1-1-1993

Competency to Refuse Psychotropic Medication: Three Alternatives to the Law's Cognitive Standard

Elyn R. Saks

Competency to Refuse Psychotropic Medication: Three Alternatives to the Law's Cognitive Standard

ELYN R. SAKS*

* Associate Professor, University of Southern California Law Center. B.A., Vanderbilt University, 1977; M. Litt., Oxford University, 1981; J.D., Yale Law School, 1986. I am grateful to Scott Altman, Richard Bonnie, Richard Craswell, John Finnis, Martin Levine, Leonard Long, Edward McCaffery, Michael Shapiro, Larry Simon, Christopher Slobogin, Lawrence Solum, Charles Weisselberg, Stephen Wizner, and the participants of the University of Miami School of Law Symposium on Competence and of the U.S.C./Oxford Legal Theory Workshop for their helpful comments on earlier drafts of this article. I also thank Steven Glovsky, Sylvia Smith, and Stephen Sonnenberg for their valuable research assistance. Finally, I thank Susan McGlamery of the U.S.C. Law Library for her valuable library assistance.

I. INTRODUCTION

Historically, involuntarily confined psychiatric patients have been an exception to the general rule that competent patients still have a right to determine their own treatment.[1] The state's confinement of psychiatric patients for purposes of treatment was thought to authorize the state to impose treatment even if the patient protested. In the last ten years or so, advocates for the mentally disabled have challenged this historical practice.[2] Their efforts to secure a right of competent patients to refuse psychotropic medication have been largely successful.[3] Although a recent Supreme Court case makes clear that the United States Constitution does not underpin a right of competent patients as such to refuse psychotropic medication (at least

1. *See, e.g.*, Washington v. Harper, 494 U.S. 210, 225-26 (1990); Price v. Sheppard, 239 N.W.2d 905, 909, 913 (Minn. 1976); *In re* Hospitalization of B, 383 A.2d 760, 763 (N.J. 1977); State *ex rel.* Jones v. Gerhardstein, 416 N.W.2d 883, 892 (Wis. 1987). Competent patients have a right to refuse treatment. *See, e.g.*, Cruzan v. Director, Missouri Dep't of Health, 497 U.S. 261; Walters v. Western State Hosp., 864 F.2d 695, 697 (10th Cir. 1988); *see also* 1 MICHAEL L. PERLIN, MENTAL DISABILITY LAW: CIVIL AND CRIMINAL § 3.12, at 217 (1989). As a threshold matter, competency was not a bar to the states' involuntary confinement of psychiatric patients. Even today, states may involuntarily confine those patients who are dangerous to themselves or others, notwithstanding their competency. *See, e.g.*, CAL. WELF. & INST. CODE § 5300 (West 1991); MINN. STAT. ANN. § 253B.18 (West Supp. 1992); WIS. STAT. ANN. § 51.20 (West Supp. 1992).

2. *See, e.g.*, Mills v. Rogers, 457 U.S. 291 (1982); O'Connor v. Donaldson, 422 U.S. 563 (1975); Rennie v. Klein, 720 F.2d 266 (3d Cir. 1983); Rogers v. Okin, 738 F.2d 1 (1st Cir. 1984); *see also* John Parry, *Decision-making Rights over Persons and Property, in* SAMUEL J. BRAKEL ET AL., THE MENTALLY DISABLED AND THE LAW 435 (1985); Michael L. Perlin, *State Constitutions and Statutes as Sources of Rights for the Mentally Disabled: The Last Frontier*, 20 LOY. L.A. L. REV. 1249 (1987).

