1-2018

Agency and Insanity

Stephen P. Garvey
Cornell Law School, spg3@cornell.edu

Agency and Insanity

STEPHEN P. GARVEY†

INTRODUCTION

Around four o'clock on January 20, 1843, Edmund Drummond, private secretary to Prime Minister Sir Robert Peel, was walking down Parliament Street, headed to Downing Street. Another man approached Drummond from behind, withdrew a pistol, put it to his back, and pulled the trigger. A policeman saw what happened, rushed over, and seized the assailant, who had in the meantime reached for another pistol. The assailant's name was Daniel M'Naghten,[1] a wood-turner from Glasgow.[2] M'Naghten, it turned out, had mistaken Drummond for his real target, Prime Minister Peel. Drummond died five days later. The ensuing case against M'Naghten would have a profound, and lasting, impact on the law of insanity in the Anglo-American world.[3]

---

† Professor of Law, Cornell Law School. Earlier versions were presented to a small group of young criminal law scholars at Northwestern University School of Law and to an audience at the Cleveland-Marshall College of Law. I thank all those present on those occasions for their thoughtful comments.

1. M'Naghten's name is spelled in different ways in different sources. *See, e.g.*, RICHARD MORAN, KNOWING RIGHT FROM WRONG: THE INSANITY DEFENSE OF DANIEL MCNAUGHTAN xi–xiii (1981); Bernard L. Diamond, *On the Spelling of Daniel M'Naghten's Name*, 25 OHIO ST. L.J. 84 (1964). The spelling followed here is the one reflected in the case bearing his name. M'Naghten's Case (1843) 8 Eng. Rep. 718. For accounts of the case, see, for example, DANIEL MCNAUGHTON: HIS TRIAL AND THE AFTERMATH (Donald J. West & Alexander Walk eds., 1977); MORAN, *supra*; RICHARD D. SCHNEIDER, THE LUNATIC AND THE LORDS (2009).

2. *See* SCHNEIDER, *supra* note 1, at 143.

3. Most U.S. jurisdictions today continue to define insanity in ways traceable to the original formulation of the *M'Naghten* Rules. *See* Paul H. Robinson et al., *The American Criminal Code: General Defenses*, 7 J. LEGAL ANALYSIS 37, 77 (2015) (stating that the "majority view" of the insanity defense provides that "[a]n

M'Naghten was delusional. He attacked Peel because he believed members of Peel's Tory party were out to get him, relentlessly persecuting him. In reality, of course, the Tories weren't after him. M'Naghten, however, wasn't in reality. He was in a world of his own, at least when it came to the Tories. Earlier, M'Naghten had told the Glasgow police commissioner that "he was the object of some persecution, and . . . that he thought it proceeded from the priests at the Catholic chapel in Clyde Street, who were assisted by a parcel of Jesuits."[4] Two days later, he told the commissioner that the "Tories had joined with the Catholics,"[5] and thereafter it was the Tories who, as some witnesses later put it, "haunted" him.[6] Today, M'Naghten would likely be diagnosed as suffering from paranoid psychosis.[7]

"The delusions were of such an intensity," M'Naghten's father testified at trial, "that the Tories were with him day and night spying on him . . . they were there every time he turned around . . . they laughed at him and shook their fists in his face."[8] M'Naghten went to the police for help. He also tried to get help from Sir James Campbell, the Lord Provost of Glasgow,[9] and Alexander Johnston, Glasgow's Member of Parliament,[10] all to no avail. According to Johnston, M'Naghten "complained of being attacked through the newspapers, and said the persons of whom he complained

---

actor is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he did not know his conduct was wrong.").

4. SCHNEIDER, *supra* note 1, at 208 (citing transcript).

5. *Id.* at 209.

6. *Id.* at 204 (Jane Drummond Patterson, in whose house M'Naghten had lodged two years before the trial, testified that M'Naghten, after three or four months, mentioned being "haunted by devils."); *id.* at 207–08 (testimony of Rev. Alexander Turner that M'Naghten "talked about being haunted" by a number of persons).

7. *See id.* at 253.

8. *Id.* at 254.

9. *See id.* at 207.

10. *See id.* at 205.

followed him night and day . . . . [H]e thought his persecutors would be satisfied with nothing less than his life."[11] M'Naghten, so far as one can tell, apparently believed he had no choice. Killing the Tory prime minister was the only way out, a last desperate attempt to end the torment.

Tried for murder, M'Naghten was acquitted.[12] When the defense finished presenting its witnesses, the presiding judge, Chief Justice Tindal, asked Solicitor General William Follett: "[A]re you prepared, on the part of the Crown, with any evidence to combat this [defense] testimony . . . , because we think if you have not, we must be under the necessity of stopping the case. Is there any medical evidence on the other side?"[13] Follett threw in the towel. "No, my Lord,"[14] he replied. Justice Tindal then all but directed an acquittal by reason of insanity.[15] M'Naghten was sent to Bethlem Hospital, also known as Bedlam.[16] He was eventually transferred to Broadmoor Asylum, where he died in 1865.[17]

M'Naghten's case would eventually produce a test for insanity bearing his name. The *M'Naghten* Rule, at least in its modern formulations, equates insanity with cognitive incapacity. Sometimes the Rule is combined with a test equating insanity with volitional incapacity, often called the "irresistible impulse" test. Together, these twin incapacities constitute what we might call the law's traditional test for insanity. Two alternative tests, proposed but never enacted into law, equate insanity in some way with irrationality. Yet

---

11. *Id.* at 205–06.

12. *See id.* at 225. M'Naghten had savings of £750 (probably somewhere between $67,000 to $97,000 today) and was able to hire the Victorian equivalent of a defense dream-team. *See id.* at 55; MORAN, *supra* note 1, at 12.

13. *Id.* at 222. Justice Tindal would later speak for the judges in announcing what became known as the *M'Naghten* Rules.

14. *Id.*

15. *See id.* at 223–24.

16. *See id.* at 225.

17. *See id.* at 265–66.

neither the traditional test, nor the irrationality proposals, are entirely persuasive.[18] Each faces fair-minded objections. But if insanity isn't incapacity or irrationality, what is it? One possibility, described and explored here, proposes that insanity be understood at bottom as a *defect of consciousness*, and in particular, as a *lost sense of agency*. If so, then perhaps insanity belongs in the same family as other such defects, like sleepwalking, hypnosis, and multiple personality. If M'Naghten was insane, then according to the lost agency theory, it wasn't really M'Naghten who killed Drummond. That sounds paradoxical, but the lost agency theory tries to explain why it's not.

The argument unfolds in four moves. Part I tackles and finds wanting the traditional test of insanity, which portrays insanity as an incapacity, either cognitive or volitional. Part II develops two different versions of the irrationality test, and finds them wanting too. Part III introduces insanity as lost agency, describing its basic features and continuity with other defects of consciousness. If the lost agency theory is true, then insanity excuses because the insane don't author the thoughts and actions their minds and bodies produce. They're not the agent of the crimes they commit. Part IV returns to Daniel M'Naghten, asking whether the lost agency theory can account for the intuition that he was indeed insane when he shot and killed Edmund Drummond on Parliament Street in 1843.

---

18. Of course, some evidence suggests that the particular test jurors are told to apply actually makes little difference to the verdict they return. Jurors tend to return the same verdict whatever the test. *See, e.g.*, HENRY M. STEADMAN ET AL., BEFORE AND AFTER HINCKLEY: EVALUATING INSANITY DEFENSE REFORM 61 (1993) (finding that California's move from the MPC test to *M'Naghten* in 1982 did "not affect use of the insanity defense"); James R.P. Ogloff, *A Comparison of Insanity Defense Standards on Juror Decision Making*, 15 LAW & HUM. BEHAV. 509, 526 (1991) ("[F]or whatever reason, the particular insanity defense standards employed do not seem to strongly influence a juror's decision making."). Even so, we should still try to understand in what insanity consists, even if the legal rules we formulate in an effort to capture or define it tend in the end to make little difference to how jurors decide concrete cases.

I. TRADITION

When courts announce new rules of law they do so in the course of deciding concrete cases or controversies. Usually, that is. The most famous test for insanity, the *M'Naghten* Rule,[19] wasn't born in the usual way. The Queen at the time, a young Queen Victoria, had been an assassin's target three years before M'Naghten tried to kill her prime minister.[20] Her assailant, Edward Oxford, had been acquitted on grounds of insanity, and the Queen was none too happy when M'Naghten was acquitted on the same ground. She told the House of Lords to ask the fifteen judges of the common law courts to explain, once and for all, what made a person insane in the eyes of English law. Their pronouncement became known as the *M'Naghten* Rule. Pass the Rule's test, and you go to a hospital; fail, and you go to prison (or worse).

Criminal lawyers tend to talk about *the M'Naghten* Rule or test, but that's misleading. In reality, we have today, not one *M'Naghten* test, but two. The *old M'Naghten* is the one the judges of the common law courts announced in 1843. The *new M'Naghten* is a recognizable, modern-day descendant of the old—encountered in many state penal codes, as well as the federal code. The two belong to the same family, or at least share the same family tree, but the words of the old *M'Naghten* differ from the words of the new in more ways than one.

Five questions were put to the judges in *M'Naghten's Case*. Their answers to three of the five have for some reason faded into history.[21] What's passed into the modern canon is

---

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

20. *See* PAUL THOMAS MURPHY, SHOOTING VICTORIA: MADNESS, MAYHEM, AND THE REBIRTH OF THE BRITISH MONARCHY (2012).

21. Those who attend to the judges' answers to *all* the questions naturally tend to speak about the *M'Naghten* Rules (plural), while those who focus only on the answer to the second and third questions, dealing with the "proper questions to be submitted to the jury, when a person alleged to be afflicted with insane delusion . . . is charged with the commission of a crime," M'Naghten's Case, 8

the single answer they gave to questions two and three. According to the judges' answer to those questions, a jury faced with a plea of insanity should be instructed as follows:

> [T]o establish a defence on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from 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.[22]

So says the old *M'Naghten*. The test is easier to state than to understand. Books have been devoted to deciphering it.[23] Its meaning remains elusive despite all that effort. Indeed, maybe it doesn't even state an all-purpose test for insanity, good for any case in which insanity is alleged. The judges might have intended to limit their answers to cases

---

Eng. Rep. at 720, tend to speak of the *M'Naghten* Rule (singular).

22. *Id.* at 722. In one commonly neglected passage, the judges expressly say, in response to the *fourth* question, that someone "labour[ing] under . . . partial delusion only, and . . . not in other respects insane, . . . must be considered in the same situation as to responsibility as if the facts with respect to which the delusion exists were real." *Id.* at 723. According to at least two influential treatises, this language doesn't add anything to the canonical language cited in the text. *See* 1 WAYNE R. LAFAVE & AUSTIN W. SCOTT, SUBSTANTIVE CRIMINAL LAW § 4.2(b)(5), at 445 (1986) ("[I]t is undoubtedly fair to conclude that this particular part of *M'Naghten* [i.e., the answer to question four] does not set up a unique formula differing from the right-wrong test."); A.P. SIMESTER ET AL., SIMESTER AND SULLIVAN'S CRIMINAL LAW: THEORY AND DOCTRINE § 19.1(ii)(f), at 723 (5th ed. 2013) ("This express provision for partial delusions [in response to question four] does not appear to add anything to the substance of the Rules."). However the judges understood the relationship between their answer to the second and third questions, and their answer to the fourth question, their answer to the fourth question tells us a person suffering from delusions is insane if, assuming his delusions had been true, he would not have been criminally liable, which is the same test associated with the delusion theory (discussed below). *See infra* notes 70, 74 and accompanying text.

23. For an in-depth analysis of the old *M'Naghten*, see HERBERT FINGARETTE, THE MEANING OF CRIMINAL INSANITY 234–42 (1972). Fingarette's theory of insanity is commonly characterized as an irrationality theory. *See infra* notes 47–48 and accompanying text. Moreover, because Fingarette "read[s] the *M'Naghten* test as saying much the same thing" as the theory he proposes, FINGARETTE, *supra* at 197, perhaps the old *M'Naghten* should be understood as an irrationality test as well.

involving defendants who, like Daniel M'Naghten, were suffering from what were known, circa 1843, as "partial delusions."[24] We won't add to the existing commentary on (and confusion over) *M'Naghten*'s original meaning. Instead, we'll set the old *M'Naghten* aside, and move onto the new *M'Naghten*.

The new *M'Naghten* comes in different shapes and sizes. Different jurisdictions formulate the test in slightly different ways. For now, let's work with the following generic statement: a person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks capacity to know the criminality of his conduct.[25] So stated, the new *M'Naghten* differs (at least superficially) from the old in a few ways, only one of which needs emphasis here.[26] The old *M'Naghten* made no explicit

---

24. M'Naghten's Case, 8 Eng. Rep. at 721–22.

25. This generic statement of the rule is, of course, similar to the corresponding language of Model Penal Code § 4.01(1). It doesn't reflect the law of any particular jurisdiction. For surveys of the law in U.S. jurisdictions, see Paul H. Robinson & Tyler Scot Williams, Mapping American Criminal Law: Variations Across the Fifty States ch. 14 (Jan. 2, 2017) (unpublished manuscript) (on file with the University of Pennsylvania Law School: Legal Scholarship Repository); Robinson *et al.*, *supra* note 3. Among other things, the generic statement sidesteps the protracted debate over the meaning of the word "wrong." *See* Walter Sinnott-Armstrong & Ken Levy, *Insanity Defenses*, *in* OXFORD HANDBOOK ON THE PHILOSOPHY OF THE CRIMINAL LAW 302–06 (John Deigh & David Dolinko eds., 2011) (describing all imaginable meanings of the word "wrong"). It simply assumes "wrong" means "criminally wrong," which is how it has been construed in the land where *M'Naghten* was born. *See* R v. Windle [1952] QB 826 at 833 (Eng.). ("The test must be whether an act is contrary to law"); ANDREW ASHWORTH & JEREMY HORDER, PRINCIPLES OF CRIMINAL LAW § 5.2(b), at 143 (7th ed. 2013).

26. The new *M'Naghten* differs from the old in two other noteworthy ways. First, it discards any reference to the accused's knowledge of the "nature and quality" of what he was doing. This prong is commonly omitted because, it's thought, it doesn't add much, if anything, to the second prong. *See, e.g.*, Clark v. Arizona, 548 U.S. 735, 753–54 (2006) ("In practical terms, if a defendant did not know what he was doing when he acted, he could not have known that he was performing the wrongful act charged as a crime."). Second, the new *M'Naghten* discards any reference to "defect of reason." This element is frequently omitted because, it's probably thought, it adds nothing to the disease-of-the-mind element. Why require a "defect of reason, *from* disease of the mind"? Aren't they

mention of capacity. It asked if the actor was laboring under a defect of reason, from disease of the mind, so as not to know the nature and quality of what he was doing, or if he did know it, that he did not know he was doing what was wrong. The new *M'Naghten*, in contrast, puts capacity front and center. An actor isn't insane unless, as a result of mental disease or defect, he not only didn't realize he was committing a crime, but was *powerless* to so realize.[27] What defeats liability, according to the new *M'Naghten*, is an actor's ignorance of the law, provided it resulted from a mental disease or defect, and provided he was *powerless* to be anything but ignorant.

Before going on, we should confront Daniel M'Naghten's responsibility directly. Ignore for a moment what you think the verdict on M'Naghten's sanity would be under this or that legal test, and consult your intuitions. Should M'Naghten have been found criminally responsible or non-responsible? Was Bethlem the right place to send him? Or should he have gone to the gallows? If you agree with many thoughtful commentators, you'd say Bethlem. M'Naghten was, to put it bluntly, obviously crazy.[28] If so, then Daniel

---

more or less the same thing? *See* FINGARETTE, *supra* note 23, at 178. Fingarette, for one, thought not, calling it a "profound mistake" to omit the defect-of-reason element from the test's formulation. *Id.* at 198.

27. *See* Sinnott-Armstrong & Levy, *supra* note 25, at 311 ("[T]he *M'Naghten* rule hinged on actual knowledge rather than ability to know."). The Supreme Court slipped easily from the original non-capacity language of the old *M'Naghten* to the capacity language of the new *M'Naghten*, without noting the difference. *See Clark*, 548 U.S. at 747 (describing the "second part" of the original language from *M'Naghten* as a test for "lack of moral capacity").

28. *See* MORAN, *supra* note 1, at 4 (describing this as the "conventional wisdom" and citing the work of Fingarette, Biggs, Goldstein, and Rollin); Michael Moore, *The Quest for a Responsible Responsibility Test: Norwegian Insanity Law After Breivik*, 9 CRIM. L. & PHIL. 645, 662 (2015) [hereinafter Moore, *After Breivik*] ("Daniel M'Naghten himself . . . would be considered in popular understanding as quite crazy."). The Queen, of course, didn't think so. Nor did the public at the time. If M'Naghten could plan, as he could and did, then he wasn't insane. Or so thought popular opinion. Then again, neither the Queen nor the public heard first-hand the evidence presented at trial. Maybe they would have thought differently if they had. Moran challenges this conventional wisdom to the extent that his aim is "to demonstrate that [M'Naghten's] mental condition and alleged

M'Naghten provides us with a litmus test for insanity tests. M'Naghten is to insanity what Linda Brown is to equal protection. Any theory of insanity certifying M'Naghten to be sane is (for that reason) a bad theory of insanity, just as any theory of equal protection asserting Brown wasn't denied it is (for that reason) a bad theory of equal protection. Call this the "test of M'Naghten." If a test for insanity certifies Daniel M'Naghten to be insane, it passes the test; if not, it fails.

The first critical volley against the new *M'Naghten* focused, not on what the test said, but on what it didn't say. Suppose, the critics imagined, someone realized doing this or that was a crime, but as a result of mental disease or defect, just couldn't help himself. He was, as one might colloquially put it, *driven* to commit the crime, or as the law sometimes puts it, he experienced an "*irresistible impulse*" to commit it.[29] If you think such a person *should* be regarded as insane in the law's eyes, a point of considerable controversy,[30] the

---

delusions of persecution were, at the very minimum, rooted in the political reality of his day." MORAN, *supra* note 1, at 5. Richard Moran, *McNaughtan, Daniel* (1802/3-1865), OXFORD DICTIONARY OF NATIONAL BIOGRAPHY (online ed. Jan. 2008) (M'Naghten was a "former actor and medical student [who] was familiar with the symptoms of insanity, and . . . may have been feigning . . . .").

29. A standard citation is to Parsons v. State, 2 So. 854, 866 (Ala. 1887).

30. Stephen Morse is the most prominent and thoughtful critic of so-called volitional or control tests for insanity. *See, e.g.*, Stephen J. Morse, *Against Control Tests for Criminal Responsibility*, *in* CRIMINAL LAW CONVERSATIONS 449 (Paul H. Robinson et al. eds., 2009) [hereinafter Morse, *Against Control Tests*]. Richard Bonnie likewise "favor[s] narrowing the defense by eliminating its so-called volitional prong or control test." Richard J. Bonnie, *The Moral Basis of the Insanity Defense*, A.B.A. J., Feb. 1983, at 196. Michael Corrado is probably the foremost defender of an approach to insanity based *entirely* on volitional impairment. *See, e.g.*, Michael Corrado, *The Case for a Purely Volitional Insanity Defense*, 42 TEX. TECH. L. REV. 481 (2009). For a recent criticism of Corrado in favor of Morse's position, see Paul Litton, *The Mistaken Quest for a Control Test: For a Rationality Standard of Sanity*, *in* THE INSANITY DEFENSE: MULTIDISCIPLINARY VIEWS ON ITS HISTORY, TRENDS, AND CONTROVERSIES 185 (Mark D. White ed., 2017).

