specifically, what the criminal law should aim to do is to maximize the general well-being, and to minimize the damage of crime and anxiety about the possibility of crime, at least human cost . . .”).

74. See DERK PEREBOOM, LIVING WITHOUT FREE WILL 174–75 (2001).

75. Much more broadly, see J. Angelo Corlett, *Forgiveness, Apology, and Retributive Punishment*, 43 AM. PHIL. Q. 25 (2006). Professor Corlett argues that:

[I]f causal determinism is true in the hard deterministic sense, then there is no sense to be made of ethics and moral responsibility, and not even moral practices such as forgiving others make much, if any, sense. For we only forgive those who are blameworthy for wrongful behavior, not those who could not have done otherwise . . . . So forgiving such “persons” seems to make little or no sense (if “persons” is not too flattering a term for them in a completely deterministic world).

*Id.* at 35.

76. See generally Gideon Rosen, *Skepticism About Moral Responsibility*, 18 PHIL. PERSP. 295 (2004).

77. See *id.* at 295.

78. See *id.*

One problem with this “agnostic” approach is that it may be difficult, in practice if not in theory, to suspend judgment as to the defendant’s moral responsibility, while not treating the defendant, for practical purposes, as non-responsible. That is, we would want to clarify the real, practical difference between withholding judgments as to responsibility, and making implicit judgments of non-responsibility. Do both approaches in fact reduce to some sort of utilitarianism in the criminal justice system? And in practice, it may be difficult to permit our impressions, one way or another, of a defendant’s responsibility to have no impact on our ultimate disposition of the case.

As well, if suspending judgments as to the defendant’s moral responsibility could have moral advantages, why could such a practice not have moral disadvantages as well? For one, leaving convicted and—on some basis—sentenced persons in a haze of officially undetermined responsibility may tend to shift the real decision-making, even on matters of any possible fundamental dignity, away from legislators, judges, and juries, and toward the administrators of whatever systems the convicted defendant is assigned to.

The legal system could bypass these serious concerns if we could be sure that determinism does not hold, at levels relevant to human behavior.<sup>79</sup> But we would then also have to remember that a failure of determinism would not guarantee a meaningful realm of freedom of the will, moral responsibility, or of sane and responsible behavior. Substituting a sphere of sheer randomness for determinism hardly suffices. Random mental processes and outcomes hardly suggest meaningful free will.<sup>80</sup> Human decisions with their origins in random processes seem no more a matter of genuine free will than “decisions” that reflect the mechanisms of some deterministic process. As Professor Peter van Inwagen has phrased the argument, “if how an agent acts is a matter of chance, the agent can hardly be said to have free will.”<sup>81</sup> A complex machine that “acted” purely at random would hardly qualify for legal responsibility.<sup>82</sup>

Thus a world of randomness, or pure chance, or even a world combining determinism in some spheres with spheres of randomness, would hardly seem to be fertile ground for a meaningful distinction between typi-

---

79. For merely one recent exploration of both measurement-induced uncertainties at a microscopic level, as well as more fundamental indeterminacies, or undefined values, that seem to be inescapably hard-wired into nature itself, see Lee A. Rozema et al., *Violation of Heisenberg’s Measurement-Disturbance Relationship by Weak Measurements*, 109 PHYSICAL REV. LETTERS 100404, 100404-1 (2012).

80. *See id.*

81. *See generally* Peter van Inwagen, *Free Will Remains a Mystery: The Eighth Philosophical Perspectives Lecture*, 34 NOÛS 1 (14 PHIL. PERSP. 1) (2000).

82. Or, for that matter, be a responsible agent, or even an agent in general, or a genuine actor, at all. *See id.* at 10. For a recent critique of Professor van Inwagen’s own views, see Lara Buchak, *Free Acts and Chance: Why the Rollback Argument Fails*, 63 PHIL. Q. 20 (2013).



cally morally responsible sane defendants, and typically morally non-responsible insane defendants.