3. I speak throughout of a right of competent patients to *refuse* medication. A right to consent to medication has always existed, even when the patient was *not* competent to make this judgment. *See, e.g.*, George J. Annas & Joan E. Densberger, *Competence to Refuse Medical Treatment: Autonomy vs. Paternalism*, 15 U. TOL. L. REV. 561, 575 (1984); Perlin, *supra* note 1, at 217; David B. Wexler, *Reflections on the Legal Regulation of Behavior Modification in Institutional Settings*, 17 ARIZ. L. REV. 132, 135-36 (1975); Bruce J. Winick, *Competency to Consent to Treatment: The Distinction Between Assent and Objection*, 28 HOUS. L. REV. 1527. As a theoretical matter, we should require competency for consent as well as refusal. But as a practical matter, we should not, because the overwhelming majority of patients found incompetent to consent will nevertheless be medicated on best interest or substituted judgment grounds. Thus, requiring a process to determine competency before medicating the patient both wastes resources and further stigmatizes him, by finding him incompetent, to no good end.

This is not to say that no procedures are needed in the case of accepters. For example, perhaps caregivers should inform all patients of their right to refuse and document the

in the case of mentally ill prisoners),[4] most courts considering the issue on state constitutional, statutory, or common law grounds have found such a right.[5]

The question then arises: what is it to be competent to refuse psychotropic medication? In an earlier article on this topic, *Competency to Refuse Treatment*,[6] I argued that the law's treatment competency standard[7] is very attractive, surviving comparison with a number of competing cognitive standards.[8] That standard, briefly stated, is that a person is competent to make a treatment decision if she comprehends the caregiver's explanation of her condition and the treatment, and forms no patently false beliefs—"delusions"—about her condition and the treatment. This standard results in some controversial conclusions, such as that a psychiatric patient is competent to refuse medication even though her refusal is based on the belief that she is not ill or the belief that she is bad and deserves to suffer. These beliefs simply are not delusions in the law. Yet the standard that justifies these conclusions seems clearly justified on philosophical grounds.

This Article evaluates alternative treatment competency stan-

---

existence of consent in some way. But once there is evidence that the consent is genuine, no further inquiry may be necessary.

Of course, competency is not the only issue. Caregivers may deny patients the right to refuse on bases other than incompetency, such as the patient's danger to others. *See, e.g.*, Davis v. Hubbard, 506 F. Supp. 915, 937 (N.D. Ohio 1980); *In re* Orr, 531 N.E.2d 64, 73 (Ill. App. Ct. 1988); Gundy v. Pauley, 619 S.W.2d 730, 731 (Ky. Ct. App. 1981); Williams v. Wilzack, 573 A.2d 809, 810 (Md. 1990). When I speak of a right of competent patients to refuse, I mean to refer to cases in which the choice is purely self-regarding.

4. *See* Washington v. Harper, 494 U.S. 210 (1990) (prisoner-patients who are mentally ill and dangerous or gravely disabled may be medicated if the medication is in their best interests, even if the protesting patient is competent).

5. For courts that have found a right of competent patients to refuse medication in non-emergencies on state law grounds, see, e.g., Nolen II v. Peterson, 544 So. 2d 863 (Ala. 1989); Riese v. St. Mary's Hosp. & Medical Ctr., 271 Cal. Rptr. 199 (Cal. Ct. App. 1987); Bradshaw v. State, 816 P.2d 986 (Idaho 1991); *In re* Orr, 531 N.E.2d 64 (Ill. App. Ct. 1988); Rogers v. Comm'r of Mental Health, 458 N.E.2d 308 (Mass. 1983); Opinion of the Justices, 465 A.2d 484 (N.H. 1983); Rivers v. Katz, 495 N.E.2d 337 (N.Y. 1986); *In re* Guardianship of Willis, 599 N.E.2d 745 (Ohio Ct. App. 1991); State *ex rel.* Jones v. Gerhardstein, 416 N.W.2d 883 (Wis. 1987). A small number of courts have failed to recognize a right of competent patients to refuse medication on the basis of statutes permitting medication in some non-emergency circumstances. *See, e.g., In re* Mental Commitment of M.P., 510 N.E.2d 645 (Ind. 1987); Williams v. Wilzack, 573 A.2d 809 (Md. 1990); State v. Nording, 485 N.W.2d 781 (N.D. 1992). To the extent that litigators continue after *Harper* to bring right-to-refuse actions on federal constitutional grounds, we must presume that they have judged state law to offer no greater protection than the federal constitution.