Morse maintains that any plausible case in which one might be inclined to believe an actor couldn't have done otherwise than commit the crime charged will typically be one in which the actor would be insane under the irrationality theory he defends. He says, for example:

new *M'Naghten* would, according to many,[31] have to be amended, and so in some places it was. The fix was to add another prong (sometimes referred to, perspicuously or not, as the "irresistible impulse" prong), to wit: a person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks capacity to conform his conduct to the requirements of law.

The new *M'Naghten*, together with the irresistible impulse addition, comprise what we'll call the traditional test for insanity. The *traditional test* has an elegant and simple structure. It takes two conditions, compelled ignorance and compelled choice, and bars criminal liability when those conditions are satisfied, provided those conditions resulted from a "mental disease or defect." Anyone who satisfies the test is said to be insane. Mental disease or defect resulting in compelled choice or compelled ignorance is the equivalent of insanity.

The traditional test doubtless has many problems, but let's focus on two. In order to set the stage for the first problem we need to say more about the concept of capacity, or incapacity, on which the traditional test depends. How do we tell if an actor lacked the capacity, or was powerless, to realize he was committing a crime, or to conform his conduct

---

> Suppose . . . that an agent's desire is so powerful and insistent that it compromises the agent's ability to think straight, to bring reason to bear on the reasons not to act. Some people in the throes of intense desire may be virtually unable to think of anything except satisfying the desire. Indeed, some addicts, for example, describe seeking and using in almost *automaton-like terms*. Their minds are blank, and seeking and using '*just happens*.' This is a textbook example of irrationality.

Morse, *supra*, at 457 (emphasis added). One doesn't want to make too much of the reference to the fact that addicts and others under the influence of powerful and insistent desire sometimes describe their experience in "automaton-like terms," but that's the language of lost agency, is it not?

31. We should note for the record that some defenders of the old *M'Naghten* Rule believed that rule, properly understood and applied, could reach and excuse those who acted on an irresistible impulse. *See, e.g.*, 2 JAMES FITZJAMES STEPHEN, HISTORY OF THE CRIMINAL LAW OF ENGLAND 167–71 (1883); JEROME HALL, GENERAL PRINCIPLES OF CRIMINAL LAW 520–22 (1960).

to the law? The usual way to test an actor's capacity is to ask a counterfactual question. If the actor *would* have realized he was committing a crime in some imagined world, then we'd say he *could* have realized he was committing a crime in this one. Likewise, if the actor *would* have conformed his conduct to the requirements of law in some imagined world, then we'd say he *could* have conformed his conduct to the requirements of law in the actual world.

If this analysis of the language of capacity is correct, then much depends on how we describe the imagined world in which we put the actor to the test. Indeed, everything depends on it. Describe that world in one way, and the accused could have chosen or believed otherwise. Describe it in another, and he couldn't have. What's the right description? The traditional test for insanity was, rightly or wrongly, commonly understood to require what was described as "total" incapacity, which would suggest that the possible world in which the accused's capacities were tested should be especially unforgiving. Rare indeed should be the case in which the requisite capacity was found lacking. In that spirit, we need to imagine some truly demanding and remote worlds.

In order to test an actor's capacity to realize he was committing a crime, let's assume a world, just like the actual world, except that a magistrate magically appears at the scene of the crime and tells him in no uncertain terms that what he's about to do is a crime. If the actor would have remained ignorant, then (and only then) would we say he lacked the capacity to realize he was committing a crime. Likewise, in order to test an actor's capacity to conform his conduct to the law (assuming he realizes he's about to commit a crime), let's assume a world, just like the actual world, except a gallows magically appears and the actor will immediately find himself hung upon it if he fails to conform. If he nonetheless would have chosen to commit it, then (and only then) would we say he lacked the capacity to conform. Those are especially unforgiving counterfactuals, but

presumably something like them is what the traditional test's supposed insistence on total incapacity would entail.

The first problem now emerges. M'Naghten was probably suffering, as we've said, from paranoid schizophrenia or paranoid psychosis. He was in pretty bad shape, believing the Tories were out to get him, haunting him day and night. If we apply the traditional test, using the counterfactuals just mentioned, then *maybe*, under the circumstances, M'Naghten was insane. *Maybe*, under the circumstances, he either couldn't have realized he was committing a crime, or if he did, couldn't have done otherwise.

Maybe, but probably not. If a judicial magistrate had miraculously and suddenly appeared before M'Naghten, telling him that shooting Drummond would offend the queen's peace, and constitute a crime, would M'Naghten have understood what he was being told? Would he then have known he was about to commit a crime? Probably. Indeed, according to some well-informed observers, M'Naghten, even without the aid of our imagined magistrate's counsel, "made no *mistakes* about what he was doing—he knew he was shooting, and he knew that he was killing—nor was he ignorant of the legal and moral prohibitions against killing."[32] If so, then M'Naghten would've struck out on the traditional test's first prong. Likewise, if the gallows had miraculously and suddenly appeared on Parliament Street as M'Naghten approached Drummond, and if M'Naghten believed he'd be immediately hung upon it if he proceeded, would he have done an about-face? Again, probably, at least according to our well-informed observers, who believe "no very persuasive case [exists] for saying that M'Naghten was *compelled* to do what he did."[33] If so, then he would've struck out on the second prong, too.

If Daniel M'Naghten is the litmus test for insanity tests, then the traditional test probably fails. Under the traditional

32. Moore, *After Breivik*, supra note 28, at 662.

33. *Id.*

test, Daniel M'Naghten would probably have been found sane and hung for his crime. Indeed, the traditional test, straightforwardly applied, probably does a pretty poor job overall sorting the intuitively sane from the intuitively insane. Worse, its likely tendency is to produce false-negatives. When all is said and done, it probably sends to prison more folks who should go to a mental hospital than it sends to mental hospitals folks who should go to prison. If anything, the traditional test allocates the risk of error in the wrong direction.

The natural response to this worry was to loosen the language, moving away from total incapacity to something less than total. The Model Penal Code (MPC) showed the way, declaring in Section 4.01(1):

> A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality of his conduct or to conform his conduct to the requirements of law.[34]

Section 4.01 preserves the basic structure of the traditional test. Vagueness was its innovation. First, it removes the word "know" and substitutes the word "appreciate," on the theory that even someone who was insane might "know" he was committing a crime, but still not "appreciate," the criminality of his conduct, where appreciation was meant to convey some deeper or broader, but ill-defined, mode of understanding.[35] Second, recognizing that capacities come in degrees,[36] the Code spurned the total incapacity thought to be required under the traditional test, providing instead that a "lack of *substantial* capacity," either

---

34. MODEL PENAL CODE § 4.01(1) (AM. LAW INST. 1985).

35. *See* MODEL PENAL CODE AND COMMENTARIES § 4.01 cmt. 3 at 169 (AM. LAW INST. 1985) ("The use of 'appreciate' rather than 'know' conveys a broader sense of understanding than simple cognition.").

36. Michael S. Moore, *The Neuroscience of Volitional Excuse*, *in* LAW AND NEUROSCIENCE 179, 188 (Dennis Patterson & Michael S. Pardo eds., 2016) [hereinafter Moore, *Neuroscience*].

cognitive or volitional, would suffice to establish the defense.[37] Demanding total incapacity, as the traditional test presumably did, was intuitively too demanding. Section 4.01 is thus the traditional test, only kinder, and more forgiving.[38]

Does Section 4.01 pass the test of M'Naghten? Would it have found M'Naghten insane? Hard to say. Some might think so; others might think not. That's the beauty of vagueness. It gives discretion, inviting reliance on intuition. Maybe that's not such a bad thing. Indeed, sometimes it's a virtue. If intuition can reliably sort the sane from the insane, what need has the law for a rule? Of course, if the law can focus and guide discretion in a way that's helpful on balance, so as to sort the cases more accurately than would reliance on intuition alone, it should. So, again, the MPC rule might or might not pass the test of M'Naghten. The MPC is softer than the traditional test, and being softer, naturally offers M'Naghten a better chance, all else being equal, to avoid the gallows based on a finding of insanity. Of course, a better chance is no guarantee. A jury instructed in the language of Section 4.01 might well find M'Naghten sane, despite the latitude Section 4.01 provides.

Even if the MPC, unlike the traditional test, passes the test of M'Naghten, it faces another problem, which the traditional test also faces. The MPC, like the traditional test, has two elements: a mental-disease-or-defect element and a liability-precluding element. The problem involves the

---

37. *See* MODEL PENAL CODE AND COMMENTARIES § 4.04 at 172 (emphasis added) (Substantial capacity means "a capacity of some appreciable magnitude when measured by the standard of humanity in general, as opposed to the reduction of capacity to the vagrant and trivial dimensions characteristic of the most severe afflictions of the mind.").

38. The story of *M'Naghten*'s evolution ends, of course, with John Hinckley's attempt on the life of then-President Reagan. Prior to Hinckley's attempt, the Model Penal Code's test was on the rise; afterward, in a reaction echoing Queen Victoria's to M'Naghten's acquittal, many jurisdictions removed any reference to the accused's capacity to conform his conduct to the requirements of law. Congress, for example, enacted 18 U.S.C. § 17 (2012), which makes no explicit reference to the accused's capacity for conformity.

relationship between them. Both tests stipulate that the mental-disease-or-defect element must *cause* a liability-precluding element. But why does the liability-precluding element preclude liability *only* when it results from a mental disease or defect? Set aside the scandalous fact that the criminal law seldom, if ever, defines "mental disease or defect." That's bad enough. What's worse is its failure to explain why the law needs it in the first place. What work does the mental-disease-or-defect requirement *do*? If someone lacked the capacity, or substantial capacity, to know (or appreciate) the law or conform to its requirements, why isn't that enough, all by itself, to defeat liability?

Maybe the mental-disease-or-defect requirement is a *proxy* for something else. Maybe it guarantees that anyone pleading insanity *qua* incapacity didn't culpably cause her incapacity. It guarantees, in other words, that the insane don't have *dirty hands*. If your incapacity arises from a mental disease or defect, then presumably you're not to blame for being so incapacitated. Mental disease just happens. No one sets out to *make* themselves insane. That's sensible, but then the mental-disease-or-defect requirement isn't really integral to insanity. The mental-disease-or-defect requirement turns out to be nothing more than a clean-hands rule.[39] Can that be right? Surely mental disorder is more tightly bound to insanity than that.[40]

Moreover, if the mental-disease requirement *is* just a

---

39. *See, e.g.*, GARY WATSON, *Excusing Addiction*, *in* AGENCY AND ANSWERABILITY: SELECTED ESSAYS 318, 333 (2004). The criminal law typically excuses someone when, as a result of *involuntary* intoxication, but *not voluntary* intoxication, he lacks the capacity to know or conform to the law. *See, e.g.*, MODEL PENAL CODE § 2.08(4). This set of rules, too, suggests that the mental-disease-or-defect requirement functions like a clean-hands proxy.

40. Perhaps anyone who lacks the capacity to know he's committing a crime, or who can't conform to the law if he does know it, must *necessarily* be suffering from something fairly called a "mental disease or defect." But that won't do, at least not for the traditional test or the MPC test. Under those tests, mental disease or defect *causes* an incapacity. If mental disease or defect is just another way to *identify* the relevant incapacity, then it can't *cause* that incapacity.

clean-hands proxy, it suffers, like any proxy, from over- and under-inclusion. First, with regard to under-inclusion, suppose Sally is faultlessly incapacitated due to mental disease or defect. When she kills Jane, insanity comes to Sally's rescue. Contrast John, who's faultlessly incapacitated due to something else. When he kills Jane, he's out of luck. No insanity defense for him. That doesn't seem fair. Why should it matter why you're incapacitated, as long as you're not at fault? Second, with regard to under-inclusion, suppose Sally is incapacitated due to mental disease or defect, but decides not to take her medicine. Without the medicine, the symptoms from which she suffers will return, and she'll be incapacitated once more. All this she knows. If Sally, now incapacitated, kills Jane, she'll still be entitled to plead insanity, at least under existing law.[41] That doesn't seem entirely fair either.

The traditional test gives us the following recipe for insanity. Step one: start with a standard liability-precluding condition, i.e., compelled ignorance or compelled choice. Step two: add a mental-disease-or-defect condition (leave undefined). Step three: require the element in step one to result from the element in step two. Voila: insanity. Maybe one lesson from the traditional test's shortcomings is to rethink the starting point. Rather than a traditional excusing condition, like compulsion or ignorance, maybe one should start instead with the idea that mental disease or defect is in itself somehow the key to insanity, that mental disease or defect can't simply be an ill-fitting proxy for clean hands.

History gives one well-known example of this approach. It comes from the Court of Appeals for the District of Columbia, in its 1954 decision in *United States v. Durham*.[42]

---

41. One possible exception to this generalization, and perhaps the only one, is Washington law, which provides that "[n]o condition of mind induced by the voluntary act of a person charged with a crime shall constitute insanity." WASH. REV. CODE § 10.77.030(3) (1998).

42. 214 F.2d. 862 (D.C. Cir. 1954).

It began with promise, but didn't end well. The court—especially *Durham*'s author, Judge David Bazelon—was dissatisfied with the traditional test. His main complaint was that it left the jury without a complete portrait of the accused's state of mind at the time of the crime, and without a complete portrait, how could the jury fairly say whether he was insane or not? *Durham* thus threw out the traditional test, and replaced it with a single-sentence alternative: "[A]n accused is not criminally responsible if his unlawful act was the product of mental disease or defect."[43]

This rule, known as *Durham* Rule, or product test, is as notable for what it doesn't say as much as for what it does. Unlike the traditional test, it makes no mention whatsoever about incapacity. Indeed, it makes no effort to identify anything that amounts to a condition precluding liability, like compelled ignorance or choice. On the contrary, what precludes liability under *Durham* is the simple fact that the actor suffered from a mental disease or defect, provided said mental disease or defect caused his unlawful act. Besides wanting a test that enabled the jury hear more about the defendant's state of mind than the traditional tests were thought to allow, why Judge Bazelon opted for the test he did is hard to know. Whatever the motivation, the *Durham* Rule itself presupposed that mental disease or defect, or at least *some* mental diseases or defects, sufficed to preclude criminal

43. *Id.* at 874–75. The *Durham* Rule doesn't explain *why* a person who wouldn't have committed a crime but for the fact that he was suffering from something called a mental disease or defect shouldn't be criminally liable. Some take the Rule to rest on what's been called a causal theory of excuse, which isn't really a theory of excuse at all. A theory of excuse presupposes some people are responsible for some choices they make, at least some of the time, and a theory of excuse sorts the excused from the unexcused. Yet the causal theory of excuse (so-called) entails that no one is responsible for any choice she makes at any time. If so, if no one is responsible for any choice she makes, including the choice to commit a crime, and insofar as punishment can permissibly be imposed only on those who are responsible for their choices, then punishment, on the causal theory, is never permissible. The causal theory of excuse leads penal abolitionism. For some this conclusion is a *reductio*; for others, the beginning of enlightenment. For the latter perspective, see, for example, DERK PEREBOOM, FREE WILL, AGENCY, AND MEANING IN LIFE 153–74 (2014).

liability, with no incapacity required, provided the mental disease or defect produced the unlawful act.

The *Durham* court may have been onto something. Alas, the court didn't see what it was, and the rot soon set in. At its source was the court's failure to define "mental disease or defect." At trial, the prosecution would present its experts, who would solemnly testify that the accused suffered from no mental disease or defect. The defense would then present its experts, who, with equal solemnity, would testify that he did, and moreover, his crime was a product of it. Who was to say who was right? The experts were, after all, the experts, at least when it came to saying what was and wasn't a mental disease or defect. The battle of the experts left the jury caught in the crossfire. Far from bringing its normative judgment to bear on the question of sanity, the jury was reduced to arbitrating credibility. Which side's experts were more believable?

When the court finally got around to defining "mental disease or defect," it missed its chance. Rather than identifying what it was about some mental diseases and defects such that suffering from them was *in itself* somehow inconsistent with criminal liability, and in some way different from the ways in which suffering from the traditional liability-precluding conditions was inconsistent with criminal liability, the court took a turn back to the traditional tests. A "mental disease of defect," the court said almost a decade after *Durham*, included "any abnormal condition of the mind which substantially affects mental or emotional processes and substantially impairs behavior controls."[44]

Ring any bells? It should. The traditional test said that a mental disease or defect had to cause some liability-precluding condition, like lacking the capacity to conform to the requirements of law. The refined *Durham* test said a

44. McDonald v. United States, 312 F.2d 847, 851 (D.C. Cir. 1962) (per curiam).

mental disease or defect had to cause "substantially impaired behavior controls," otherwise it wasn't a "mental disease or defect" at all, at least not so far as the law of insanity was concerned. When push came to shove, the D.C. Circuit thus did an about face and returned to tradition,[45] with something like compulsion—"substantially impaired behavioral controls"—coming to the fore.[46] The gravitational pull of the old ways proved hard to escape.

II. IRRATIONALITY

The D.C. Circuit eventually abandoned *Durham* and embraced the Model Penal Code in its stead. Following the acquittal of John Hinckley, on grounds of insanity, for his

45. The Court's ongoing efforts to patch up the *Durham* Rule in response to problems arising from its application eventually led the Court to abandon it altogether in favor of the then-recently promulgated MPC rule. *See* United States v. Brawner, 471 F.2d 969, 1010 (D.C. Cir. 1972). The product rule survives in New Hampshire, where it arguably first took root, and in the Virgin Islands. *See* State v. Pike, 49 N.H. 399, 438 (1869); Petric v. People, 61 V.I. 401, 408 (2014).

46. When *Durham* was finally laid to rest in *Brawner*, Judge Bazelon took another stab at it. Here's what he came up with: "[A] defendant is not responsible if at the time of his unlawful conduct his mental or emotional processes were impaired to such an extent that he cannot justly be held responsible for his act." *Brawner*, 471 F.2d at 1032 (Bazelon, C.J., concurring in part and dissenting in part). Bazelon thus went one more step (or more) beyond the MPC on the path toward vagueness. Better, perhaps, to dispense with anything like a test for insanity altogether; instead, formulate a "test" that gets all the facts in front of the jury and invites its members to decide the ultimate question of responsibility using common sense and intuition. That was more or less also the majority recommendation of the Royal Commission on Capital Punishment in 1953. *See* ROYAL COMMISSION ON CAPITAL PUNISHMENT, REPORT, 1949–1953, ¶ 333(iii), at 116 (1953) ("[A] preferable amendment of the law would be to abrogate the [*M'Naghten*] Rule and to leave the jury to determine whether at the time of the act the accused was suffering from disease of the mind (or mental deficiency) to such a degree that he ought not to be held responsible."). Getting the ultimate question to the jury, without any more by way of guidance from the law, was probably the hoped-for effect of the New Hampshire doctrine as well, which is usually, though perhaps mistakenly, taken to be equivalent to the *Durham* Rule. *See* John Reid, *Understanding the New Hampshire Doctrine of Criminal Insanity*, 69 YALE L.J. 367, 396 (1960) (claiming that the "ends sought by New Hampshire [were] . . . to discard legal presumptions and give the jury the full fact-finding duty").

attempted assassination of President Reagan, many states went back to *M'Naghten*. The tradition, in one guise or another, prevailed. Still, dissatisfaction with the traditional test lives on. At bottom, this discontent rests on the conviction that the traditional test misunderstands insanity because it sees nothing *special* about the insane. That, so say the discontented, must be wrong. The traditional test takes the oldest excuses in the book—ignorance and compulsion—and says that insanity appears when those excuses happen to result from a mental disease or defect, which is then left mysteriously undefined. That can't be right. Insanity must be something more, and something else. But what?