This is of course not to try to assess the strength or weakness of determinist theories, or of world views with a randomness component. The important point is simply that the current popularity of such views, and the difficulties of accounting for an insanity defense on any such bases, again reflects the fraying and unraveling of the jurisprudence of the insanity defense.

## 2. *The Complications of Compatibilism and Further Doctrinal Unraveling*

The temptation to try to somehow reconcile moral responsibility, free will, and the familiar distinction between legal sanity and insanity with a purely determinist worldview is, for many contemporary thinkers, overwhelming. The belief that a form of free will, and usually the distinction between responsibility and non-responsibility, can be reconciled with universal determinism is the view known as “compatibilism.”<sup>83</sup>

How free will, responsibility, and the sanity/insanity distinction are to be reconciled with determinism is approached by the various compatibilists in different ways. The complications are almost innumerable, and will not be explored herein. Familiar psychology ordinarily finds a meaningful distinction between, for example, the sane and insane;<sup>84</sup> the addicted and the unaddicted;<sup>85</sup> compulsive hand washers and those who are not;<sup>86</sup> and those cooperating in crime only under a credible threat of imminent death and those far removed from such obvious, direct physical coercion.<sup>87</sup> The problem is to somehow validate the basic differences in how we judge and treat all of the above categories if we assume that each category is subject to a relentless causal determinism.<sup>88</sup>

Let us very briefly consider possible compatibilist solutions. Can a distinction between compulsion or coercion on the one hand, and mere universal causal determinism on the other, somehow validate the basic ways in which the law differently treats those it regards as sane and insane?<sup>89</sup> Surely there is some sort of difference between coercion at gunpoint and the workings of neurological processes. Or we might try another tack: Would a distinction between those of our causally determined acts that we somehow identify with, accept, or endorse, as distinct from those we do not, better serve to distinguish the legally sane and in-

---

83. See, e.g., Alfred R. Mele, *EFFECTIVE INTENTIONS: THE POWER OF CONSCIOUS WILL* 153 (2009).

84. See *id.* at 154.

85. See *id.*

86. See *id.*

87. See *id.*

88. See *id.*

89. See *id.*

sane?<sup>90</sup> But what if some insane persons strongly identify with beliefs the law considers to be insane beliefs? What if those identifications themselves, though not socially coerced, are part of the universal deterministic network of causes and effects? It seems reasonable to believe that not every such endorsement of one's legally insane beliefs necessarily must be a further element of one's insanity.

This is again not to attempt to evaluate any form of compatibilism.<sup>91</sup> The aim is instead merely to suggest that the contested popularity of compatibilist views and their rivals amounts to another dimension along which the standard logic of the insanity defense is not only fraying, but unraveling.

To its credit, modern compatibilism has not tried to reduce moral and legal responsibility, or the insanity defense in particular, to the narrowly cognitive question of merely what the criminal defendant knew or understood. Compatibilists typically recognize the relevance as well of emotions, desires, drives, and cravings to determinations of sanity and responsibility. They more specifically recognize that:

[To] be morally responsible someone must not be influenced exclusively by some irrational desire but must exhibit some kind of rationality in the formation of their intentions. A psychopath or someone acting solely under the influence of an overwhelming craving for heroin or agoraphobia without having any moral beliefs about the matter is not . . . morally responsible.<sup>92</sup>

The standard approaches to the insanity defense, apart from those that refer to the possibility of an "irresistible impulse,"<sup>93</sup> instead often

90. See, e.g., JOHN MARTIN FISCHER & MARK RAVIZZA, *RESPONSIBILITY AND CONTROL: A THEORY OF MORAL RESPONSIBILITY* 11 (1998); Harry G. Frankfurt, *Freedom of the Will and the Concept of a Person*, 68 J. PHIL. 5 (1971). For some complications, see, e.g., Neil Levy, *Luck and History-Sensitive Compatibilism*, 59 PHIL. Q. 237 (2009). For discussion of Fischer & Ravizza's "moderate (moral) reasons-responsiveness" criterion, see, e.g., Neal Judisch, *Reasons-Responsive Compatibilism and the Consequences of Belief*, 11 J. ETHICS 357 (2007); Gary Watson, *Reasons and Responsibility*, 111 ETHICS 374 (2001).