6. Elyn R. Saks, *Competency to Refuse Treatment*, 69 N.C. L. REV. 945 (1991).

7. I use the phrase the "law's treatment competency standard" as arbitrary shorthand for my reading of a set of cases in the United States in the twentieth century.

8. *Id.* at 947-48.

dards that, unlike the law's standard, place more weight on noncognitive impairments. Although it seems true that severe cognitive impairments impact one's ability to make decisions, it is too late in the day to ignore that serious mental illness can grossly impair one's functioning on levels other than the cognitive; and perhaps those impairments also have an important impact on one's decisionmaking ability.[9] Sufferers of serious mental illness may experience severe dysfunctionality in their mood and behavior as well as in their thinking. Thus, a manic-depressive patient may swing from utter despair to wild elation; and decisions consistent with her mood may be inevitable. Similarly, an impulse-control disordered patient may be completely at the mercy of his impulses. His choice to gratify those impulses may not really be a choice in the true sense. Competency standards that give a more central role to mental illness, and *all* the impairments it may cause, are well worth considering.

This Article draws on the competency literature for three such standards: a "different person" standard, a "volitional impairment" standard, and a "product of mental illness" standard. The different person standard finds incompetent those whom mental illness has so transformed that they are as if different persons, and so make "inauthentic" choices that do not reflect their true selves. The volitional impairment standard finds incompetent those who are so powerfully moved to act that we cannot expect them to resist. The product of mental illness standard finds incompetent those whose choices are a product of their mental illness. The law has endorsed each of these standards in one form or another. The first appears in cases involving wills,[10] while the second and third are familiar from the criminal law.[11] Commentators have also approved of these standards.[12] Each seems prima facie appropriate to measure treatment

9. *See, e.g.*, United States v. Smith, 404 F.2d 720, 725 (6th Cir. 1968); Durham v. U.S., 214 F.2d 862, 870-71 (D.C. Cir. 1954); *see also* S. SHELDON GLUECK, MENTAL DISORDER AND THE CRIMINAL LAW 138-39 (1925); MANFRED S. GUTTMACHER & HENRY WEIHOFEN, PSYCHIATRY AND THE LAW 403-08 (1952); ROYAL COMMISSION ON CAPITAL PUNISHMENT 1949-1953, Report 73-129, Report 80 (1953); S. Sheldon Glueck, *Psychiatry and the Criminal Law*, 12 MENTAL HYGIENE 569, 574 (1928); Simon E. Sobeloff, *Insanity and the Criminal Law: From M'Naghten to Durham, and Beyond*, 41 A.B.A. J. 793, 794 (1955).

10. *See infra* note 22.

11. *See infra* parts III.B., IV.A.

12. Ruth Macklin, *Treatment Refusals: Autonomy, Paternalism, and the "Best Interest" of the Patient*, *in* ETHICAL QUESTIONS IN BRAIN AND BEHAVIOR 46-47 (Donald W. Pfaff ed., 1983); Alexander M. Meiklejohn, *Contractual and Donative Capacity*, 39 CASE W. RES. L. REV. 307, 345-52 (1988-89); Bruce L. Miller, *Autonomy & the Refusal of Lifesaving Treatment*, HASTINGS CENTER REP., Aug. 1981, at 22; Henry Weihofen, *Mental Incompetency to Contract or Convey*, 39 S. CAL. L. REV. 211, 220 (1966); Comment, *Mental Illness and the Law of Contracts*, 57 MICH. L. REV. 1020, 1031-36 (1959). *Cf.* E. ALLAN

competency.