Enter the irrationality theory. The theory's champions naturally agree that irrationality is somehow the touchstone of insanity, but their accounts differ in detail and nuance, making it hard to tell exactly what the theory stands for. Maybe it would be better to talk about irrationality *theories*, which bear a familial resemblance to one another, but differing in important ways.[47] As usual, the devil in is the

---

47. Michael Moore and Stephen Morse are today's most prominent irrationality revisionists. *See infra* note 67. They don't agree on everything, however. For one thing, they embrace different theories of insanity as irrationality, as discussed in the text. Moore believes insanity is a status defense or exemption that, when applicable, bars liability for all of an accused's acts or choices committed while insane; Morse believes insanity is an excuse that, when applicable, excuses some acts or choices, but not all. Moreover, Moore would make room for a separate excuse based on lack of capacity to control one's choices, or conform one's choices to the requirements of law. *See* Moore, *Neuroscience*, *supra* note 36, at 179 (offering an analysis of "volitional excuse"). Morse sees no need for such an excuse, or if such a need exists, apparently believes its costs would exceed its benefits, so the law shouldn't adopt it.

For earlier discussions commonly included in the irrationality camp, see JOEL FEINBERG, *What Is So Special About Mental Illness?*, *in* DOING AND DESERVING: ESSAYS IN THE THEORY OF RESPONSIBILITY 272 (1970); Herbert Fingarette, *Insanity and Responsibility*, 15 INQUIRY 6 (1972); FINGARETTE, *supra* note 23, at 175–94; Herbert Morris, *Criminal Insanity*, 17 INQUIRY 345 (1974) (reviewing HERBERT FINGARETTE, THE MEANING OF CRIMINAL INSANITY (1972)). Robert Schopp is also commonly included. *See* ROBERT F. SCHOPP, AUTOMATISM, INSANITY, AND THE PSYCHOLOGY OF CRIMINAL RESPONSIBILITY: A PHILOSOPHICAL INQUIRY 215 (1991). Lumping all these theorists in the same camp makes sense insofar as they all stand opposed to the traditional test, and insofar as they all make some appeal

details.

What supposedly unites all irrationality theories is, once more, the proposition that insanity somehow involves irrationality. We might therefore formulate the irrationality theorist's test for insanity thusly: an actor is insane if he's irrational. Alas, if that's all the theory tells us, it doesn't tell us nearly enough.[48] M'Naghten was, let's agree, irrational in some sense, but he knew how to put two and two together. He knew how to plan an attack on the prime minister. He knew how to work a gun. Wasn't he rational, at least instrumentally? Conversely, we're all irrational more or less, more or less of the time.[49] Weakness of will, for example, is an all-too-common form of practical irrationality. The same goes for self-deception, which is an all-too-common form of epistemic irrationality. What we need to know is when irrationality veers into insanity. Without an answer, insanity risks being like pornography: we know it when we see it.

So let's try to make irrationality more precise. Let's

---

to the idea of irrationality, but it also risks obscuring important differences among them.

For example, some passages in Feinberg's important essay on insanity echo the lost-agency theory:

> '[S]enseless' desires, because they do not cohere, are likely to seem alien, not fully expressive of their owner's essential character. When a person acts to satisfy them, it is as if he were acting on somebody else's desires. And, indeed, the alien desires may have a distinct kind of unifying character all their own, *as if a new person were grafted onto the old one.*

Feinberg, *supra*, at 288 (emphasis added). Indeed, in a footnote to this passage, Feinberg writes: "Hence the point of the ancient metaphor of 'possession.'" *Id.* at 288 n.4.

48. *See* FINGARETTE, *supra* note 23, at 179 ("[T]he word 'irrational' is used in a number of very different senses."); Sinnott-Armstrong & Levy, *supra* note 25, at 317 ("[T]he term 'rational' is vague and controversial."). *See generally* Walter Sinnott-Armstrong, *Insanity vs. Irrationality*, 1 PUB. AFF. Q. 1 (1987) (criticizing irrationality theories proposed by Feinberg, Moore and Fingarette).

49. Stephen Morse emphasizes that rationality is a "continuum concept." *See, e.g.*, Stephen J. Morse, *Diminished Rationality, Diminished Responsibility*, 1 OHIO ST. J. CRIM. L. 289, 301 (2003).

distinguish two different irrationality theories: *irrationality-as-unintelligibility* and *irrationality-as-delusion*.[50] Both theories say that insanity is in some way related to *psychosis*. The unintelligibility theory says that insanity consists in being psychotic, where being psychotic is a "degree-vague assessment of severity in a variety of mental abilities" indicating "*severe* abnormalities of behavior."[51] The delusion theory says (as a first approximation) that insanity consists in psychotic choices or acts, where a psychotic choice or act is a choice or act *based on* delusion.

A. *Unintelligibility*

Irrationality is a matter of degree. According to the unintelligibility theory, the insane are just that much crazier than the rest of us. The "irrationality associated with insanity is a deep[], . . . radical kind of irrationality."[52] But "deep" and "radical" how? The unintelligibility theory portrays insanity's irrationality as deep and radical in two ways.[53] First, insanity is deep, insofar as it constitutes a

---

50. Prior to Thomas Erskine's defense of James Hadfield in 1800, existing English authority, such as it was, appears to have equated legal insanity with what was then called "total" insanity, or what's here called irrationality-as-unintelligibility. Erskine argued that "total" insanity wasn't necessary for legal insanity. What was then called "partial" insanity should, he thought, suffice. As Erskine said in *Hadfield*, in some cases "reason is not driven from her seat [total insanity], but . . . [instead] distraction sits down upon it along with her, holds her, trembling, upon it, and frightens her from propriety [partial insanity]." Trial of James Hadfield, 27 How. St. Tr. 1281, 1313 (1800). If partial insanity is equivalent to what's described here as irrationality-as-delusion, then the *M'Naghten* Rules reflected the delusion theory in the judges' answer to the fourth question put to them by the House of Lords. For some historical analyses that can be read in support of this claim, see, for example, Joel Peter Eigen, *Delusion in the Courtroom: The Role of Partial Insanity in Early Forensic Testimony*, 35 MED. HIST. 25 (1991); Richard Moran, *The Modern Foundation for the Insanity Defense: The Cases of James Hadfield (1800) and Daniel McNaughtan* (1843), 477 ANNALS AM. ACAD. POL. & SOCIAL SCI. 31 (1985).

51. Moore, *After Breivik*, *supra* note 28, at 656.

52. FINGARETTE, *supra* note 23, at 179.

53. Moore's latest statement can be found in Moore, *After Breivik*, *supra* note 28. Earlier statements can be found in MICHAEL S. MOORE, LAW AND PSYCHIATRY:

property, not of acts, but of actors. It attaches not just to what a person does. It goes deeper, attaching to the person himself. Second, the irrationality must be so radical that the person becomes unintelligible to us, so *unintelligible* that it no longer makes sense to look upon him as a moral agent at all.

When a person ceases to be an agent, it makes no sense to get angry at him for anything he does, crimes included. Resentment and indignation are misplaced, just as they are when an animal causes harm, inasmuch as animals aren't moral agents. One might say that the insane, as non-agent persons, are "beyond good and evil,"[54] or perhaps "not operating within the realm of reason at all."[55] They are, as the unintelligibility theory's foremost proponent has put it, "stranger to us than birds in our garden."[56] We should treat them with compassion, as we should any sentient creature. We might want to confine them if they're dangerous. But blaming them, being resentful or indignant toward them, makes no sense. Blame presupposes moral agency, and calling someone insane is just another way of saying she's not a moral agent at all.

The mental illness typically afflicting the unintelligible is some form of schizophrenia. Schizophrenia is commonly associated with hallucinations and delusions, but its positive symptoms also manifest in "word salad," "thought-blocking," and neologisms, as well as agitated and repetitive movements; or, at the other extreme, catatonia.[57] These are

---

RETHINKING THE RELATIONSHIP 217–45 (1984); Michael S. Moore, *Mental Illness and Responsibility*, 39 BULL. MENNINGER CLINIC 308 (1975), *reprinted in* MICHAEL S. MOORE, PLACING BLAME: A THEORY OF CRIMINAL LAW 595–609 (1997).

54. Moore, *After Breivik*, supra note 28, at 678.

55. R.A. DUFF, ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW 289 (2007). Duff suggests that "if a person is so disordered that he is not operating within the realm of reason at all, he should be described as 'a-rational' or 'a-reasonable' rather than irrational or unreasonable." *Id.*

56. Moore, *After Breivik*, supra note 28, at 678 (quoting Manfred Bleuler, described as the "'father' of schizophrenia").

57. *See* AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF

the images of the insane that the unintelligibility theory most naturally evokes. Or think about some of the more florid inmates of the Bridgewater Asylum for the Criminally Insane, depicted in Frederick Wiseman's 1967 documentary *Titicut Follies*. Other words come to mind: raving, rambling, bizarre, deranged, delirious, incoherent, and so on. Or perhaps: "[T]otally deteriorated, drooling, hopeless[ly] psychotic[]."[58] You get the basic idea. Most of us doubtless would have a hard time regarding such creatures as fellow travelers. We are, as the unintelligibility theory says, apt to look upon them as something less than human, beings so disintegrated or disordered as to be non-agents.

Insanity's history is replete with analogies. The unintelligibility theory analogizes the insane to wild beasts or brutes.[59] The analogy is usually traced to Judge Robert Tracy's jury charge in the 1724 trial of Edward Arnold, but its antecedents go back to 1256.[60] Early English law

---

MENTAL DISORDERS 99 (5th ed. 2013); U.S. DEP'T HEALTH & HUM. SERVS., NAT'L INST. MENTAL HEALTH, SCHIZOPHRENIA 3–4 (2015), https://infocenter.nimh.nih.gov/pubstatic/NIH%2015-3517/NIH%2015-3517.pdf.

58. GREGORY ZIBOORG, MIND, MEDICINE AND MAN 273 (1943).

59. Moore, *After Breivik*, *supra* note 28, at 678–79.

60. *See* Anthony Michael Platt & Bernard L. Diamond, *The Origins and Development of the "Wild Beast" Concept of Mental Illness and Its Relation to Theories of Criminal Responsibility*, 1 J. HIST. BEHAV. SCI. 355, 361 (1965) (tracing the history of the "wild beast" concept back from Tracy to Hale to Coke and finally to Bracton). According to Platt and Diamond:

> The "wild beast" concept emerged finally in the seventeenth century as a result of a mistranslation of "*brutis*" . . . Bracton's use of "brutis" was in fact shorthand for "brutis animalibus," which may literally be translated as "brute" or "dumb animals." This expression and idea was used particularly by the canonists to distinguish man from animals. Bracton correctly observed that "animals which lack reason" do not possess the will or intent to do harm; consequently, they are not to be considered legally accountable. It was quite legitimate for Bracton to include the insane in this conceptual framework because they too, like dumb animals, were considered lacking in will and understanding.

Id. at 360–61. The wild beast analogy might actually fit the lost-agency theory as well as, if not better than, the unintelligibility theory. The basic idea would be that animals in general lack a sense of themselves as agents insofar as they lack a sense of agency.

distinguished between "total insanity" and "partial insanity,"[61] and the wild beast analogy was meant to illuminate what it was to be totally insane. As Judge Tracy said: a person isn't insane unless "*totally* deprived of his understanding and memory, and doth not know what he is doing, no more . . . than a brute, or a wild beast."[62] The unintelligibility theory, which embraces the wild beast analogy,[63] thus limits insanity to total insanity. Total insanity, however, is a Procrustean theory of insanity.

Make no doubt about it. If someone qualifies as insane under the unintelligibility theory, then insane he is. That's not the problem. False positives are the problem. The unintelligibility theory is under-inclusive: it will too often put the seal of sanity on the certifiably insane. Take Daniel M'Naghten. M'Naghten was crazy, but not *so* crazy we can't make any sense of him. We might not understand why he thought the Tories were out to get him, but if killing the prime minster was the only way to stop them, then killing the prime minister was an intelligible thing to do. Nor was M'Naghten so crazy that he couldn't figure out how to make the attempt. He was able to come up with a plan and put it into action. He wasn't "totally deprived" of reason. Yet if we agree that an adequate theory of insanity should certify M'Naghten insane, the unintelligibility theory fails the test. That's one strike.[64]

---

61. *See, e.g.*, Eigen, *supra* note 50, at 27 ("Legal tracts written by Sir Matthew Hale and Lord Coke, together with judicial instructions dating to the late eighteenth century, reveal that juries were traditionally instructed that only a total want of memory and understanding—a *total* insanity—would satisfy the law's criterion for exemption from culpability.").

62. Trial of Edward Arnold, 16 How. St. Tr. 693, 765 (1724) (emphasis added).

63. Moore, *After Breivik*, *supra* note 28, at 678–79.

64. Moore maintains that M'Naghten *would* qualify as insane under the unintelligibility theory. *See* Moore, *After Breivik*, *supra* 28, at 662. Others would reasonably disagree. What the disagreement highlights, perhaps, is another point Moore makes; namely, that the "line separating persons from non-persons," the intelligible from the unintelligible, is not so much a line as a gray zone. Still, portraying M'Naghten as having passed some threshold into *unintelligibility*, as having entered a zone wherein he's "stranger to us than birds in our garden,"

Next, suppose M'Naghten had been insane under the unintelligibility theory. Suppose he *was* like a wild beast. If so, he shouldn't be held criminally liable for the attempt on Drummond's life. The problem is that, if he'd been as crazy as the unintelligibility theory requires, he probably wouldn't have committed the crime in the first place. Crimes don't require cunning, but they do usually require coherence. The non-agents of the unintelligibility theory are, one imagines, so disorganized it would be miraculous if they could pull it together enough to even be able pull a trigger.[65] Maybe some can, but not many. If so, then the unintelligibility theory is a

---

doesn't seem quite right. History's descriptions of M'Naghten, and the unintelligibility theory's description of the insane, don't mesh very well. Here is one account of how M'Naghten behaved while in prison awaiting trial:

> What would become increasingly interesting about M'Naughten was the fact that his demeanor throughout his time in prison was always precisely the same. He always appeared calm and composed. He had a hearty appetite and ate well. He appeared very attentive to conversations between other people in the jail: other prisoners, attorneys, police guards, and is said to have frequently laughed at any jocular observations that were made.

SCHNEIDER, *supra* note 1, at 29–30. On its face, that doesn't sound like someone who's "stranger to us than birds in our garden." Morse makes much the same observation more generally about people suffering with psychotic disorders. *See* Stephen J. Morse, *Moore on the Mind*, *in* LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL MOORE 240 (Kimberly Kessler Ferzan & Stephen J. Morse eds. 2016) [hereinafter Morse, *Moore on the Mind*] ("People with psychotic disorders may be perfectly capable of substantive and instrumental rationality in some, indeed most areas of their lives . . . [T]hey are not 'stranger to us than the birds in our gardens,' nor are they 'beyond good and evil' . . .. Each is recognizably one of us.") (quoting Michael Moore, *The Quest for a Responsible Responsibility Test: Norwegian Insanity Law After Breiuik*, 9 CRIM. L. & PHIL. 645, 678 (2015)).

65. FINGARETTE, *supra* note 23, at 206 ("[T]here are those whose mental powers have generally so deteriorated for one reason or another that the individual has become permanently incapable of the most elementary self-care or interpersonal intercourse."); Morse, *Moore on the Mind*, *supra* note 64, at 241 Morse, *infra* note 64, at 241 ("[T]hose people who are omni-disabled are usually too disorganized to engage in criminal conduct other than simple assaultive or disorderly conduct, for which no sensible defendant raises an insanity defense."). As Thomas Erskine put it in his defense of Hadfield, cases in which "reason is not merely disturbed, but wholly driven from her seat . . . are not only extremely rare, but never can become the subjects of judicial difficulty." Trial of James Hadfield, 27 How. St. Tr. at 1313.

test for insanity that will seldom, if ever, be used or needed. That's another strike.

Now, the unintelligibility theorist might insist that M'Naghten *was* irrational enough to be called insane. That seems unlikely, but suppose he was. Now suppose M'Naghten, as delusional as ever, decides, not to kill the prime minister, but instead to steal a bowler, not because he was irresistibly driven to it, but just because he wants to look dapper and doesn't feel like paying. He knows full well that he's stealing and stealing is a crime. His theft has nothing to do with his delusions. He doesn't, for example, believe stealing a bowler will be a blow against the Tory establishment. Still, he remains as delusional as he was on the day he shot Drummond. Is M'Naghten guilty of theft? What says the unintelligibility theory? It would be obliged to say M'Naghten is not guilty, because he was insane. Insanity attaches to persons, not to their choices or acts; or perhaps better: insanity attaches to persons, and because it attaches to persons, it attaches to *all* their choices and acts, including M'Naghten's imagined theft of the bowler. If M'Naghten was insane, then according to the unintelligibility theory, he could literally do no wrong.

That's a hard bullet to bite.[66] M'Naghten's delusions had

66. FINGARETTE, *supra* note 23, at 208 ("[T]here seems no good moral reason why, in general, a person who is persistently irrational about food should not nevertheless be held responsible in connection with his business dealings."); Morse, *Moore on the Mind*, *supra* note 64, at 241 ("[P]eople with severe mental disorders . . . may be competent or morally responsible for some conduct."); Anthony Kenny, *Can Responsibility Be Diminished?*, *in* LIABILITY AND RESPONSIBILITY 1, 24–25 (R.G. Frey & Christopher Morris eds., 1991) ("Treating madness as a status rather than a factor . . . gives a certified mental patient a license which is not given to others: he knows there are certain things he may do without being held criminally responsible, while all others not of the same status will be held responsible."). Morse and Kenny also argue that treating insanity as a status defense or exemption is also, in one way or another, stigmatizing. *See* Kenny, *supra*, at 24–25 ("Treating madness as a status . . . attaches stigma to insanity by assuming, without any need of proof, that insanity predisposes to criminal action."); Morse, *supra* at 243 (expressing concern that treating insanity as a status excuse "would contribute, albeit marginally, to common misunderstandings and fear of mental disorder that continue to stigmatize and

nothing to do with hats. A new chapeau wouldn't end his persecution. If so, would the state really be out of line if it condemned him for pinching the bowler? If not, then insanity can't be based on status, as the unintelligibility theory says. Insanity can't be like infancy. Children of a certain age might be able to do no wrong, but the same doesn't go for the insane, unless, of course, insanity is limited to those who really are unintelligible to us. Yet, if insanity is so circumscribed—if only the unintelligible are insane—then the insane will be lucky enough to survive in the world, unless someone's looking after them. They're unlikely to be going around committing crimes. The unintelligibility theory is left in a bind. If it says M'Naghten was sane, it fails the test of M'Naghten. If it says M'Naghten was insane, then M'Naghten gets a pass for *any* crime he commits. Neither horn has great appeal. Strike three.