91. For a concise overview of the major forms of compatibilism, see Michael McKenna, *Compatibilism*, STAN. ENCYCLOPEDIA PHIL. (Edward N. Zalta ed., rev. ed. 2009) (2004), available at <http://plato.stanford.edu/entries/compatibilism> (last visited Feb. 15, 2014). For the classic critique of compatibilism, see IMMANUEL KANT, *CRITIQUE OF PRACTICAL REASON* § 5.96, at 80–81 (Mary J. Gregor ed., 1997) (1788) (referring to compatibilism as "wretched subterfuge" or as "petty word-jugglery"). For a contemporary critique of compatibilism, see, e.g., FREE WILL AND ILLUSION, *supra* note 64, at 49–53 ("There are *two* major inherent difficulties with compatibilism: the shallowness of desert and value, and the issue of injustice and victimization."). "Without libertarian free will, no one is ultimately responsible for being a criminal . . ." *Id.* at 49.

92. RICHARD SWINBURNE, *MIND, BRAIN, AND FREE WILL* 216 (2013) (characterizing predominant view among contemporary compatibilists).

93. For discussion, see, e.g., Carrido, *supra* note 53, at 314–17, 316 nn.60–64 (citing *Orndorff v. Commonwealth*, 691 S.E.2d 177, 179 n.5 (Va. 2010)); see also *Morgan v. Commonwealth*, 646 S.E.2d 899, 902 (Va. App. 2007) (explaining irre-

seem to focus in large measure on broadly cognitive matters. These matters include what the defendant knew or had the “capacity” to know, understand, grasp, judge, or appreciate regarding facts and the relevant legal standard.<sup>94</sup> A defendant’s knowledge and understanding differ from the defendant’s impulses, drives, motives, desires, and intentions. Knowledge and understanding also differ from acting and willing. Compatibilists often require, beyond the standard merely cognitive elements of sanity and insanity, something like intention, willing, and defendants’ “adequate capacity and opportunity” not to have acted as they did.<sup>95</sup>

Requiring something like a defendant’s capacity to have acted otherwise typically makes sense. Would we really deny the possibility of an insanity plea to a defendant—regardless of what the defendant knew or understood—if the defendant were held bound by rigidly fixed “internal” or “external” constraints upon what the defendant could will or intend? Would we hold sophisticated but deterministically programmed robots legally responsible if we believed that they “knew” and “understood” what they were quite unavoidably doing?<sup>96</sup> Presumably not. But, crucially, this judgment, extended to human defendants, would not depend upon whether the defendant suffered from anything we called a “disease” or not.

At some point, the courts must take at least an implicit stand on the basic claim of compatibilism: that even strict causal determinism in the broad natural realm is compatible with meaningful moral and legal responsibility. More specifically, courts must at some point eventually consider whether, in a determined universe, defendants are capable of being

---

sistible impulse defense as requiring specifically that “disease” have been responsible for defendant’s total, and not merely partial, inability to control his actions).

94. See, e.g., *supra* note 7 (discussing classic *M’Naghten* insanity test). Compare the standard judicial test for due process in the context of competency to stand trial, which includes, among other elements, that the defendant not have “a mental condition . . . such that he lacks the capacity to understand the nature and object of the proceedings against him.” *United States v. Mitchell*, 709 F.3d 436, 439 (5th Cir. 2013) (quoting *Drope v. Missouri*, 420 U.S. 162, 171 (1975)).