Although simply relying on the literature for appropriate standards is less satisfactory than deriving them from a careful consideration of clinical reality, each of these standards seems to plausibly capture something of what is wrong with the decisions of impaired mentally ill people. For example, the decisions of mentally ill people may appear problematic because the people do not seem to be "themselves." When mental illness transforms a quiet, restrained, inhibited person into an outgoing, uninhibited, seductive person, one may question the right and competency of the latter to speak for the former. Perhaps we should not respect his choice to wildly proposition others, when his "true" self is likely to return to suffer the consequences. In this case, a different person standard may seem appropriate.

Similarly, the decisions of some mentally ill people seem problematic because one suspects they cannot control themselves. Consider, for example, the purchasing decisions of a manic patient on a shopping spree or the pyromaniac's decision to burn down a building. A volitional impairment standard may be warranted to address these cases.

The product of mental illness standard may also capture something, though in a more inchoate manner, of what is wrong with the choices of the mentally ill. This Article examines what might be problematic about such choices. Thinking about the role of mental illness in producing the choice, and all the impairments mental illness causes, may advance our inquiry in a way that merely focusing on particular impairments does not.

I bill these three standards, in contrast to the law's standard, as "noncognitive." A cognitive competency standard focuses on impairments in thinking. This includes impairments in absorbing and manipulating information, or, in other words, in comprehension and reasoning. It also includes impairments in the ability to assess evidence or to form acceptable beliefs. The treatment context most often implicates the latter ability because information about psychotropic medication is relatively simple and straightforward.

The distinction between cognitive and noncognitive impairments is not sharp, and nothing turns on its being sharp. Mood-disordered people, for example, may suffer clear cognitive impairments.[13] Some

---

FARNSWORTH, FARNSWORTH ON CONTRACTS § 4.6 (1990) (describing volitional impairment test in contracts context).

13. The same may be true of people with behavioral impairments. For example, impulse control-disordered people may systematically overvalue immediate gratification, and this may involve systematic cognitive errors (e.g., about the likelihood of their feeling a certain way in the future). Similarly, obsessive-compulsive people who repetitively perform certain rituals

disorders, such as schizo-affective disorder, explicitly comprise mood and thought defects.[14] Even "pure" mood disorders may be accompanied by either mood-congruent or mood-incongruent psychotic features such as delusions.[15] Thus, standard psychiatric nosologies acknowledge that mood and thought defects may coexist.

More fundamentally, however, one wonders whether the very idea of a disordered mood that does not involve any cognitive distortions even makes sense. To feel low may be necessarily to tell oneself discouraging things about one's worth and/or to feel pessimistic about the future, and these thoughts may be distorted or false. Similarly, to feel high may be necessarily to have an exalted sense of one's worth and/or to feel optimistic about the future. These thoughts, too, may be false. The very notion of a low or high mood may refer essentially to disordered cognition, that is, to beliefs about oneself and the world that distort reality.[16] The idea of a purely noncognitive impairment may, therefore, be a myth.

The view that no bright line separates cognitive and noncognitive impairments seems fairly persuasive. Given that the mind is a unity, it would be surprising if such a bright line did exist. This means that it is something of a misnomer to bill the three competency standards considered in this Article purely "noncognitive." Indeed, I come in the end to gloss the product of mental illness standard in fairly cognitive terms. Still, the standards were not chosen at random. They

---

may have distorted beliefs about what might or will happen if they do not perform their rituals.

14. *See* AMERICAN PSYCHIATRIC ASSOCIATION, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS (3d ed. rev. 1987), at 208-10 [hereinafter DSM-III-R].

15. *See, e.g., id.* at 216, 218, 220, 223. Like thought disorders, mood disorders may involve impairments in the form no less than the content of thought. Thus, a manic patient may exhibit "flight of ideas," *id.* at 214-15, and a depressed patient may exhibit slowed thinking, *id.* at 219. As a result, each may suffer in his ability to comprehend information. I focus here on delusions because it is delusions that most often compromise decisions about psychotropic medication.