B. *Delusion*

Let's tack to the other irrationality theory: irrationality-as-delusion.[67] For starters, let's agree that delusions are

---

exclude people with such disorders").

67. Morse has set forth his account in various places over the past twenty years. The language he uses to describe what makes an actor irrational sometimes differs slightly from one presentation to another, but so far as one can tell, nothing of substance is meant to turn on these differences. *See, e.g.*, Morse, *Moore on the Mind*, *supra* note 64; Stephen J. Morse, *Culpability and Control*, 142 U. PA. L. REV. 1587 (1994); Stephen J. Morse, *Diminished Rationality, Diminished Responsibility*, 1 OHIO ST. J. CRIM. L. 289 (2003); Stephen J. Morse, *Rationality and Responsibility*, 74 S. CAL. L. REV. 251 (2000); Stephen J. Morse & Morris B. Hoffman, *The Uneasy Entente Between Legal Insanity and Mens Rea: Beyond: Clark v. Arizona*, 97 J. CRIM. L. & CRIMINOLOGY 1071, 1124 (2008); Stephen J. Morse, *Uncontrollable Urges and Irrational People*, 88 VA. L. REV. 1025 (2002).

Morse sometimes says that an actor is insane if he lacks sufficient ability to grasp and be *guided* by good reasons. *See, e.g.*, Morse, *Against Control Tests*, *supra* note 30, at 454 ("The capacity to grasp and be *guided* by good reason is the heart of normative rationality.") (emphasis added). Referring to the ability to be *guided* by good reason has led some commentators to argue that Morse smuggles a volitional or control test into his irrationality test. *See, e.g.*, Michael S Moore, *Compatibilism(s) for Neuroscientists*, *in* LAW AND THE PHILOSOPHY OF ACTION 1,

irrational beliefs. They're irrational inasmuch as they don't respond to evidence contradicting them, no matter how rationally compelling the countervailing evidence may be. M'Naghten, for example, believed the Tories were out to get him. Time and again he was told the Tories were not out to get him, to no avail. Nothing, let's agree, would have convinced him otherwise.

Delusions come in different shapes and sizes. Some are self-contained. If you suffer from what's known as Capgras delusion, you delusionally believe your spouse is an imposter, but the delusion usually doesn't prompt violence. You don't, for example, decide to confront the imposter, or go out looking for your real spouse.[68] Other delusions don't sit idly

---

51 (Enrique Villanueva ed., 2014) ("Morse's catch basket here, 'irrationality,' includes a 'rational capacity to be guided by reasons on the particular occasion of wrongdoing, in addition to the general irrationality constitutive of true craziness. This particular incapacity returns Morse to his own can't/won't distinction, however much he wishes to step away from that distinction."); Michael Louis Corrado, *Morse on Control Tests*, *in* CRIMINAL LAW CONVERSATIONS 461, 462 (Paul H. Robinson et al., eds. 2009) ("The capacity to grasp is indeed an element of rationality; but the capacity to be guided by reasons is a control notion."). That's an understandable interpretation, but probably not the one intended. What Morse probably means is that an actor can't be guided by good reasons if he can't grasp them in the first place, and someone who's delusional can't grasp good reasons, or at least can't grasp any reason capable of dislodging his delusions.

When the dust settles, the irrationality-as-delusion theory and the judges' answer in *M'Naghten* to question four, end up, so far as one can tell, saying more or less the same thing; namely, that an accused suffering from delusions isn't criminally liable if, assuming his delusions were real, he wouldn't be guilty of the crime charged. For discussions of *M'Naghten* emphasizing the language from the answer to question four, see, for example, *Parsons v. State*, 2 So. 854, 865 (Ala. 1887) ("The rule in *McNaghten's Case* . . . is that the defense of insane delusion can be allowed to prevail in a criminal case only when the imaginary state of facts would, if real, justify or excuse the act."); Davis v. State, 28 S.W.2d 993, 996 (Tenn. 1930) ("Under [*McNaughten's Case*] . . . a homicide committed under an insane delusion is excusable, if the notion embodied in the delusion and believed to be a fact, if a fact indeed, would have excused the defendant."); Sinnott-Armstrong & Levy, *supra* note 25, at 301–02 ("[T]he judges in *M'Naghten*'s case proposed a counterfactual test" according to which an accused "'labour[ing] under . . . partial delusions only . . . must be considered in the same situation as to responsibility as if the facts with respect to which the delusion exists were real.'") (quoting M'Naghten Case (1843) 8 Eng. Rep. 718, 723).

68. Every rule has an exception. Frith and Johnstone describe "one extreme

by. Delusionally believing the Tories were out to get him, M'Naghten decided to take action. He came to believe that his only option was to assassinate Peel. M'Naghten's delusion was irrational, but once that delusion became fixed in his mind, the decision to kill Peel looks not only intelligible, but perfectly rational. M'Naghten did the cost-benefit analysis. Better to kill than (eventually) be killed. Still, at the root of his rational decision to go after Peel was the irrational delusion that Peel and his ilk were coming after him.

The delusion theory differs from the unintelligibility theory in two ways. First, irrationality is understood, not as a property of persons, but as a property of choices or acts. The unintelligibility theory exempts persons from criminal liability. The delusion theory excuses choices or acts. Second, a choice or act is irrational insofar as it rests on an irrational belief, i.e., a delusion. M'Naghten's choice to make an attempt on Peel's life was insane because the practical reasoning leading him to that choice was irrational, and the practical reasoning leading to that choice was irrational because it rested on an irrational belief: the delusion that the Tories were out to get him.[69] Thus, so far as one can tell, the delusion theory tells us that a choice is insane if a delusion has infected the practical reasoning leading to it. M'Naghten was insane because his choice to kill Peel rested on a delusion of persecution.

Alas, the traditionalist is apt to be unmoved. He'll gladly concede that M'Naghten's choice to make an attempt on Peel's life rested on delusion. He tried to kill Peel only because he irrationally believed the Tories were ruining his

---

case [in which] a [Capgras] patient who believed that his step-father had been replaced by a robot subsequently decapitated him to look for batteries and controls in his head." *See* CHRISTOPHER FRITH & EVE JOHNSTONE, SCHIZOPHRENIA: A VERY SHORT INTRODUCTION 140 (2003).

69. Morse, *Moore on the Mind*, *supra* note 64, at 240 ("We excuse [on grounds of insanity] when substantive irrationality [i.e., delusion] impairs the agent's practical reasoning in the context in question.").

life. But, the traditionalist will ask, who cares? If we assume that M'Naghten, despite his delusions, could have realized he was committing a crime, and if he could have conformed to the law and not have committed it, what difference does his delusion make? In other words, the traditionalist wants to know *why* a delusion-inspired choice, without more, should preclude liability, when *ex hypothesi* the choice itself was neither compelled nor made in compelled ignorance. For the traditionalist, what should matter is whether M'Naghten, delusional or not, had the capacity to realize killing Peel was a crime, and if so, whether he had the capacity to stop himself.

The delusion theorist might at this point reply that he's been misunderstood. Not *all* choices based on delusions are insane. Some are; some aren't. A deluded choice is insane only if, had the actor's delusions been *real*, the choice wouldn't have been a crime.[70] For example, assuming

---

70. Morse might agree with this statement of the irrationality-as-delusion theory, but one can't be entirely sure. *Compare* Morse & Hoffman, *supra* note 67, at 1129 ("Some agents . . . act for such irrational reasons that it would be unjust to blame and punish them *whether or not* they would be justified if the facts and circumstances were as they delusionally believed.") (emphasis added), *with* Morse, *Moore on the Mind*, *supra* note 64, at 243 (suggesting that a delusional actor should be convicted if he would still be guilty of a crime "even if all the facts and circumstances as he believes them to be were true"). Morse writes: "For years I waffled on this issue, but now believe that the delusional spouse must be convicted." Morse, *Moore on the Mind*, *supra* note 64, at 243. The delusional spouse is someone whose decision to kill his wife is based on "delusional jealousy," but who would still be guilty of a crime even if his delusions were true. If Morse does indeed embrace the delusion theory, that embrace would become hard to reconcile with the claim that insanity is a matter of more or less, inasmuch as irrationality is a matter of more or less. *See* Morse & Hoffman, *supra* note 67, at 1118 ("The capacity for rationality . . . is clearly a continuum concept."). The delusion theory asks if the accused would be guilty of a crime if his delusions were true. That question would appear to invite a yes-or-no answer, however much disagreement might exist over whether yes or no is the right answer.

For those familiar with the criminal-law literature, this statement of the delusion theory will sound familiar. It sounds like Christopher Slobogin's "integrationist" alternative to any affirmative test for insanity. Slobogin proposes (roughly) abolishing any affirmative defense of insanity and instead applying existing criminal-law doctrines to the facts as the defendant believed them to be, whether his beliefs are delusional or not. *See* Christopher Slobogin, *An End to*

M'Naghten's delusions about the Tory spies were real, his decision, say, to steal a bowler, would still be a crime. Stealing the bowler would do nothing to end the persecution, and M'Naghten would therefore be guilty of larceny, despite his delusions. In contrast, eliminating Peel was, M'Naghten delusionally believed, the only way he could end his torment. If we assume that killing Peel really was the only way M'Naghten could end his torment, would he be guilty of murder? That, according to the delusion theory, is the critical question.

What's the answer? It depends. Let's take self-defense off the table. M'Naghten attacked Drummond, not the other way round. M'Naghten never claimed Drummond was an imminent threat to his life or limb.[71] M'Naghten might nonetheless claim that Peel's death, compared to his own continued suffering, was the lesser evil. In other words, M'Naghten might try to justify his attempt on Peel's life in the name of necessity. Would that have worked under the law of necessity circa 1843? Probably not, if only because

---

Insanity: Recasting the Role of Mental Disability in Criminal Cases, 86 VA. L. REV. 1199, 1199–1208 (2000) [hereinafter Slobogin, *An End to Insanity*]. Alas, "subjectivizing" the criminal law to the extent the integrationist test requires would necessitate substantial reform of existing doctrine. *See, e.g.*, Morse & Hoffman, *supra* note 67, at 1128 (stating that Slobogin is "incorrect about the extent of subjectivization current law accepts"); Christopher Slobogin, *A Defense of the Integrationist Test as a Replacement for the Special Defense of Insanity*, 42 TEX. TECH. L. REV. 523, 532–33 (2009) (noting that if the law isn't sufficiently "subjective," then a "'special defense' for offenders with mental illness would still be needed, albeit one that focuses solely on whether, at the time of the crime, they lacked subjectified *mens rea* or believed they were confronted by circumstances that would be justifying or coercive if true") [hereinafter Slobogin, *A Defense*]. As Slobogin notes, his integrationist theory is "virtually identical" to the "Partial Delusion defense announced . . . in M'Naghten" in response to question four. Christopher Slobogin, *The Integrationist Alternative to the Insanity Defense: Reflections on the Exculpatory Scope of Mental Illness in the Wake of the Andrea Yates Trial*, 30 AM. J. CRIM. L. 315, 335 (2003).

71. *See* Slobogin, *An End to Insanity*, *supra* note 70, at 1203 ("Whether M'Naghten would have been acquitted under the . . . [integrationist] approach would depend on whether he believed the harassment would soon lead to his death or serious bodily harm and whether he thought there was any other way to prevent that occurrence.").

common-law necessity didn't extend to murder.[72] Nor would it have worked today, since English law still doesn't recognize necessity as a defense to murder.[73] Much the same would go for common-law duress, which likewise didn't extend to murder. M'Naghten might have had better luck in a different time or place. The Model Penal Code, for example, contemplates necessity and duress, under the right circumstances, as defenses to homicide, though most jurisdictions don't follow the Code on that score. If a jury were asked to decide if M'Naghten, assuming his delusion were true, chose the lesser evil, or killed under duress, who knows what it would do.[74]

The delusion theory rests its verdict, sane or insane, on the law applied to the accused's delusional world. It requires stepping into the actor's crazy world and applying the law to the facts as they exist in that crazy world. Take the case of James Hadfield of England, circa 1800.[75] Hadfield believed his death was necessary in order to bring forth the second coming of Christ. He also believed that his death had to be at someone's hand other than his own, else the second coming would not come. So Hadfield decided he had to kill King George III, which would cause the King's successor to

---

72. One can get picky about this. According to one contemporary treatise on English law, "[u]nder current law, *Dudley [and Stephens]* is regarded as authority that necessity is unavailable for murder. But, on the facts, *Dudley* need not have decided any specific rule for murder." SIMESTER ET AL., *supra* note 22, § 21.3(ii)(e), at 805. *Dudley and Stephens* is, of course, the famous English cannibalism case, but it wasn't decided until 1884, over forty years after *M'Naghten*.

73. *See id.*

74. One might think any claim of lesser evils or duress would founder because M'Naghten, even within his delusional world, had other options to protect himself from the Tories short of killing the prime minister. Why, for instance, didn't he go to the police and tell them what the Tories were doing to him? One reply, of course, is that he did, and was told nothing could be done. Perhaps from within his delusional world, M'Naghten believed the state had abandoned him. An analogy might be to a battered spouse who kills her sleeping abuser believing that the state, after her repeated but failed efforts to get help from the authorities, will not protect her.

75. Trial of James Hadfield, 27 How. St. Tr. 1281 (1800).

execute him for treason, and thus would Christ come again. Was Hadfield, under the delusion theory, sane or not?

According to the delusion theory, it depends: is it a crime to kill the king if killing the king is necessary to bring Christ into the world? Well, once again, necessity probably wasn't at the time a defense to homicide, let alone when the target is the King. Then again, it is Christ we're talking about.[76] Is it permissible to sacrifice the King (in order to sacrifice oneself in order) to bring about the Second Coming? It probably depends on whether you believe in Christ in the first place, but set that complication aside. Isn't the real problem the question itself? The delusion theory throws up the question, but the question itself seems a little crazy. The criminal law is created for our world, not delusional ones.[77] All else being equal, a test that avoids crazy questions is better than one that turns on them.

Time to switch gears. Each of the tests for insanity discussed above—the traditional test, the MPC, *Durham*, and the irrationality twins—are tests *for the accused*, which he can either pass and be labeled insane, or fail and be

---

76. James Fitzjames Stephen (a judge at the time) had this to say about the verdict in *Hadfield's Case*: "My own opinion . . . is that, if a special Divine order were given to a man to commit murder, I should certainly hang him for it unless I got a special Divine order not to hang him. What the effect of getting such an order would be is a question difficult for any one to answer till he gets it." 2 STEPHEN, *supra* note 31, at 160 n.1.

77. Slobogin tries to address this objection in Slobogin, *A Defense*, *supra* note 70, at 539–42. Judge Ladd, in *State v. Jones*, 50 N.H. 369 (1871), wasn't impressed with the delusion theory. Referring to the answer given to question four in *M'Naghten*, he wrote:

> The doctrine thus promulgated as law has found its way into the text books, and has doubtless been largely received as the enunciation of a sound legal principle since that day. Yet it is probable that no ingenuous student of the law ever read it for the first time without being shocked by its exquisite inhumanity. It practically holds a man confessed to be insane, accountable for the exercise of the same reason, judgment, and controlling power, that is required of a man in perfect health. It is, in effect, saying to the jury, the prisoner was mad when he committed the act, but he did not use sufficient reason in his madness.

Id. at 387–88.

labeled sane. The traditional test and the MPC, for example, ask if the accused had the power, more or less, to know or appreciate he was committing a crime, or whether he had the power, more or less, to do otherwise. *Durham* asks if the accused's mental disease or defect caused him to commit the crime. The irrationality tests ask if the accused was so irrational he was no longer an agent at all, or if his crime would not have been criminal if his delusions had been true. For all their differences, each of these tests shares one thing in common. They each presuppose that the accused was the one in control of his choices at the time of the crime. They each presuppose that the accused was the author of his actions. What, though, if he wasn't? That brings us to insanity as lost agency.

III. LOST AGENCY

When you move your hand you experience your hand moving, and you experience yourself moving your hand. Your hand is moving, and you're the one moving it. In other words, when you move your hand you experience a sense of *ownership* (it's *my* hand) and a sense of *agency* (it's *me* moving it).[78] These experiences then lead you to believe that the moving hand is yours, and that the author or agent of its movements is you. Our experiences of ownership and agency are commonplace. We take them for granted. So much so that they fade into the phenomenal background. We hardly ever

78. Research on the sense of agency is a relatively new development. *See* Tim Bayne & Elisabeth Pacherie, *Narrators and Comparators: The Architecture of Agentive Self-Awareness*, 159 SYNTHESE 475, 475 (2007) ("Until recently, neither philosophers nor psychologists had much interest in the awareness of one's own agency. This is no longer the case, and there is now a burgeoning literature on the mechanisms underlying 'the sense of agency.'"). For recent collections, see, for example, AGENCY AND SELF-AWARENESS: ISSUES IN PHILOSOPHY AND PSYCHOLOGY (Johannes Roessler & Naomi Eilan eds., 2003); DECOMPOSING THE WILL (Andy Clark, Julian Kiverstein & Tillmann Vierkant eds., 2013); NEUROPSYCHOLOGY OF THE SENSE OF AGENCY: FROM CONSCIOUSNESS TO ACTION (Michela Balconi ed., 2010); SENSE OF AGENCY: EXAMINING AWARENESS OF THE ACTING SELF (Nicole David, James W. Moore & Sukhvinder Obhi eds., 2015); THE SENSE OF AGENCY (Patrick Haggard & Baruch Eitam eds., 2015).

notice them. We take them for granted. Simple, right?

Not always. Take Dr. Strangelove. The character in Stanley Kubrick's 1964 film of the same name was an odd duck, but perhaps the oddest thing about him was that hand. It seemed to have a mind and will of its own. The good doctor's affliction has a name: alien hand syndrome.[79] Those, like the doctor, who suffer from this syndrome experience and believe the moving hand is theirs. What they don't experience or believe is that they're the one moving it. They experience a sense of ownership without a sense of agency. Their hand is moving, but they're not moving it. Someone, or something, external or alien to them, to their sense of self, is in control. Many "refer to their alien hands as 'imps' or 'devils.'"[80]

Most of us are lucky. We don't suffer from alien hand syndrome. But most of us have driven a car, and driving a car, or any other over-learned behavior, can provide a glimpse of what it's like to lose one's sense of agency, though doubtless well short of the real McCoy.[81] We sometimes go on auto-pilot. We drive, but have no sense of driving. We might talk about "highway hypnosis." Psychologists might talk

---

79. To be more precise, loss of a sense of agency *without* loss of a sense of ownership is called anarchic hand syndrome. *See, e.g.*, Elisabeth Pacherie, *The Anarchic Hand Syndrome and Utilization Behavior: A Window onto Agentive Self-Awareness*, 22 FUNCTIONAL NEUROLOGY 211 (2007). Loss of a sense of agency *and* ownership is called alien hand syndrome. *See* Clelia Marchetti & Sergio Della Sala, *Disentangling the Alien and Anarchic Hand*, 3 COGNITIVE NEUROPSYCHIATRY 191 (1998).