95. Moore, *supra* note 59, at 243. While Professor Moore seeks to bypass broad questions of the truth or falsity of determinism, he expresses one element of the remaining problem in these terms: “Neuroscience’s concrete version of the lack of free will challenge(s) should be put as: *in whatever sense and to whatever extent macro-sized natural events and natural processes are caused*, so too are both human actions, and the choices that cause them, caused by brain states.” *Id.* at 263. *But cf.* FISCHER & RAVIZZA, *supra* note 90, at 39.

96. See, e.g., Morse, *supra* note 56, at 116–17. Professor Morse argues, however, that free will is not crucial to criminal responsibility, and that the “folk psychology” on which the criminal law is largely based is compatible with causal determinism, even though the law presumes the capacity of most persons to both understand and to be genuinely guided by good reasons. See *id.* at 117, 119. *But see supra* notes 33–38 and accompanying text. The idea of being moved by a “reason” is itself murky. For reference to a narrow contrary phenomenon called automatism or unconscious behavior, see, e.g., *Smith v. State*, 663 S.E.2d 155, 156–57 (Ga. 2008).

moved by and responsive to what are called “reasons.”<sup>97</sup> Again, it is not our aim to take sides on this or any other relevant controversy. But this controversy over the meaning and role of reasons will inevitably further contribute to the fraying and unraveling of our standard insanity jurisprudence.

Consider in this context an obvious question for the compatibilists, and others, who claim that a capacity to be moved by reasons should often be decisive in insanity and other kinds of cases. Are reasons in this sense simply links in a naturalistic, perhaps entirely determined, chain of causes and effects, unavoidably producing the defendant’s act? That would preserve the naturalistic and deterministic elements of the challenge to the traditional logic of insanity. But to do the work required of what we call a “reason,” must not reasons also amount to abstractions as well? That one’s acting in a particular way would result in more good than harm, for example, could be a reason for acting in that way. But that hypothetical effect of a possible future action is in itself just an abstract idea. How does a pure abstract idea insinuate itself, effectively, into an ongoing chain of “natural” or material causes and effects?<sup>98</sup>

If, in sum, we “naturalize” our idea of a reason, it becomes far more difficult to explain why a reason—let alone, some sort of more general capacity for reasoning, in this naturalistic sense—should make any crucial difference to cases of insanity versus sanity, and of legal culpability. Someone with the capacity to reason in the naturalized sense is simply lucky to have that capacity, if we regard it as a good thing, or unlucky if we do not.<sup>99</sup> Luck would seem to be a doubtful ground for imposing or not imposing criminal blame. And if it is argued that some people work in order to develop sufficient reasoning capacity to become morally responsible, the troubling answer may then be that having that very capacity, or that very disposition, is itself a matter of luck.<sup>100</sup> Regardless of how one

---

97. For further discussion from among the large relevant literature, see R.A. Duff, *What Kind of Responsibility Must Criminal Law Presuppose?*, in *FREE WILL AND MODERN SCIENCE* ch. 11, at 178 (Richard Swinburne ed., 2011); Stephen J. Morse, *Rationality and Responsibility*, 74 *S. CAL. L. REV.* 251 (2000); Christopher Šlobogin, *An End to Insanity: Recasting the Role of Mental Disability in Criminal Cases*, 86 *VA. L. REV.* 1199, 1227–31 (2000).

98. See SUSAN BLACKMORE, *CONVERSATIONS ON CONSCIOUSNESS* 198, 205 (2007) (conversation with John Searle) (discussing idea that genuinely rational deciding and acting require free will); see also Peter Westen, *Getting the Fly Out of the Bottle: The False Problem of Free Will and Determinism*, 8 *BUFF. CRIM. L. REV.* 599, 636 (2005).

99. For a very concise treatment of this and related issues by the renowned contemporary philosopher, see Galen Strawson, *Luck Swallows Everything*, *TIMES LITERARY SUPPLEMENT* (June 26, 1998), <http://www.naturalism.org/strawson.htm>.