80. SAM KEAN, THE TALE OF DUELING NEUROSURGEONS 257 (2014).

81. *See* Tim Bayne & Neil Levy, *The Feeling of Doing: Deconstructing the Phenomenology of Agency*, *in* DISORDERS OF VOLITION 49, 56 (Natalie Sebanz & Wolfgang Prinz eds., 2006) ("Although it seems to be true that experiences of authorship are recessive or dampened in the context of automatic actions (or, at least, actions which we experience as automatic), we doubt that the phenomenology of authorship is entirely lacking from such experiences."); Patrick Haggard, *Conscious Intention and the Sense of Agency*, *in* DISORDERS OF VOLITION 69 (Natalie Sebanz & Wolfgang Prinz eds., 2006); Matti Vuorre & Janet Metcalfe, *The Relation Between the Sense of Agency and the Experience of Flow*, 43 CONSCIOUSNESS & COGNITION 133 (2016).

about automaticity.[82] Some musicians and athletes might describe it as "flow," or being "in the zone."[83] They lose a sense of being in control of what they're doing. A virtuoso might say, in order to try to capture the experience, that the "music just took over." Of course, zoning out is a far cry from alien hands. For one thing, the flow stops if something interrupts it, and once interrupted, the actor regains the sense of agency.[84] You zone back in. Focus won't restore the sense of agency to an anarchic hand, however.

Insanity, on the account offered below, is alien hand syndrome writ large. What makes an act insane is the fact that actor has no sense of agency at the time he performs it. He experiences the action as the action of his body, but fails to experience himself as its author. He doesn't experience himself as in control of what he's doing, because he doesn't experience himself as doing it. Imagine Dr. Strangelove's hand hits you. You'd be upset, but you wouldn't (once you calm down) blame *him*. *He* didn't hit you. His hand did. Likewise, if insane Jane kills John, Jane's not to blame. Jane didn't kill John. Jane's body did. Blame and censure presuppose responsibility, responsibility presupposes agency, and agency presupposes a sense of agency. Jane lacked agency, because she lacked a sense of agency, on the theory of insanity proposed here.

This theory needs a name. Call it the *lost-agency theory*.[85] Of course, like most things having to do with the

---

82. John A. Bargh & Tanya L. Chartrand, *The Unbearable Automaticity of Being*, 54 AM. PSYCHOLOGIST 462, 468 (1999).

83. *See, e.g.*, LÁSZLÓ HARMAT ET AL., FLOW EXPERIENCE: EMPIRICAL RESEARCH AND APPLICATIONS (2016); Vuorre & Metcalfe, *supra* note 81.

84. Some psychologists might describe this phenomenon as shifting from what they call System 2 (the automatic system) back to System 1 (the deliberative system), where the shift itself is under the command of System 2. Something happens to interrupt the flow, at which point System 2 relinquishes control back to System 1. On System 1 and System 2 generally, see DANIEL KAHNEMAN, THINKING, FAST AND SLOW 19–30 (2011).

85. John Deigh hints at the lost-agency theory in his overlooked essay, *Moral Agency and Criminal Insanity*, *in* EMOTIONS, VALUES, AND THE LAW 196, 207–10

mind, one's sense of agency can doubtless be diminished or reduced, as opposed to being lost altogether.[86] For ease of

(2008) (An insane action "does not issue from the actor's exercise of his agency. He is not its author."). Deigh draws on the work of Harry Frankfurt, who's famously associated with an account of moral responsibility consisting in a mesh or harmony between what Frankfurt calls one's higher-order volitions and one's effective first-order desires. Basically, your choices are free (and you're therefore responsible for them) if your first-order desires, which move you to act, mesh with your higher-order volitions; your choices are unfree (and you're therefore not responsible for them) if your high-order volitions and first-order desires don't mesh. Moreover, if they don't mesh, that means you qua your higher-order volitions are *alienated* from you qua your first-order desires. It goes without saying that the sense of alienation involved in anarchic hand syndrome and lost agency is of a far different and more profound sort than the sense Frankfurt had in mind.

Compare the lost-agency theory to theories that locate insanity in a broader character theory of excuse, according to which (roughly) insanity excuses because, when someone who's insane commits a crime, his act is, as a result of mental disease or defect, "out of character." *See* LAWRIE REZNEK, EVIL OR ILL? JUSTIFYING THE INSANITY DEFENCE 237 (1997) ("When illness makes a person act out of character, we are inclined to excuse him on the basis of a change in moral character."). Character theories of excuse, found more in the criminal-law literature than in the philosophical literature, have long faced stern challenges. Applied to insanity, for example, one might ask about the case in which the person has *long* been insane, such that whatever wrong he does, far from being *out* of character, is actually *in* character. Should he be excused? The character theory of insanity would presumably say no, but that seems counter-intuitive. Unlike the character theory, which says insanity excuses because crimes committed while the accused was insane are out of character for the accused, the lost-agency theory says insanity bars liability because the accused didn't commit the crime in the first place.

86. Besides being a matter of more or less, lost agency can probably come and go. *See* Matthis Synofzik et al., *The Experience of Agency: An Interplay Between Prediction and Postdiction*, 4 FRONTIERS IN PSYCHOL. 1, 6 (2013) ("Agency attribution in patients with delusions of influence [or control] . . . fails only episodically and only in certain contexts.") [hereinafter Synofzik et al., *Experience of Agency*].

Criminal law theorists within the irrationality camp debate whether insanity is an excuse or an exemption (or status excuse). *See infra* note 47 and accompanying text. Excuses are pleas that attach to acts. Exemptions are pleas that attach to actors. If insanity is an excuse, it defeats liability for particular actions. If insanity is an exemption, it defeats liability for any action the actor performs over whatever period of time he's exempt, i.e., during whatever period of time he occupies the responsibility-defeating status. The lost-agency theory elides this dichotomy. On the one hand, when an actor's sense of agency is lost, he's not on the hook for anything he does within that interval. In that sense, insanity looks

exposition, however, let's introduce the theory on the assumption that the sense of agency is all-or-nothing.[87]

A. *The Demon Within*

Metaphors are a risky business. They can illuminate, but they can also mislead. Having put that disclaimer on the table, here's a metaphor for lost-agency.

Call it the *alien-self metaphor*. Before an actor loses his sense of agency the only self around is the actor himself. Call that self the *real self*. When the sense of agency is lost, another self emerges. Call that self the *alien self*. Now we have two selves, so to speak, in the same body. Before the split the real self was in charge. Once the sense of agency is lost the real self loses command. The alien self takes over. The real self becomes a passive bystander to the actions of a body he continues to experiences as his own, but no longer

---

like an exemption. Assuming lost agency isn't usually, if ever, a permanent condition, that it can come and go, one needs to ask whether it was present or absent at the time of the crime. In that sense, insanity looks like an excuse.

87. The neuropsychological process by which the sense of agency arises has produced an enormous literature. The dominant model is known as the comparator model. *See, e.g.*, Sarah-Jayne Blakemore et al., *Abnormalities in the Awareness of Action*, 6 TRENDS COGNITIVE SCI. 237 (2002); Christopher D. Frith et al., *Abnormalities in the Awareness and Control of Action*, 355 PHIL. TRANS. ROYAL SOC'Y LONDON B: BIOLOGICAL SCI. 1771 (2000); Christopher D. Frith, *Explaining Delusions of Control: The Comparator Model 20 Years On*, 21 CONSCIOUSNESS & COGNITION 52 (2012). For more or less accessible introductions, see ANIL ANANTHASWAMY, THE MAN WHO WASN'T THERE: INVESTIGATIONS INTO THE STRANGE NEW SCIENCE OF THE SELF 112–15 (2015); Markus Schlosser, *Agency*, STAN. ENCYCLOPEDIA PHIL. 29–30 (2015). For recent criticisms of the comparator model, and attempts to upgrade it, see, for example, Matthis Synofzik, *Comparators and Weightings: Neurocognitive Accounts of Agency*, *in* THE SENSE OF AGENCY, *supra* note 78, at 289; Bayne & Pacherie, *supra* note 78; Matthis Synofzik et al., *Beyond the Comparator Model: A Multifactorial Two-Step Account of Agency*, 17 CONSCIOUSNESS & COGNITION 219 (2008) [hereinafter Synofzik et al., *Two-Step Account*]; Synofzik et al., *Experience of Agency*, *supra* note 86. Of course, telling when someone has acted without a sense of agency isn't easy. Studies of the phenomenon usually rely on self-reporting or something called "intentional binding." *See, e.g.*, James W. Moore & Sukhvinder S. Obhi, *Intentional Binding and the Sense of Agency: A Review*, 21 CONSCIOUSNESS & COGNITION 546 (2012). The former is vulnerable to fabrication; the latter can only be tested in a laboratory.

experiences himself as the author of what his body does. His actions are thus at once both his and not his. His body is acting. Intentions are being formed and volitions are being executed to produce bodily movement, but he doesn't experience himself as the author of any of it.

The alien-self metaphor has antecedents in the history of insanity. Once upon a time the insane were, as we've seen, compared to wild beasts and children. Once upon a time as well, however, they were compared to something else: those who were *possessed*.[88] An insane person was likened to someone demonically possessed, someone whose body an alien self had overtaken: someone with a demon within.[89]

---

88. *See, e.g.*, ROY PORTER, MADNESS: A BRIEF HISTORY 10–33 (2002) (chapter entitled "Gods and demons"); ROY PORTER, MIND-FORG'D MANACLES: A HISTORY OF MADNESS IN ENGLAND FROM THE RESTORATION TO THE REGENCY 63 (1987) ("One mark of possession by the Devil was insanity."); DANIEL N. ROBINSON, WILD BEASTS & IDLE HUMOURS: THE INSANITY DEFENSE FROM ANTIQUITY TO THE PRESENT 74–112 (1996) (chapter entitled "Possession & Witchcraft"); ANDREW SCULL, MADNESS IN CIVILIZATION: A CULTURAL HISTORY OF INSANITY FROM THE BIBLE TO FREUD, FROM THE MADHOUSE TO MODERN MEDICINE 67–69 (2015) (section entitled "Demonic Possession and Spiritual Healing"). For a recent historical account of possession and exorcism in sixteenth and seventeenth century Europe, see BRIAN P. LEVACK, THE DEVIL WITHIN: POSSESSION & EXORCISM IN THE CHRISTIAN WEST (2013).

89. Consider Andrea Yates, who drowned her children to save their souls. According to one account: "Mrs. Yates believed that *Satan was within her* and tormented her and the children. She thought after she drowned her children, she would be arrested and executed. She indicated that Satan would be executed along with her." Phillip J. Resnick, *The Andrea Yates Case: Insanity on Trial*, 55 CLEV. ST. L. REV. 147, 149 (2007) (emphasis added). Yates, like James Hadfield over 150 years earlier, killed wanting, or at least believing, she would be executed as a result. Yates killed to kill Satan; Hadfield, to summon Christ.

Or consider Andrew Goldstein who, in January 1999, "picked . . . up [Kendra Webdale], and threw her onto the tracks in front of . . . [a New York City subway] train just as it entered the station." CHARLES PATRICK EWING, INSANITY: MURDER, MADNESS, AND THE LAW 116 (2008). Goldstein described his experience in a videotaped statement to the police:

> As I was standing on the platform there was a woman standing waiting for the train. She was facing the incoming train and I was standing behind her. I got the urge to push, kick or punch. I pushed the woman who had blond hair. I don't recall what she looked like. But I know she was a white female. When I pushed her she fell onto the track and was struck by the train . . .

The lost-agency theory gives new meaning to this crazy old idea. "The devil made me do it" is a lame excuse,[90] but when it comes to insanity it may contain a kernel, or perhaps more than a kernel, of truth. In fact, possession may have been behind the old wild-beast analogy, inasmuch as "animals and the insane were both considered embodiments of evil spirits."[91]

When the alien self takes charge, the real self is still on-board, but helpless. The alien self responds to reasons as it

---

> I feel like an aura, or a sensation like you're losing control of your motor system. And then, you lose control of your sense and everything. *And then you feel like something's entering you. Like you're being inhabited.* I don't know. But—and then, and then it's like an overwhelming urge to strike out or push or punch . . .

Id. (emphasis added). Thanks to Joshua Kleinfeld for the reference to Goldstein's remarks.

What about John Hinkley? How would he have fared under the lost-agency theory? Without knowing all the facts, offering a prediction is, of course, hazardous. Having said that, the conventional wisdom, so far as one can tell, was then, and is now, that Hinckley was sane, thus the adverse reaction to his acquittal on grounds of insanity. In the popular imagination, Hinckley is thought to have suffered from erotomania, also known as de Clerambault's syndrome, i.e., he delusionally believed Jodie Foster was in love with him, which somehow prompted his assassination attempt. In fact, Hinckley realized Foster was *not* in love with him. *See* RICHARD J. BONNIE ET AL., A CASE STUDY IN THE INSANITY DEFENSE: THE TRIAL OF JOHN W. HINCKLEY, JR. 41–42 (3d ed. 2008). Though not suffering from the delusions associated with erotomania, Hinckley may nonetheless have been delusional. The experts who testified at trial disagreed. The defense psychiatrists testified that Hinckley was psychotic, suffering in particular from delusions of reference; the Government's psychiatrists testified that Hinckley was non-psychotic. *See* BONNIE ET AL., *supra*, at 31. ("Perhaps the most significant divergence of clinical opinion about Hinckley's mental disorder at the time of the offense pertained to whether or not Hinckley had 'delusions' and 'ideas of reference.'"). If the Government's witnesses were right, then one red flag indicating lost-agency would have been missing in Hinckley's case. That's about as much as one can honestly say looking at the case from afar.

90. *See* Moore, *Neuroscience*, *supra* note 36, at 201.

91. Platt & Diamond, *supra* note 60, at 363 ("Catholic dogma distinguished human beings from animals, assigning to Man the possibility of an immortal soul. At the same time, animals were considered appropriate objects of punishment and excommunication. One means of reconciling these views, endowing animals with intelligence and souls without contradicting Christian dogmas, was to assume that they were incarnations of evil spirits.").

goes about its business, which might include the business of crime. It can plan and premeditate, and indeed, its plans can be quite intricate. It can look out for the police. It can lock doors, flee from danger, and so forth. It can also be impulsive, and act with what appears to be utter disregard for the consequences. The alien self has a mind and will of its own. It might respond to reasons the real self would never respond to, and fail to respond to reasons to which the real self would consistently respond. The alien self's reasons don't integrate or cohere with the real self's reasons, making it natural to describe the insane as dis-integrated or in-coherent.

Lost agency isn't the traditional control, or irresistible impulse, test of insanity in disguise. It's easy to confuse the two. Someone claiming to be insane under the traditional test might well say, "I couldn't control myself." Someone claiming to be insane under the lost-agency theory might well say the same thing. The similarity is superficial, however. The sense in which control is lost under the traditional test, and the sense in which it's lost under the lost-agency theory, are very different. Under the traditional theory the question is whether the actor could have chosen otherwise than he did. Under the lost-agency theory, in contrast, the question is whether the actor at the time of choice experienced himself as the one doing the choosing. "I wasn't in control," is a plea of a different order from "I couldn't control myself." The former pleads lost-agency; the latter, compulsion.

The real self can react to its lost sense of agency in different ways, and different actors can describe the experience of lost agency in different language. Faced with the "vague and strange experience"[92] of having one's body move without moving it, one possibility is simple bewilderment. The real self knows his body performed an

92. Synofzik et al., *Experience of Agency*, *supra* note 86, at 5; Synofzik et al., *Two-Step Account*, *supra* note 87, at 228 (Lost agency is experienced as "strange, peculiar and not fully done by me.").

action, but having failed to experience himself as its author, might well have no idea why he did what he did, i.e, why his body did what it did. Or he might report having acted for one reason or another but with no experience of those reasons as his reasons. He might describe himself as an onlooker or detached observer. If asked why he did what he did, he might reply: "It wasn't me." "I don't know what happened." "It just happened." "I don't know why I did it." Or he might be reduced to (apparent) paradox: "I did it, but I didn't do it," as in "I (my body) did it, but I (as agent) didn't."

Or he might go a step further, into delusion.[93] Faced with

93. The literature on delusion, neglected by criminal-law theorists, is very large. For helpful overviews, see, for example, Lisa Bortolotti, *Delusion*, STAN. ENCYCLOPEDIA PHIL. (2013); Lisa Bortolotti & Kengo Miyazono, *Recent Work on the Nature and Development of Delusions*, 10 PHIL. COMPASS 636 (2015); Max Coltheart et al., *Delusional Belief*, 62 ANN. REV. PSYCHOL. 271 (2011) [hereinafter Coltheart et al., *Delusional Belief*]; *Forum—Phenomenological and Neurocognitive Perpsectives on Delusions*, 14 WORLD PSYCHIATRY 164 (2015). For book-length treatments, see, for example, LISA BORTOLOTTI, DELUSIONS AND OTHER IRRATIONAL BELIEFS (2010); PHILIP GERRANS, THE MEASURE OF MADNESS: PHILOSOPHY OF MIND, COGNITIVE NEUROSCIENCE, AND DELUSIONAL THOUGHT (2014); JENNIFER RADDEN, ON DELUSION (2011); LAWRIE REZNEK, DELUSIONS AND THE MADNESS OF THE MASSES (2010).

The literature tends to organize around four questions. First: Are delusions beliefs or something else? For defenses of the traditional doxastic account against critics, see, for example, Tim Bayne & Elisabeth Pacherie, *In Defence of the Doxastic Conception of Delusions*, 20 MIND & LANGUAGE 163 (2005); Lisa Bortolotti, *In Defence of Modest Doxasticism About Delusions*, 5 NEUROETHICS 39 (2011). Second: Is the formation and maintenance of a delusion best explained along the lines of the so-called two-factor model or the one-factor (prediction-error) model? *Compare, e.g*, Martin Davies et al., *Monothematic Delusions: Towards a Two-Factor Account*, 8 PHIL., PSYCHIATRY & PSYCHOL. 133 (2002) (two-factor model), *and* Oren Griffiths et al., *Delusions and Prediction Error: Re-Examining the Behavioural Evidence for Disrupted Error Signaling in Delusion Formation*, 19 COGNITIVE NEUROPSYCHIATRY 439 (2014) (same), *with* Philip R. Corlett et al., *Toward a Neurobiology of Delusions*, 92 PROGRESS IN NEUROBIOLOGY 345 (2010) (one-factor theory based on prediction error). Third: Can the two models be reconciled? Kengo Miyazono et al., *Prediction-Error and Two-Factor Theories of Delusion Formation: Competitors or Allies?*, *in* ABERRANT BELIEFS AND REASONING 34 (Niall Galbraith ed., 2015). Fourth: Can the two-factor-model account for polythematic, as well as monothematic, delusions? *See, e.g.*, Max Coltheart, *On the Distinction Between Monothematic and Polythematic Delusions*, 28 MIND & LANGUAGE 103, 110–11 (2013) (suggesting possible ways in which the two-factor theory can explain polythematic delusions).

the experience of his body moving but not being the author of its movement, he might entertain the proposition that he was—must have been—under the control or influence of some alien or external force. Moreover, having entertained the proposition that someone or something external to his real self has taken charge of what he does, he might come to believe it. Indeed, he might hold onto that belief no matter what the evidence to the contrary. He might, in other words, form and maintain what are known as delusions of alien control or influence.

According to a 29-year-old shorthand typist diagnosed with schizophrenia:

> When I reach my hand for the comb it is my hand and arm which move, and my fingers pick up the pen, but I don't control them. . . . I am just a puppet who is manipulated by cosmic strings. When the strings are pulled my body moves and I can't prevent it.[94]

In the mind of someone so afflicted, the metaphor of alien control will no longer be just a metaphor. It will become reality. Someone or something external to the real self has taken over: a demon, a device, God, and so forth.

The sense of agency can be lost over the mind as well as the body.[95] One might experience a thought in one's head but have no experience of having put it there. The thought is instead experienced as having been inserted into one's consciousness. This experience might in turn give rise to its own particular brand of delusion, commonly known as "thought insertion."[96] The actor comes to believe that

---

94. ANDREW C.P. SIMS, SYMPTOMS IN THE MIND: AN INTRODUCTION TO DESCRIPTIVE PSYCHOPATHOLOGY 153 (2d ed. 1995).

95. *See, e.g.*, Joelle Proust, *Is there a Sense of Agency for Thought?*, *in* MENTAL ACTIONS 253 (Lucy O'Brien & Matthew Soteriou eds., 2009).

96. The literature on thought insertion, so far as one can tell, focuses on two questions. First: Does thought insertion involve a lost sense of agency or a lost sense of ownership? *Compare* G. LYNN STEPHENS & GEORGE GRAHAM, WHEN SELF-CONSCIOUSNESS BREAKS: ALIEN VOICES AND INSERTED THOUGHTS (2000) (agency), *and* Patrizia Pedrini, *Rescuing the "Loss-of-Agency" Account of Thought Insertion*, 22 PHIL., PSYCHIATRY & PSYCHOL. 221 (2016) (same), *and* Shaun Gallagher,

someone or something is putting thoughts in his head. According to another 29-year-old, also diagnosed with schizophrenia:

> I look out the window and think the garden looks nice and the grass cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts, only his . . . . He treats my mind like a screen and flashes his thoughts into it like you flash a picture.[97]

One might likewise experience a voice in one's head but have no experience of being its author. The actor loses his sense of agency over his own "inner speech," attributing it to some external source or entity. This phenomenon too has a name: "auditory verbal hallucination."[98] Thought insertion

---

Relations Between Agency and Ownership in the Case of Schizophrenic Thought Insertion and Delusions of Control, 6 REV. PHIL. PSYCHOL. 865 (2015) (same), *with* Lisa Bortolotti & Matthew Broome, *A Role for Ownership and Authorship in the Analysis of Thought Insertion*, 8 PHENOMENOLOGY & COGNITIVE SCI. 205 (2009) (ownership), *and* Michelle Maiese, *Thought Insertion as a Disownership Symptom*, 14 PHENOMENOLOGY & COGNITIVE SCI. 911 (2015) (same), *and* Jean-Remy Martin & Eliszbeth Pacherie, *Out of Nowwhere: Thought Insertion, Ownership and Context Integration*, 22 CONSCIOUSNESS & COGNITION 111 (2013) (same).

Second: Does the same mechanism that explains lost agency over action explain lost agency over thoughts? *Compare, e.g.*, Philip Gerrans, *The Feeling of Thinking: Sense of Agency in Delusions of Thought Insertion*, 2 PSYCHOL. CONSCIOUSNESS: THEORY, RES., & PRAC. 291 (2015) (yes), *with* Agustin Vincete, *The Comparator Account on Thought Insertion, Alien Voices and Inner Speech: Some Open Questions*, 13 PHENOMENOLOGY & COGNITIVE SCI. 335 (2014) (raising questions).

97. SIMS, *supra* note 94, at 152.

98. For a sampling of the literature on auditory verbal hallucinations, see STEPHENS & GRAHAM, *supra* note 96; Raymond Cho & Wayne Wu, *Mechanisms of Auditory Verbal Hallucination in Schizophrenia*, 4 FRONTIERS IN PSYCHIATRY 1 (2013); Remko van Lutterveld et al., *The Neurophysiology of Auditory Hallucinations—A Historical and Contemporary Review*, 2 FRONTIERS IN PSYCHIATRY 1 (2011); Lauren Swiney & Paulo Sousa, *A New Comparator Account of Auditory Verbal Hallucinations: How Motor Prediction Can Plausibly Contribute to the Sense of Agency for Inner Speech*, 8 FRONTIERS IN HUM. NEUROSCI. 72 (2014); Rachel Upthegrove et al., *Understanding Auditory Verbal Hallucinations: A Systematic Review of Current Evidence*, 133 ACTA PSYCHIATRICA SCANDINAVICA 352 (2016); Sam Wilkinson & Ben Alderson-Day, *Voices and Thoughts in Psychosis: An Introduction*, 7 REV. PHIL. & PSYCH. 529, 531 (2016) ("[T]he orthodox view of AVHs [is that] they are to be understood as the result of disrupted monitoring of inner speech."). For thoughts on how the law

and auditory verbal hallucinations can each be grounded in a lost sense of agency, not over the movements of one's body, but over the movements of one's mind.

One form of auditory verbal hallucination, so-called command hallucinations, have attracted special attention from the law, at least when God is doing the commanding. According to the "deific decree" doctrine, if God tells you to kill, you're insane.[99] However the doctrine might relate to the traditional test,[100] the fact that an accused heard the voice of God commanding him to commit the crime charged is some evidence that the real self wasn't in charge. Hearing God's command might mean the actor has come to see himself as an instrument of God's hand, and seeing oneself as an instrument is one way to try to make sense of lost agency. Of course, insofar as the deific doctrine extends only to those who hear God's voice, it sweeps too narrowly. It shouldn't matter who's doing the commanding. First Amendment worries aside,[101] what matters is that the experience of being commanded is good evidence of what really matters: to wit, the experience of not being the agent of the acts adding up to

---

can distinguish true from malingered auditory verbal hallucinations, see Simon McCarthy-Jones & Phillip J. Resnick, *Listening to Voices: The Use of Phenomenology to Differentiate Malingered from Genuine Auditory Verbal Hallucinations*, 37 INT'L J.L. & PSYCHIATRY 183 (2014).

99. For a nice overview of the doctrine's history, see Lundgren v. Mitchell, 440 F.3d 754, 784–87 (6th Cir. 2006) (Merritt, J., dissenting).

100. *See, e.g.*, JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 25.04[C][1][a][iv], at 348 (7th ed. 2015) (discussing "deific decree" doctrine).

101. Wilson v. Gaetz, 608 F.3d 347, 354 (7th Cir. 2010) ("[T]o distinguish between 'deific' and all other delusions and confine the insanity defense to the former would present serious questions under the First Amendment[.]") (Posner, J.); Grant H. Morris & Ansar Haroun, *"God Told Me to Kill": Religion or Delusion?*, 38 SAN DIEGO L. REV. 973, 992, 996 (2001). For commentary on the relationship between command hallucinations and violence, see, for example, Louise G. Braham et al., *Acting on Command Hallucinations and Dangerous Behavior: A Critique of the Major Findings of the Last Decade*, 24 CLINICAL PSYCHOL. REV. 513, 522–26 (2004); Dale E. McNiel et al., *The Relationship Between Command Hallucinations and Violence*, 51 PSYCHIATRIC SERVICES 1288 (2000); Abraham Rudnick, *Relation Between Command Hallucinations and Dangerous Behavior*, 27 J. ACAD. PSYCHIATRY & L. 253 (1999).

the crime.

Of course, not all delusions are delusions of alien control and inserted thoughts. Delusions come in many shapes and sizes. Some are monothematic, involving "a single delusional belief or a small set of delusional beliefs that are all related to a single theme."[102] Others, of the variety typically seen in schizophrenia, for example, are polythematic, involving "delusional beliefs about a variety of topics unrelated to each other." Some are downright bizarre.[103] The real John Nash, for example, believed he would become the Emperor of Antarctica, not to mention being the left foot of God on Earth.[104] Other delusions don't reach such bizarre heights. Some are self-contained, leaving other beliefs untouched. Others metastasize, infecting the whole mind, such that other beliefs form around the delusion. Some of these beliefs are themselves delusional (secondary delusions); others seem like rational responses to the irrationality of the primary delusion. Delusions can compose a strange and motley crew.

---

102. Coltheart et al., *Delusional Belief*, *supra* note 93, at 280, 283–88. Monothematic delusions include the Capgras delusion, Cotard delusion, Fregoli delusion, mirrored-self misidentification, somatoparaphrenia, delusion of alien control, and delusion of thought insertion.

103. How might one explain the bizarre content of some delusions? According to one account:

> Bizarre delusions . . . represent inappropriate use of metaphor in an attempt to establish some inter-subjective meaning, albeit futile. During the formative delusional mood, the world becomes ineffable. Prodromal patients use relative terms (similes) to describe their experiences: "It is *as if* people are actors, walking down the street wearing masks." As these experiences persist, the relative terms subside (people *are* wearing masks, they *are* in disguise); the simile becomes a metaphor as the delusion develops and the metaphor becomes a top-down prior around which perception and cognition are organized.

Philip R. Corlett, *Answering Some Phenomenal Challenges to the Prediction Error Model of Delusions*, 14 WORLD PSYCHIATRY 181, 182 (2015).

104. *See* Donald Capps, *John Nash's Delusional Decade: A Case of Paranoid Schizophrenia*, 52 PASTORAL PSYCHOL. 193, 200–01 (2004). The Nash character was also depicted as having visual hallucinations, but these are very rare. The real Nash's hallucinations were only auditory.

Delusions of alien control and thought insertion reflect under-attribution of agency. When the actor ought to experience himself as agent, he doesn't. Delusions can also reflect the opposite error: over-attribution. When the actor ought *not* to experience himself as an agent, he does. So-called delusions of reference (or megalomania) are an example.[105] Such delusions "involve beliefs that unrelated or commonplace phenomena in the world (events, objects, or other people) refer directly to oneself and carry a special personal significance."[106] The character John Nash in Ron Howard's *A Beautiful Mind* suffered from delusions of reference, searching papers, magazines and whatnot, looking for secret messages and codes only he could decipher. Although delusions of alien control and delusions of reference might seem at opposite ends of the spectrum, delusions of reference, like delusions of alien control, might likewise originate in the experience of lost agency.

The basic idea is simple. A lost sense of agency arises when the internal mechanism by which we gain a sense of agency fails to work as it should.[107] Delusions of alien control arise when the actor tries to make sense of that experience, and some further defect prevents the actor from rejecting the hypothesis that some alien force is in control of his thoughts and actions.[108] Another response, when one's internal mechanism for self-monitoring fails, is to rely on external

---

105. For more on delusions of reference, see Mike Startup et al., *Delusions of Reference: A New Theoretical Model*, 14 COGNITIVE NEUROPSYCHIATRY 110 (2009); Mike Startup & Sue Startup, *On Two Kinds of Delusions of Reference*, 137 PSYCHIATRY RES. 87 (2005).

106. Coltheart et al., *Delusional Belief*, *supra* note 93, at 277.

107. *See id.* at 288.

108. This presupposes a so-called bottom-up account, according to which a delusion arises from an abnormal experience. So-called top-down accounts put the causal relation in the opposite direction, with abnormal experience arising from the delusion. One challenge top-down theories face, of course, is the need to explain where the delusion comes from in the first place. Bottom-up accounts claim that the abnormal experience, which gives rise to the delusion, is itself a result of some defect in how the brain works. *See* Bortolotti, *Delusion*, *supra* note 93, § 3.2, at 19–21.

cues in an effort to compensate.[109] These external cues then take on out-sized significance and meaning. They assume aberrant salience. In order to make sense of it all, the actor assigns them special and personal importance. Delusions of reference result: mundane objects "speak" to the actor, for example.

Delusions of reference commonly go hand-in-hand with delusions of persecution.[110] As with delusions of reference, the link between lost agency and persecutory delusions isn't obvious or direct. Still, a link might well exist.[111] For

---

109. ANANTHASWAMY, *supra* note 87, at 117 (noting that schizophrenic patients "have to rely more heavily on their judgments about the external environment to augment their sense of agency[,]" because their internal mechanism for self-monitoring is defective); Bayne & Pacherie, *supra* note 78, at 486; Shitij Kapur, *Psychosis as a State of Aberrant Salience: A Framework for Linking Biology, Phenomenology, and Pharmacology in Schizophrenia*, 16 AM. J. PSYCHIATRY 13, 15 (2003); Synofzik et al., *Experience of Agency*, *supra* note 86, at 6 ("[A]s a consequence of giving up the usually most robust and reliable internal action information source, i.e., internal predictions, the sense of agency in psychotic patients is at constant risk of being misled by *ad-hoc* events, invading beliefs, and confusing emotions and evaluations."); Matthis Synofzik et al., *Misattributions of Agency in Schizophrenia Are Based on Imprecise Predictions About the Sensory Consequences of One's Actions*, 133 BRAIN 262, 270 (2010) (Schizophrenic "patients might over-attribute external events to their own agency whenever stronger weighted external agency cues are in fact not veridical and misleading."); Martin Voss et al., *Altered Awareness of Action in Schizophrenia: A Specific Deficit in Predicting Action Consequences*, 133 BRAIN 3104, 3110 (2010).

110. DANIEL FREEMAN & PHILIPPA GARETY, PARANOIA: THE PSYCHOLOGY OF PERSECUTORY DELUSIONS 122 (2004) ("Internal sensations of significance and reference may lead to delusions of reference that are understood within a persecutory belief system."); Coltheart et al., *Delusional Belief*, *supra* note 93, at 277 ("Delusions of persecution and reference commonly co-occur (often along with hallucinations)."); Startup & Startup, *supra* note 105, at 112 ("[I]t appears that the traditional association between referential delusions and persecutory delusions applies primarily to referential delusions of observation, [not communication.]").

111. FREEMAN & GARETY, *supra* note 110, at 117 (developing a model of persecutory delusion in which "[i]nternal anomalous experiences are important," including "thoughts being experienced as voices; actions being experienced as unintended; more subtle cognitive alternations such as perceptual anomalies; depersonalization; or a sense of significance or reference"); Daniel Freeman & Philippa Garety, *Advances in Understanding and Treating Persecutory Delusions: A Review*, 49 SOC. PSYCHIATRY & PSYCHIATRIC EPIDEMIOLOGY 1179,

example, when an actor loses a sense of agency over his thoughts and inner speech, let alone his body, anxiety is a natural response. He might come to believe that the thoughts he experiences, and the words he hears in his head, aren't just external: they might be part of a malevolent conspiracy.[112] Compared to seeing oneself as someone's or something's puppet, it might be easier to see oneself as the object of persecution.[113] Persecutory delusions typically draw content from the actor's pre-existing stock of beliefs.[114] M'Naghten's belief that the Tories were after him may, for example, have been rooted in what, according one author, was his "hatred of the Tories and the policies they represented."[115]

We've come a long way. We started with the experience of lost-agency, and delusions of alien control and thought insertion to which that experience can give rise. Those aren't uncommon delusions to find in cases where the defendant looks like an exemplar of insanity. Other delusions, also not uncommon in folks who look, at least intuitively, like they're

---

1182 (2014); Daniel Freeman, *Suspicious Minds: The Psychology of Persecutory Delusions*, 27 CLINICAL PSYCHOL. REV. 425, 432 (2007); Robyn Langdon et al., *The Cognitive Neuropsychological Understanding of Persecutory Delusions*, *in* PERSECUTORY DELUSIONS: ASSESSMENT, THEORY, AND TREATMENT 221, 229 (Daniel Freeman et al., 2008) ("[N]europsychological impairments . . . might precipitate a train of thought leading (more or less directly) to a persecutory delusion."); Jennifer Radden, *Defining Persecutory Paranoia*, *in* RECONCEIVING SCHIZOPHRENIA 255 (Man Cheung Chun et al. eds., 2007).

112. *See, e.g.*, Paul C. Fletcher & Chris D. Frith, *Perceiving is Believing: A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia*, 10 NATURE REVS. NEUROSCIENCE 48, 56 (2009) ("Ultimately, someone with schizophrenia will need to develop a set of beliefs that must account for a great deal of strange and sometimes contradictory data. Very commonly they come to believe that they are being persecuted: delusions of persecution are one of the most striking and common of the positive symptoms of schizophrenia . . .").

113. FREEMAN & GARETY, *supra* note 110, at 120 ("Believing that something is wrong with them (for instance, that they are becoming mad) may be a more distressing belief (and less plausible and compelling) than that they are being persecuted.").

114. *Id.* at 119 ("In the search for meaning, pre-existing beliefs about the self, others, and the world will be drawn upon.").

115. MORAN, *supra* note 1, at 45.

insane, such as delusions of reference and persecution, can likewise arise from lost-agency, or so it seems. If so, then insanity doesn't always come with a demon within. Lost agency can manifest in delusions the alien-self metaphor fails to capture. Nonetheless, even in these cases, the real self isn't at the controls. Alien self or not, what matters in the end is the lost sense of agency—of not being in control, of being a passive onlooker—not the content of the particular delusions, if any, it generates.

Not all delusions, we should emphasize, spring from the experience of lost agency. Some delusions are rooted in other abnormal experiences. For example, someone with Capgras delusion recognizes a familiar face, but doesn't experience the affective reactions that usually go along with it. That missing affective experience, together with some further defect in the way the actor tests hypotheses, produces a delusion: the person with the familiar face (one's mother, for example) must really be a robot or an imposter.[116] When someone commits a crime in the grip of delusion, red flags should go up. With delusion comes at least the possibility, and sometimes probability, that the real self wasn't the self in charge at the time of the crime. Delusion or no, if the choice to commit a crime was made when the sense of agency was gone, the real culprit wasn't the real self. Whether or not the actor delusionally believed an alien self was in command, the real self wasn't.

The traditional test portrays insanity as consisting in a

116. *See, e.g.*, Neralie Wise, *The Capgras Delusion: An Integrated Approach*, 15 PHENOMENOLOGY & COGNITIVE SCI. 183 (2016) (offering an account of the Capgras delusion that draws on what the author calls the phenomenological and analytic traditions). The Cotard delusion is the belief that one is dead, or that parts of one's body are rotting. "[T]he experiential states underlying the Capgras and Cotard delusions are different: whereas the Capgras delusion appears to involve a fairly focal impairment in face processing, the Cotard delusion seems to involve a global alteration in affective experience. Rather than experiencing only familiar faces as alien, the Cotard patient experiences *everything* as strange, devoid of meaning and lifeless." Tim Bayne, *Delusions*, *in* THE OXFORD COMPANION TO CONSCIOUSNESS 218, 220 (Tim Bayne et al. eds., 2009).

traditional excusing condition, i.e., compelled ignorance or choice, provided it results from a "mental disease or defect" (left undefined). The irrationality theory, in each of its forms, portrays insanity as consisting, more or less, in one particular feature characteristic of different mental disorders, namely, psychosis (i.e., delusions and hallucinations).[117] In contrast, the aptly-named lost-agency theory portrays insanity as consisting in the loss of one's sense of agency. Delusions and hallucinations are a common upshot of lost agency, but they're not coextensive with it. If responsibility presupposes agency, and if agency presupposes a sense of agency, then the insane, lacking a sense of agency, aren't responsible.