100. See *id.* Relatedly, it is sometimes argued that a defendant’s insanity or insufficient capacity for reason, in whatever sense, means that judicial punishment is unlikely to deter future antisocial conduct by that defendant. See, e.g., LAFAVE, *supra* note 8, at 393; Arenella, *supra* note 36, at 273 (“futility”). Whether this traditional belief about insanity is actually well-founded, or is instead a fragile dogma of traditional insanity jurisprudence, amounts to yet another point of fraying and future unraveling.

thinks of these arguments, the unraveling of our traditional ideas of sanity and responsibility must take this additional form.

3. *Illusionism and Fictionalism About Free Will as a Final Source of Insanity Defense Unraveling*

Charles Darwin, in elaborating on the text chosen for our epigram,<sup>101</sup> argued as follows:

We cannot help loathing a diseased offensive object, so we view wickedness. . . . Yet it is right to punish criminals; but solely to *deter* others. . . . [O]ne deserves no credit for anything. . . . [N]or ought one to blame others. This view will do no harm, because no one can be really *fully* convinced of its truth, except [the] man who has thought very much, [and] he . . . will not be tempted, from knowing every thing [sic] he does is independent of himself, to do harm.<sup>102</sup>

Darwin thus crucially acknowledges that our legal and moral blaming practices do not make sense by our own ordinary standards. Still, those practices should, supposedly, remain stable and effective, because the few of us who grasp Darwin's insights will continue to promote the perceived common good, and the rest of us will go on as before, unaware of the fundamental illusions embedded in our common sense views.

In this regard, Darwin is a forerunner of the increasingly influential schools of thought known respectively as illusionism<sup>103</sup> and fictionalism.<sup>104</sup> The merits, and the stability in practice, of illusionist and fictional-

101. See *supra* note 1 and accompanying text.

102. DARWIN, *supra* note 1, at 608 (entry of September 6, 1838). For brief discussion, see Galen Strawson, *The Impossibility of Ultimate Responsibility?*, in FREE WILL AND MODERN SCIENCE ch. 8, at 135 (Richard Swinburne ed., 2011).

103. See, e.g., FREE WILL AND ILLUSION, *supra* note 64; Saul Smilansky, *Free Will and Moral Responsibility: The Trap, the Appreciation of Agency, and the Bubble*, 16 J. ETHICS 211, 217 (2012) (“[I]llusory beliefs are in place, and . . . the role they play is largely positive. Humanity is fortunately deceived on the free will issue, and this seems to be a condition of civilised [sic] morality and personal value.”); Smilansky, *supra* note 64. For a brief critique of Smilansky, see, e.g., OWEN M. FLANAGAN, *THE REALLY HARD PROBLEM: MEANING IN A MATERIAL WORLD* 36 (2007); WALLER, *supra* note 44; WEGNER, *supra* note 63; Cashmore, *supra* note 62; Greene & Cohen, *supra* note 36, at 1783 (“Free will, as we ordinarily understand it, is an illusion. . . . As consequentialists, we can hold people responsible for crimes simply because doing so has, on balance, beneficial effects through deterrence, containment, etc.”). This latter claim should be, and likely will be, increasingly controversial among ethicists, judges, attorneys, and legal commentators, in and beyond the context of insanity defenses.

104. The broader approach to fictionalism in various areas of philosophy, including moral philosophy generally, as distinct from a fictionalism regarding claims made specifically about insanity and criminal responsibility, is illustrated in MARK ELI KALDERON, *MORAL FICTIONALISM* (2005) and RICHARD JOYCE, *THE MYTH OF MORALITY* (2007). For the only vaguely related idea of “legal fictions,” see classically LON L. FULLER, *LEGAL FICTIONS* (1967).

ist approaches to criminal responsibility and to insanity in criminal contexts in particular are unclear, but potentially important.<sup>105</sup> The rise of various forms of fictionalism and illusionism contributes to the unraveling of confidence in traditional insanity jurisprudence.