Schizophrenia is probably first among equals when it comes to the mental disorders most commonly associated with insanity. On the lost-agency theory, that comes as no surprise, for the "delusion of alien control is particularly associated with schizophrenia."[118] Indeed, lost agency appears capable of underwriting a number of delusions characteristic of schizophrenia.[119] Yet not everyone we might intuitively regard as insane is schizophrenic, or at least we shouldn't assume that to be true. What, for example, about those afflicted with manias, philias and phobias, i.e.,

---

117. The unintelligibility variant equates insanity with severe psychosis. The delusion variant equates it with psychosis of whatever severity, provided the crime the accused committed wouldn't have been a crime had the actor's delusions been real.

118. Coltheart et al., *Delusional Belief*, *supra* note 93, at 288; *id.* at 278 ("[C]ontrol delusions . . . are considered to be more specifically characteristic of schizophrenia.").

119. FRITH & JOHNSTONE, *supra* note 68, at 141 ("Many of the delusions reported by patients with schizophrenia seem to result from a combination of an abnormal experience with a willingness to develop extremely unlikely explanations for that experience."); JOELLE PROUST, THE PHILOSOPHY OF METACOGNITION: MENTAL AGENCY AND SELF-AWARENESS 243–64 (2013) (discussing sense of agency in schizophrenia); Philip Gerrans, *Passivity Experience in Schizophrenia*, *in* DISTURBED CONSCIOUSNESS: NEW ESSAYS ON PSYCHOPATHOLOGY AND THEORIES OF CONSCIOUSNESS 325 (Rocco J. Gennaro ed., 2015).

disorders whose stock-in-trade is intense, unremitting, and unwanted desire? When an actor commits a crime in the grip of a desire associated with these disorders, who commits the crime? An alien self or the real self?

If we attend to the way those suffering from such afflictions sometimes describe their experiences, it might be tempting to believe, for example, that the kleptomaniac didn't choose to set the fire, because his desire to set it *bypassed* his will.[120] What, though, could that possibly mean? How can desires move the body without the will's help? When the doctor taps your patellar ligament your knee goes up without your will, but desires don't move the body the way reflexes do. Desires cause intentions; intentions cause volitions; and volitions cause the body to move. That's how desires move bodies. Desires don't move bodies all by themselves. They require the will to form intentions and volitions in their service, since volitions are in the end what make the body move, at least when its motion isn't due to mechanical reflex. If so, then the idea of desires bypassing the will is incoherent, or at best metaphorical.

Any sense of alienation involved in disorders of desire is probably, at least in most cases, an alienation of a more common and pedestrian kind, compared to the lost sense of agency necessary for insanity. Rather than the alienation of lost agency, disorders of desire more likely reflect the alienation arising from desires we wish we didn't have. Still, we can't completely eliminate the possibility that the metaphor of the bypassed will isn't, at least sometimes, more than just a metaphor.[121] Sometimes, perhaps, an actor might

120. Moore, *Neuroscience, supra* note 36, at 195–97 (discussing the idea of desires "bypassing" the will). When Joel Feinberg talks about the desires associated with kleptomania he describes them as "'senseless' . . . because they do not cohere, are likely to seem alien, not fully expressive of their owner's essential character. When a person acts to satisfy them, *it is as if he were acting on somebody else's desires.*" FEINBERG, *supra* note 47, at 288 (emphasis added).

121. Bayne & Levy, *supra* note 81, at 52 ("It is sometimes suggested that one of the pathological features of the phenomenology of addiction and obsessive-compulsive spectrum disorders is that the individuals concerned experience their

not experience the relevant desire as his desire, nor the intention arising from that desire as his intention, nor the volition arising from that intention as his volition. If so, then talk of bypassed wills becomes another way of saying the actor's sense of agency was lost.[122]

B. *Defects of Consciousness*

The traditional test of insanity invites us to see insanity as a defect of reason (cognition) or will (volition). The lost-agency theory puts insanity in a different light. It's neither (just) a defect of reason, nor will. Instead, it's a defect of consciousness: a lost sense of agency. So understood, insanity doesn't stand alone. It belongs in the same camp as other defects of consciousness. Three such defects have captured the criminal law's imagination: hypnosis, somnambulism (sleepwalking), and multiple personality disorder (now known as dissociative personality disorder).[123] These phenomena are the exotica of the criminal law: interesting to think about but rarely seen in the wild of the real world. Criminal lawyers usually see them as belonging to one species, and insanity as an entirely different animal. The

---

actions as caused by their desires and urges rather than as having their source in *them*. We suspect that as one begins to experience one's movements as caused by one's mental states, one no longer experiences them as one's own actions.").

122. What about psychopaths? Are they insane under the lost-agency theory? Students of the criminal law, along with ethicists, have spent a lot of time wondering if psychopaths are capable of bearing moral responsibility for the crimes they commit. The idea is that if they can't be morally responsible, then they can't be criminally responsible either. One couldn't possibly do justice here to the enormous literature dealing with psychopaths and psychopathy. *See, e.g.*, BEING AMORAL: PSYCHOPATHY AND MORAL INCAPACITY (Thomas Schramme ed., 2014); HANDBOOK ON PSYCHOPATHY AND LAW (Kent A. Kiehl & Walter P. Sinnott-Armstrong eds., 2013); RESPONSIBILITY AND PSYCHOPATHY: INTERFACING LAW, PSYCHIATRY, AND PHILOSOPHY (Luca Malatesti & John McMillan eds., 2010). Suffice it to say that, so far as one can tell, psychopathy doesn't appear to involve any claim of lost agency.

123. Epileptic fugue, associated with petit mal seizures, is another example. For a fascinating discussion of such cases in nineteenth century England, see JOEL PETER EIGEN, UNCONSCIOUS CRIME: MENTAL ABSENCE AND CRIMINAL RESPONSIBILITY IN VICTORIAN LONDON (2003).

lost-agency theory suggests we should instead see them as members of the same species, belonging in the same doctrinal camp.

1. Hypnosis

In Robert Wiene's 1920 film *The Cabinet of Dr. Caligari*, hypnotism leads a character named Cesare (under the control of Dr. Caligari) to commit a string of murders.[124] In the 1959 novel *The Manchurian Candidate,* post-hypnotic suggestion leads Raymond Shaw to attempted assassination. That's film and fiction.[125] Not the real world. Hypnosis today is more about stage entertainment and smoking cessation. One searches in vain for real crimes committed under hypnosis.[126] Maybe that's because hypnosis can push a person only so far: no one will do under hypnosis something he wouldn't do on his own.[127] Still, the prospect of hypnotic

---

124. *See, e.g.*, Bernard Williams, *The Actus Reus of Dr. Caligari*, 142 U. PA. L. REV. 1661, 1670 (1994).

125. STEFAN ANDRIOPOULOS, POSSESSED: HYPNOTIC CRIMES, CORPORATE FICTION, AND THE INVENTION OF CINEMA (2008); Deirdre Barrett, *Hypnosis in Film and Television*, 49 AM. J. CLIN. HYPNOSIS 13 (2006).

126. *See* Michael Heap, *Hypnosis in the Courts*, *in* THE OXFORD HANDBOOK OF HYPNOSIS 745–66 (Michael R. Nash & Amanda J. Barnier eds., 2008); Graham F. Wagstaff, *Hypnosis and the Law: Examining the Stereotypes*, 35 CRIM. JUST. & BEHAV. 1277, 1279 (2008). Newspapers, including the *New York Times*, reported in 1895 that Thomas McDonald, acquitted of murder, had claimed at trial to have been under the hypnotic control of Anderson Gray. *See, e.g.*, *The Hypnotist Made Principal: His Subject Found Guiltless of Murder by a Kansas Court While He Bears the Penalty for the Crime*, N.Y. TIMES, Apr. 7, 1895. Just a few days later, however, the *Times* ran another story, reporting "that the defense of hypnotism was not raised at the trial; that no evidence concerning hypnotism was given, and that the word 'hypnotism' was mentioned but once, in a remark of McDonald's counsel, after Gray, the first man tried, had been convicted." *Hypnotism Not a Factor: Gray Made McDonald a Murderer Merely by Argument*, N.Y. TIMES, Apr. 15, 1895. In any event, McDonald was acquitted, apparently on grounds of self-defense. *See id.* Gray was convicted and hung. *See* State v. Gray, 39 P. 1050, 1054 (Kan. 1895) (affirming conviction).

127. *See, e.g.*, MODEL PENAL CODE AND COMMENTARIES § 2.01 cmt. 2 at 221 (AM. LAW INST. 1985) ("The widely held view that the hypnotized subject will not follow suggestions repugnant to him was deemed insufficient to warrant treating his conduct while hypnotized as voluntary; his dependency and helplessness are too pronounced.").

crime has become, for whatever reason, part of the legal imagination.

The experts disagree about what really happens when someone enters a hypnotic trance.[128] According to one school of thought, what happens isn't much. Hypnotized subjects just get more relaxed and willing to follow orders. They play-act. If so, if nothing special happens when someone is hypnotized, then a hypnotic crime is no different than any other crime, and neither are the culprits who commit them. The other school makes hypnosis more interesting. Hypnosis is a dissociated state in which the hypnotist takes control of the subject's mind and actions.[129] The subject becomes "mesmerized."[130] If so, then hypnotic crimes aren't just like the rest. Hypnotized villains are innocent pawns. Their manipulators bear all the guilt.

Go back to the metaphor of the alien self. If we take the dissociation account as true, the family resemblance between hypnotism and insanity becomes easy to see. The insane and the hypnotized both experience a lost sense of agency. Indeed, what hypnotists call the "classic suggestion effect" just is a lost sense of agency.[131] This lost sense of agency

---

128. *Compare* Erik Z. Woody & Pamela Sadler, *Dissociation Theories of Hypnosis*, *in* THE OXFORD HANDBOOK OF HYPNOSIS: THEORY, RESEARCH, AND PRACTICE 81 (Michael R. Nash & Amanda J. Barnier eds., 2008) (discussing state theories), *with* Steven Jay Lynn et al., *Social Cognitive Theories of Hypnosis*, *in* THE OXFORD HANDBOOK OF HYPNOSIS: THEORY, RESEARCH, AND PRACTICE 111 (Michael R. Nash & Amanda J. Barnier eds., 2008) (discussing non-state theories).

129. Vince Polito et al., *Sense of Agency Across Contexts: Insights from Schizophrenia and Hypnosis*, 2 PSYCHOL. OF CONSCIOUSNESS: THEORY, RES., & PRAC. 301, 309–10 (2015) ("[T]he 'dissociated control' and 'dissociated experience' theories of hypnotic responding are conceptually very similar to the comparator model account of passivity experiences in schizophrenia.").

130. So named after Franz Anton Mesmer. *See* ALAN GAULD, A HISTORY OF HYPNOTISM (1992).

131. David A. Oakley & Peter W. Halligan, *Hypnotic Suggestion: Opportunities for Cognitive Neuroscience*, 14 NATURE REVS.: NEUROSCIENCE 565, 568 (2013) ("[A]n effect is considered a 'classical suggestion-effect' only if it is experienced as involuntary; as 'happening all by itself.'"); Vince Polito et al., *Measuring Agency Change Across the Domain of Hypnosis*, 1 PSYCHOL. OF CONSCIOUSNESS: THEORY,

divides the self. The real self becomes disassociated from the alien self, and the alien self takes control.[132] The hypnotized real self, like the insane real self, becomes a passive observer to what his mind and body do under the alien self's control. Indeed, those under hypnosis sometimes attribute what they do to alien control, just like the insane sometimes do.[133]

Still, the phenomenology of hypnosis differs from insanity in at least three ways. First, the loss of agency in insanity is associated with mental disease or defect; the loss of agency in hypnosis is associated with, because it results from, hypnotic induction. Second, the alien self of insanity controls the real self, but no one in turn controls the alien self, whereas the alien self of hypnosis is under the hypnotist's control. The hypnotist is the puppeteer, and the alien self the puppet. The real self is a helpless bystander.[134] Finally, when the actor is told to snap out of it, the lost sense of agency ends, the alien self disappears, and the self re-integrates. Reintegration doesn't come about so easily for the insane.

Again, hypnotic crimes, or at least alleged hypnotic

RES., & PRAC. 3, 3 (2014) ("[E]xperiencing . . . actions in hypnosis as occurring without effort or conscious volition. . . . has been considered an essential element of hypnotic responding.").

132. *See* Woody & Sadler, *supra* note 128, at 89 ("[A]cross [the] diverse matrix of hypnotic behavior there is an essential denominator: in hypnosis all these behaviors are accompanied by the subjective experience that the self is not the origin of the response."); *id.* at 92 (suggesting that the dissociation arising from hypnosis involves a "breakdown" in the same mechanism that produces a loss of the sense of agency in "psychotic disorders, such as schizophrenia."); *id.* at 94 ("[D]issociation theories [of hypnosis] hypothesize that for highly hypnotizable people, hypnosis transiently brings about a disruption of [the] mechanism . . . for discriminating the internal versus external origins of events.").

133. Michael H. Connors, *Hypnosis and Belief: A Review of Hypnotic Delusions*, 36 CONSCIOUSNESS & COGNITION 27, 37–38 (2015); Rochelle E. Cox et al., *An Hypnotic Analogue of Alien Control: Modeling the Delusion and Testing Its Impact on Behavior and Self Monitoring*, 1 PSYCHOL. OF CONSCIOUSNESS: THEORY, RES., & PRAC. 407, 425 (2014).

134. Hypnotic actions are at the crossroads between lost agency and manipulation, i.e., the hypnotist induces a lost sense of agency and manipulates the actions of the emergent alien self.

crimes, are, at best, offbeat events. If and when they ever occur, most students of the criminal law, at least inasmuch as the dissociation theory accurately depicts the hypnotic subject's state of mind, would agree that criminal liability would be out of bounds. The disagreement is *why*. According to some, the actor hasn't acted. He hasn't voluntarily moved his body. Others find this implausible. How can it be that the actor hasn't acted when his body performed whatever complex action it takes to commit the crime? Volitions must have formed, which means the resulting movements were voluntary. If so, the reason for letting the hypnotic criminal go must be something other than the alleged fact that he didn't act.

Both sides have a point. Indeed, they might be talking past one another. The no-act camp looks at the real self, rightly observing that the real self didn't act. The real self was just a bystander, albeit a bystander to the voluntary movement of his own body.[135] The act camp looks at the alien self, rightly observing that the alien self did act, albeit at the hypnotist's behest. Mainly at stake in this dispute is the burden of proof. If the no-act camp wins, the state bears the burden. If the act camp wins, the defense bears it. Yet rather than make the burden question turn on who you look at—real self or alien self—the burden question should be answered directly. Are the reasons for assigning the burden to one side or the other better served if the state must prove the accused was in command, or if the defense must prove he wasn't?

2. Sleepwalking

Unlike hypnotic crimes, which exist only in fiction, people do actually commit crimes while asleep.[136] Mrs.

---

135. *See* Tim Bayne, *The Sense of Agency*, *in* THE SENSES: CLASSIC AND CONTEMPORARY PERSPECTIVES 368 (Fiona Macpherson ed., 2014) ("Many people are reluctant to regard physical or mental happenings that are unaccompanied by a basic 'experience of doing' as actions.").

136. Sleepwalking is caused by partial arousal from stage 3–4 NREM sleep.

Cogdon laid out her 19-year-old daughter Pat's pajamas, put a hot-water bottle in her bed, and a glass of milk on her bedstand.[137] Then she went to bed. She later "left her bed, fetched an axe from the woodheap . . . and struck two accurate forceful blows on [Pat's] head with the blade of the axe, thus killing her."[138] The year was 1950. The place was Australia, and the Korean War was waging not far away. Mrs. Cogdon had told Pat her worries about the war before going to sleep. When she awoke after the killing, Mrs. Cogdon reported having "dreamt that 'the war was all around the house,' that soldiers were in Pat's room, and that one soldier was on the bed attacking Pat."[139] She recalled nothing else. Tried for murder, she was acquitted.

Then there was Ken Parks. Parks had a gambling problem, which had strained his marriage. He got no sleep on May 22, 1987. On the following day he went to sleep around 1:30 a.m. with much on his mind. The next thing Parks remembered was "looking down at his mother-in-law's face. Her mouth and eyes were open and she had a

---

Sleepwalking "episodes typically take place during the first third of the night when slow-wave [or deep] sleep is predominant." Antonio Zadra et al., *Somnambulism: Clinical Aspects and Pathophysiological Hypotheses*, 12 THE LANCET: NEUROLOGY 285, 285 (2013). "Sleepwalking" or "somnabulism" is distinguished from what's called "RBD," for "REM Sleep Behavior Disorder. *See* Naoko Tachibana, *REM Sleep Behavior Disorder*, 6 SLEEP MED. CLINICS 459, 459 (2011). RBD is a "unique parasomnia characterized by dream enactment behavior during REM sleep." Ronald B. Postuma et al., *REM Sleep Behavior Disorder: From Dreams to Neurodegeneration*, 46 NEUROBIOLOGY OF DISEASE 553, 553 (2012). Cases involving sleepwalking can also involve the use of sleep aids, or alcohol. *See* Christopher Daley et al., *"I Did* What*?" Zolpidem and the Courts*, 39 J. AM. ACAD. PSYCHIATRY L. 535, 535 (2011); Shreeya Popat & William Winslade, *While You Were Sleeping: Science and Neurobiology of Sleep Disorders and the Enigma of Legal Responsibility of Violence During Parasomnia*, 8 NEUROETHICS 203, 207–09 (2015).

137. *See* Norval Morris, *Somnamulistic Homicide: Ghosts, Spiders, and North Koreans*, 5 RES JUDICATAE 29, 29 (1951). *Fain v. Commonwealth*, 78 Ky. 183, 183–85 (Ct. App. 1879), is also frequently cited.

138. Morris, *supra* note 137, at 30.

139. *Id.*

'frightened "help-me" look.'"[140] After falling asleep, Parks had driven from his home to that of his in-law's, fourteen miles away. He stabbed his mother-in-law to death, and strangled his father-in-law unconscious. He then got back into his car and drove to a police station. He remembered bits and pieces of what happened, but only bits and pieces. The jury acquitted Parks, believing he was indeed asleep throughout it all.[141]

Sleepwalking, like hypnosis, is a state of altered or impaired consciousness.[142] The alien-self metaphor can once again help explain what happens, and why sleepwalkers, like the hypnotized, are kin to the insane. The consciousness of the insane and the somnambulist are altered insofar as the sense of agency is lost in both. Neither experiences himself as the author of what he does, and when agency is lost, the alien self is born. The real (waking) self becomes dissociated from the alien (sleeping) self.[143] The alien self takes charge, and the real self becomes a passive bystander. The alien self may be acting out a dream, or something like a dream.[144] The

---

140. Roger Broughton et al., *Homicidal Somnabulism: A Case Report*, 17 SLEEP 253, 255 (1994) (detailed discussion of Parks case); Kenneth J. Weiss et al., *Parasomnias, Violence, and the Law*, 39 J. PSYCHIATRY & L. 249 (2011).

141. R. v. Parks, [1992] 2 S.C.R. 871, 880 (Can.).

142. *See* Zadra et al., *supra* note 136, at 285 ("Somnambulism is defined as a series of complex behaviours that are usually initiated during arousals from slow-wave sleep and culminate in walking around with an altered state of consciousness and impaired judgment.") (internal quotations omitted).

143. Dev Banerjee & Angus Nisbet, *Sleepwalking*, 6 SLEEP MED. CLINICS 401, 410 (2011) ("[S]leepwalking and other non-REM parasomnias might arise from a dissociation of sleep and wakefulness occurring across different brain regions . . . ."); Mark W. Mahowald et al., *State Dissociation: Implications for Sleep and Wakefulness, Consciousness, and Culpability*, 6 SLEEP MED. CLINICS 393, 395–96 (2011) ("Disorders of arousal are the most impressive and most frequent of the NREM sleep-state dissociation/admixture phenomena . . . . Disorders of arousal simply represent the simultaneous occurrence of [wakefulness] and NREM sleep.").

144. *See, e.g.*, Delphine Oudiette et al., *Dreamlike Mentations During Sleepwalking and Sleep Terrors in Adults*, 32 SLEEP 1621, 1626 (2009) ("[D]reamlike mentations (mostly brief, frightening visual images) may occasionally exist during sleepwalking and sleep terrors, suggesting that a

alien self of sleepwalking isn't usually dangerous. Usually, but not always. Sometimes something happens that precipitates a crime.[145] Sleep crimes tend to be violent, but when awake, the sleepwalkers who commit them tend not to be.

Again, sleepwalkers aren't exactly like the insane. First, the loss of agency in sleepwalking is associated with the psychology of sleep, not mental disease or defect. Second, unlike the hypnotized subject and the insane, the somnambulist may not be able to report what happened. Perhaps he doesn't remember, or perhaps he never forms memories about what happened in the first place. Still, sometimes the real self appears to get glimpses of what happened, and can recall bits and pieces,[146] as did Parks when he recalled the look on his mother-in-law's face. Finally, when the actor wakes up, the lost sense of agency ends, the alien self disappears, and the self re-integrates. It takes more for the insane to reintegrate than waking up.[147]

---

complex mental activity takes place during SWS. Sleepwalking may thus represent acting out of the corresponding dreamlike mentations."); Zadra et al., *supra* note 136, at 288 ("[E]mpirical evidence suggests that sleep mentation is not only frequently part of the main experience of somnambulism, but also can modulate motor behavior during an episode. . . . Furthermore, the mentation reported by patients was congruent with recorded nocturnal behavior, suggesting that sleepwalking might be the acting out of dreamlike mentations.").

145. Weiss et al., *supra* note 140, at 280 ("It appears that violent behavior can occur when NREM sleep is interrupted, during somnabulistic episodes, upon incomplete arousal from sleep.").

146. *See, e.g.*, Mark R. Pressman, *Sleepwalking, Amnesia, Comorbid Conditions and Triggers: Effects of Recall and Other Methodological Biases*, 36 SLEEP 1757, 1757 (2013) (noting that "recent reports of dream-like mentation associated with sleepwalking episodes and even the incorporation of elements of perceptual environment and behavior have suggested that amnesia for at least some patients and some episodes is not as complete as has been previously accepted"); Zadra et al., *supra* note 136, at 288 ("[M]any patients can and do recall at least portions of episodes upon awakening, and thus [data] suggest[s] that complete amnesia is not standard for adult sleepwalkers.").

147. Sleepwalking cases raise the same doctrinal question as do hypnosis cases: Is the accused not guilty because he didn't act, or because he acted but isn't responsible for some other reason? *See infra* note 135 and accompanying text.

3. Multiple Personality

In the popular imagination, the portrait of multiple personality disorder (now officially called dissociative identity disorder, or DID) is the character Sybil, who was alleged to have had *sixteen* different personalities or personality states.[148] Sybil didn't commit any crimes, nor did any of her personalities. Some multiples do. Take Bridget Denny-Shaffer. Denny-Shaffer was a delivery nurse at Rehoboth Hospital in Gallup, New Mexico.[149] On May 10, 1991, she went to the Memorial General Hospital in Las Cruces, and identified herself as Linda, a medical student from the University of New Mexico, who was doing a pediatric rotation.[150] She then took one of the babies in the Hospital's nursery, and headed for Texas.[151] Charged with kidnapping, Denny-Shaffer, who was diagnosed with multiple personality disorder (MPD), argued that she never kidnapped anyone.[152] Linda did.[153]

When we talk about multiple personality disorder we need to be careful. What exactly is going on? One theory, at

148. The character Sybil was based on Shirley Ardell Mason, who probably didn't actually suffer from MPD. *See* DEBBIE NATHAN, SYBIL EXPOSED: THE EXTRAORDINARY STORY BEHIND THE FAMOUS MULTIPLE PERSONALITY CASE (2011). An "alter" is thought to emerge in response to a particular emotional episode, which the alter can deal with better than the host. MPD is thought to arise as a reaction to trauma: the self splits in order to manage a traumatic event.

149. United States v. Denny-Shaffer, 2 F.3d 999, 1002 (10th Cir. 1993).

150. *Id.*

151. *Id.*

152. *See id.* at 1010.

153. *See id.* at 1002. The story is actually more complicated. Everyone agreed the accused's host personality, named "Gidget," wasn't in charge at the time of the kidnaping; one of the alters, either "Rina" or "Bridget," was. Anyhow, whoever was in charge at the time must have lied when she identified herself to the staff at the Memorial General Hospital in Las Cruces as "Linda." Appellate opinions addressing the criminal liability of defendants diagnosed with MPD are rare. The reported cases include Kirkland v. State, 304 S.E.2d 561 (Ga. Ct. App. 1983); Commonwealth v. Roman, 606 N.E.2d 1333 (Mass. 1993); State v. Grimsley, 444 N.E.2d 1071 (Ohio Ct. App. 1982).

least compared to the next, is metaphysically extravagant.[154] Assuming a simple case of MPD, with one "alter" and one "host," this extravagant account tells us we have two different people occupying the same body seriatim over time. If so, what should happen when, in a case like Denny-Shafer, the alter was in charge at the time of the crime? What should the verdict be? Guilty or not?

Two possibilities pop out. One would be to judge the mind of the alter. After all, the alter was the one in charge at the time of the crime. If she satisfies the elements of the crime charged and has no defense, then the verdict should be guilty. Of course, that would mean the host goes to jail too. Yet the host was (on this story) innocent, assuming she wasn't complicit in the alter's crime. That doesn't seem fair. Another possibility would be to judge the mind of the host (even though the alter was in charge at the time). If the host wasn't complicit, then the verdict should be not guilty, though that would mean the guilty alter goes free too, assuming of course that the alter had no defense of his own. That doesn't seem fair either. The metaphysically extravagant account, besides being extravagant, throws up some hard choices.

The second theory of MPD avoids these conundrums, because it doesn't rely on the unlikely metaphysics of the first. It doesn't hypothesize two different people in the same body. Instead, it supposes that multiple personality disorder involves one body and one person.[155] What gets multiplied

---

154. *See, e.g.*, ELYN R. SAKS, JEKYLL ON TRIAL 42–51 (1997) (describing metaphysically extravagant theory); Elyn R. Saks, *Multiple Personality Disorder and Criminal Responsibility*, 10 S. CAL. INTERDISC. L.J. 185, 189–90 (2001) (same).

155. *See* Jeanette Kennett & Steve Matthews, *Delusion, Dissociation and Identity*, 6 PHIL. EXPLORATIONS 31, 33 (2003) ("[E]ach of the symptoms in the modern cases [of MPD] may be assimilated to other well recognized psychiatric conditions. . . . [W]hat we have [in cases of MPD] are single persons with a serious mental illness which, like other serious mental illnesses, impairs the development and exercise of unified autonomous agency."); *id.* at 34 ("The evidence suggests . . . that alter personalities are mere person-fragments, and not

isn't the number of people in the body, but the number of selves in the person. When it comes to bodies and persons, the rule is "one to a customer."[156] One body, one person. Your person can handle multiple selves but your body can manage only one person per lifetime. For present purposes, let's use Occam's razor and eliminate unnecessary persons. Let's assume, therefore, that the second theory is right. Multiple personalities are really one person with multiple selves.[157]

The alien-self metaphor can help shed light on how MPD works according to the second account. In MPD-speak, the alter is the alien self, and the host is the real self. When the real self (host) loses its sense of agency,[158] the alien self (alter) emerges and takes control. The real self continues to experience a sense of ownership. He continues to recognize the body the alter is occupying as his body, but he loses any sense of agency over it.[159] Agency and control are instead

---

in the sense of being short-lived fully-fledged persons, but in the sense that alters are one-dimensional and lacking in character development."); Robert F. Schopp, *Multiple Personality Disorder, Accountable Agency, and Criminal Acts*, 10 S. CAL. INTERDISC. L.J. 297, 298 (2001) ("[T]o the extent that DID exculpates criminal defendants, it does so for the same reasons that support the exculpatory significance of impaired consciousness more generally.")

156. DANIEL DENNETT, CONSCIOUSNESS EXPLAINED 422 (1991).

157. The shift in the *Diagnostic and Statistical Manual of Mental Disorders* from "Multiple Personality Disorder" to "Dissociative Identity Disorder" reflects the shift in the psychiatric profession from the first theory to the second. As one of the psychiatrists behind the change put it: "Multiple personality carries with it the implication that [those with the disorder] really have more than one identity, . . . [but the real] problem is fragmentation of identity." Clyde Haberman, *Debate Persists Over Diagnosing Mental Health Disorders, Long After "Sybil,"* N.Y. TIMES, Nov. 24, 2014 (quoting Stanford psychiatrist David Spiegel).

158. *See* Kennett & Matthews, *supra* note 155, at 37–38 (noting that the experience of "depersonalization" is common to dissociative disorders and that depersonalization involves a "loss of feelings of agency"). Those who suffer from "depersonalization disorder" experience lost agency, but don't go onto form the delusions associated with MPD.

159. *See id.* at 33 ("[T]here is co-consciousness (or the so-called 'looking-on phenomenon). Some personalities claim they have phenomenological access to other personality states. It is not completely clear what this involves, but those patients' so-called alters who claim to experience it say they have an intimate and immediate observer-role in relation to other alters' thoughts and actions."); *id.* at 42 (suggesting "that amnesia with regard to important personal

vested in the alter, with the real self again reduced to the status of observer or onlooker. Sometimes the real self can't recollect what the alien self did, but not always. Sometimes, like the sleepwalker, he catches and recalls snippets of what happened. Of course, an alien self doesn't *really* take control. That's what the first theory says: it takes the metaphor literally. The second theory, in contrast, takes the alien-self metaphor as just that: a metaphor used to make sense of lost agency.

Compare Denny-Shaffer to Jane. Jane has been diagnosed with schizophrenia, including delusions of alien control. She too kidnaps a baby, experiencing no sense of agency at the time. She delusionally attributes her acts to the work of some force external to herself. She might not know what to make of this external force, let alone give it a name and personality. All she states with confidence is that she wasn't the one in command when the baby was taken. Unlike Jane, who doesn't give her alien overseer a name, Denny-Shaffer did. Indeed, she did more than give it a name. She delusionally invested it with a personality different, though perhaps not entirely different,[160] from her own. Still, Denny-Shafter's alien self was arguably just a delusion, albeit an exceptionally elaborate one, not a separate person commandeering her body from time to time.[161]

Multiple Personality Disorder is sometimes thought to send the law of insanity into a tailspin, and indeed it does,

---

information in DID is often to be understood in terms of the difficulty of incorporating traumatic and depersonalized or delusional experiences into autobiographical memory").

160. Alters tend to be stock characters. "One study reports that in 85% of the cases of DID there is a child alter, in 53% of the cases there is an opposite gender alter, in 52% of the cases there is a promiscuous alter; 22% of alters were judged to be hypomanic or manic and 38% were judged to be psychotic." *See id.* at 35.

161. Jeanette Kennett & Steve Matthews, *Identity, Control and Responsibility: The Case of Dissociative Identity Disorder*, 15 PHIL. PSYCHOL. 509, 511 (2002) ("[S]omeone with DID is an individual human person whose psychiatric symptoms . . . are akin to a species of *global self-delusion.* So-called alter personalities are not to be regarded as metaphysically separate entities from the person, but rather count as altered states of that person.") (emphasis added).

but only if multiples are multiple people in one body (the first theory), and only if insanity is, as the traditional and MPC tests say, a matter of incapacity resulting from mental disease or defect. If the alter who commits the crime suffers from no such incapacity, and if the host, when back in charge, suffers from no such incapacity, where's the insanity? In contrast, if MPD involves multiple selves in one person (in one body) (the second theory), and if insanity is understood as consisting in lost agency, the problem goes away. Far from being a *problem* case for insanity, MPD cases rather turn out to be *paradigm* cases of insanity. Sybil turns out to be insanity's poster child.

IV. M'NAGHTEN, AGAIN

We end where we began, with Daniel M'Naghten. When M'Naghten shot Drummond, did he act with a sense of agency, or had that sense abandoned him? If he retained his sense of agency, then he was sane; if not, then insane.

Alas, it's hard to tell. Someone watching him from afar would surely think he was in command as he approached Drummond, removed the pistol from his coat, fired, and then tried to reach for another pistol to fire again. Still, appearances can deceive. M'Naghten's body could act with purpose even if M'Naghten wasn't at the helm.[162] Indeed, though seldom remarked upon,[163] M'Naghten's defense at trial was based almost entirely on the claim that (in some

162. The expert defense witnesses testified that M'Naghten suffered primarily from persecutory delusions, as well as delusions of reference, both of which can arguably rise from a lost sense agency. Dr. Munro testified that M'Naghten stated that "he had seen paragraphs in *The Times* newspaper containing allusions which he was satisfied were directed at him; he had also seen articles in the *Glasgow Herald*, beastly and atrocious, insinuating things untrue and insufferable of him . . . ." SCHNEIDER, *supra* note 1, at 212–13.

163. For an exception, see Eigen, *supra* note 50, at 45 ("The most articulate pairing of delusion with loss of self-control was made only at the very end of the period under review, during the trial of Daniel McNaughtan.").

sense) he couldn't control himself,[164] not that he didn't know he was committing a crime. Consider the following bits of uncontested testimony from the defense doctors:

> Dr. Munro: "The act with which he is charged, coupled with the history of his past life, leaves not the remotest doubt on my mind of the presence of insanity sufficient to deprive the prisoner of all self-control."[165]
>
> Dr. Morison: "[His delusions] deprived the prisoner of all restraint over his actions."[166]
>
> Dr. M'Clure: "I consider when he fired at Mr. Drummond, at Charing Cross, he was suffering from an hallucination which deprived him of all ordinary restraint."[167]
>
> Dr. Hutcheson: "The prisoner had lost all self-control at the moment he fired at Mr. Drummond. The act flowed immediately from the delusion. . . . [T]he act was the consequence of the delusion, which was irresistible."[168]

What were these doctors saying? Someone versed in the long-standing debate between cognitive and volitional insanity tests might think the answer is obvious, and maybe it is. Maybe the doctors were saying M'Naghten suffered from an "irresistible impulse" or lacked the power to control himself when he shot Drummond. M'Naghten, due to a mental disease or defect, couldn't have chosen otherwise. Simple as that.[169] If so, then however delusional he was, he

---

164. The opinion in *M'Naghten*, in its summary of the "substance" of the medical testimony, said: "[I]t was of the nature of the disease with which the prisoner was affected, to go on gradually until it had reached a climax, when it burst forth with irresistible intensity: that a man might go on for years quietly, though at the same time under its influence, but would all at once break out into the most extravagant and violent paroxysms." M'Naghten's Case (1843) 8 Eng. Rep 718, 719.

165. SCHNEIDER, *supra* note 1, at 213. Dr. Munro was a physician with thirty years' experience practicing at Bethlem.

166. *Id.* at 220. Dr. Morison was "a physician at St. Luke's Hospital and also affiliated with Bethlem Hospital and the Surrey Asylum." *Id.* at 219.

167. *Id.* at 220. Dr. M'Clure was a "London surgeon who had accompanied Munro and Morison in the prison examination of M'Naghten." *Id.*

168. *Id.* Dr. Hutcheson was a "physician to the Royal Lunatic Asylum in Glasgow." *Id.*

169. Or maybe their testimony should be taken as a testament to the idea that

should have been convicted under the traditional test if he realized he was committing a crime, and if, for example, he would've stopped himself if a gallows had suddenly materialized before his eyes as he approached Drummond from behind, and if he believed he'd be hung from it if he pulled the trigger.

Yet maybe the doctors were saying, or trying to say, something very different. Maybe all the talk about self-control and restraint wasn't really about M'Naghten's capacity to conform to the law. Maybe instead it was about lost agency. Maybe the doctors were trying to say, albeit using the language of self-control, that M'Naghten wasn't in command at the time he killed Drummond. Someone or something else was. That something or someone else might or might not have pulled the trigger if the gallows appeared. But that doesn't matter, according to the lost-agency theory. If M'Naghten wasn't at the helm when his body shot Drummond, the alien self's capacity to have chosen otherwise is beside the point. All that matters is that M'Naghten wasn't in control. For if he wasn't in control, he wasn't responsible for pulling the trigger. He was insane.

We usually assume when someone's body moves, he or she is the one moving it. That assumption usually holds, but not when insanity takes hold. The insane actor's mind and body commit the crime, but the mind and body committing the crime are not under his command. They've been commandeered. The choices his mind makes, the reasons moving his mind to make those choices, and the bodily movements resulting from those choices, are no longer experienced as *his* choices or *his* reason or *his* movements. An alien self is the one pulling the strings. It would therefore make no more sense to blame him for the crime resulting from those choices, reasons and movements than it would be to blame you or me. Blame presupposes a sense of agency,

---

cognitive and volitional impairments amounting to insanity can't, in the end, so easily be distinguished and kept separate from each other.

and insanity precludes blame because insanity defeats the sense of agency.

Everyone agrees that insanity blocks criminal liability. The challenge has long been to explain *why*. Perhaps insanity blocks liability because insanity is incapacity, and incapacity blocks liability. So says the traditional theory. Or perhaps insanity blocks liability because insanity is irrationality, and irrationality blocks liability. So say the irrationality theories. Neither of these accounts fully satisfies. That dissatisfaction prompts the search for another explanation. Perhaps, instead, insanity blocks liability, not because the insane are compelled or irrational, but because they choose and act without a sense of agency. So says the lost-agency theory. An insane actor is, quite literally, out of *his* mind.[170]

170. If insanity does indeed consist in lost agency, the upshot is ironic. Here's the irony: Determinism tells us that we're not really in control of what we do, at least not if being in control means having the contra-causal capacity to choose otherwise, and at least not if we lack that capacity. Be that as it may, our brains trick us into thinking that we do have it. The only ones whose brains don't trick them are those that lack a sense of agency. So if it's crazy to think we have contra-causal powers, then the only ones who aren't crazy are the insane. *Cf.* KEAN, *supra* note 80, at 264 ("[V]ictims of alien hand syndrome and other syndromes may have simply lost the illusion of free will for part of their bodies. In some sense, they might be closer to the reality of how the brain works than the rest of us. Makes you wonder who's really deluded.").