Some might reply that at least in some areas of the law, the idea of a “legal fiction,” in the sense classically explored by Lon Fuller, has generated only limited controversy. Professor Fuller thought of legal fictions as useful or expedient,<sup>106</sup> but as neither truthful, nor as lies, nor as mere error,<sup>107</sup> and perhaps not even as consciously false<sup>108</sup> assumptions. In the context of criminal responsibility and insanity, though, it is far from clear that broad, widely, known basic judicial falsification, conscious or otherwise, will somehow avoid eventually unraveling. Punishing, or declining to punish, where the logic of doing so is admittedly basically defective, if in some sense still supposedly useful, is likely to seem in other respects unfair, and to be unstable over time.

In the end, the various versions of illusionism and fictionalism offer not a stable solution to insanity defense dilemmas, but only a further, and for our purposes final, aspect of the gradual unraveling of traditional insanity defense doctrine.

#### IV. CONCLUSION: INSANITY DEFENSE DOCTRINE AS INCREASINGLY A MATTER OF THREADS UNRAVELING

For the reasons explored above, it should now be clear that the most crucial difficulties with the insanity defense are not at the level of competing definitions, problems of proof, or of the mere existence of complex or difficult cases. These sorts of problems attach to all kinds of litigation, and have existed in one form or another at most times and in most places. The unstable condition of the current law of the insanity defense is different. The current law of insanity, admittedly, may well make some viable and useful distinctions. But despite these viable distinctions, the most fundamental distinction—that between persons who are legally classified as insane, and thus presumably as not bearing criminal moral responsibility, and persons not so classified, and thus who are presumably criminally responsible—is in crucial respects fraying and unraveling, for good or ill.

---

105. Illusionist and fictionalist approaches to civil, as distinct from criminal, insanity may or may not have immediately dramatic consequences. *See, e.g.,* *Dougherty v. Rubenstein*, 914 A.2d 184, 193 (Md. Ct. Sp. App. 2007) (will contest case based on testator’s alleged insanity); *see also* *Prop. & Cas. Ins. Co. of Hartford v. Davenport*, 907 F. Supp. 2d (D. Vt. 2012) (insurance coverage limitations case).

106. *See* FULLER, *supra* note 104, at 7.

107. *See id.* at 5.

108. *See id.* at 7.

This unraveling process does not impeach everything we might want to say for any purpose about mental illness,<sup>109</sup> mental capacity,<sup>110</sup> or the determination of sanity and insanity.<sup>111</sup> But we need to be collectively aware of the complex unraveling process discussed above. Where possible, we should try to manage the unraveling process in such a way as to validate our worthiest intuitions about fairness in the criminal justice system. Most assuredly, this unraveling process should not prevent us from identifying, critiquing, and remedying relevant structural inequalities, and economic and social injustices more broadly. A proper concern for structural and institutional justice is broader than the context of the insanity defense. But as is illustrated above, the law of insanity faces fundamental conflicts over even the near term future. This unraveling process calls for our most thoughtful collective responses.

---

109. See, e.g., *Durrence v. State*, 695 S.E.2d 227, 230 n.2 (Ga. 2010); see also David Kupfer & Darrel Regier, *To DSM-5 User Community*, AM. PSYCH. ASS'N, [www.dsm5.org/Pages/Default.aspx](http://www.dsm5.org/Pages/Default.aspx) (last visited Feb. 15, 2014) (discussing revisions in mental illness classifications). But cf. Thomas Insel, *Transforming Diagnosis*, NAT'L INST. MENTAL HEALTH: DIRECTOR'S BLOG (Apr. 29, 2013), [www.nimh.gov/about/director/2013](http://www.nimh.gov/about/director/2013).

110. See, e.g., *Durrence*, 695 S.E.2d at 230 n.3.

111. See the interesting New Hampshire approach summarized in *State v. Fichera*. See *State v. Fichera*, 903 A.2d 1030, 1035 (N.H. 2006) (“[T]he test for insanity does not define or limit the varieties of mental diseases or defects that can form the basis for a claim of insanity.” (citation omitted)).