16. The law's cognitive standard refers to *severe* distortions of reality. Although some mood-disordered people suffer from these and receive a special code in DSM-III-R to indicate the presence of psychotic features, *see* DSM-III-R, *supra* note 14, at 218, 223, most do not. Thus, although disordered mood may necessarily involve disordered cognition, I do not suggest that it involves *severely* disordered cognition of the kind with which the law's standard is concerned. This raises the possibility that noncognitive standards seem attractive because they allow evaluators to find people incompetent or insane because of disordered cognition that does not rise to the level that the law's standard requires. Thus, it might be more accurate to refer to these standards as "weak cognitive standards" rather than as "noncognitive standards." Although some might wish for a competency standard that disables people with weakly disordered thinking, and thus might find the standards considered here attractive because they do, I am not in this category. I find the law's cognitive standard itself attractive and noncognitive standards tempting only insofar as they refer to deficiencies that are not essentially cognitive. It is these that I focus on in this Article.

share the feature of referring to defects that are not *purely* or *essentially* cognitive. Even though impaired mood, for example, may involve impaired cognition to some extent, the latter does not exhaust the former. It is the residue that noncognitive standards are concerned with.

It is important to stress as well that these standards may not be mutually exclusive. The first and the third are commonly proposed as *the* standard that should govern competency. The volitional impairment standard, by contrast, is commonly recommended as a supplement to a cognitive standard. In principle, however, some or all of these standards might be warranted to capture all of the people whose decisions should not be legally respected. For example, although each standard might be generally applicable to a large range of cases, some of the standards might be easier to apply to particular cases. Alternatively, each standard might be necessary to capture a narrow range of cases. Without them, one might be forced to overlook some deserving cases. There may be reasons to fear a world with multiple incompetency standards. Evaluators would thereby have more authority to disrespect people's decisions. As it turns out, however, each of these standards is so deficient in and of itself that it must be rejected.

Although this Article is concerned with how we should understand the concept of competency to refuse treatment, it is worth pointing out that, as a normative matter, granting competent psychiatric patients the right to refuse seems sound. If mental illness were always simply unconventionality, or if the drugs were in a deep sense problematic—ineffective, unduly risky, perhaps even evil—this would be an easy question. Granting patients the right to refuse would allow them to avoid something clearly harmful at no cost. Although many legal and other commentators seem to hold these beliefs about the illness and the drugs,[17] I believe quite the opposite: that mental illness is often crippling and that the medication is generally effective, poses acceptable risks, and is certainly not evil. As a result, this is a harder question for me because a right to refuse would permit competent

17. For commentators who seem to believe that mental illness may be simply unconventionality, see, e.g., BRUCE J. ENNIS, PRISONERS OF PSYCHIATRY: MENTAL PATIENTS, PSYCHIATRISTS, AND THE LAW 216-17 (1972); THOMAS S. SZASZ, LAW, LIBERTY, AND PSYCHIATRY 15-16, 43 (1963). For commentators who seem to believe the drugs are ineffective, risky, and perhaps even evil in some sense, see, e.g., PETER R. BREGGIN, PSYCHIATRIC DRUGS: HAZARDS TO THE BRAIN 2-8, 168-70 (1983); Robert Plotkin, *Limiting the Therapeutic Orgy: Mental Patients' Right to Refuse Treatment*, 72 NW. U. L. REV. 461, 474-79 (1977); Jessica Litman, Comment, *A Common Law Remedy for Forcible Medication of the Institutionalized Mentally Ill*, 82 COLUM. L. REV. 1720, 1726-27 (1982). *See generally* Mary McCanon, *The Right to Refuse Antipsychotic Drugs: Safeguarding the Mentally Incompetent Patient's Right to Procedural Due Process*, 73 MARQ. L. REV. 477 (1990).

patients to make what to me is clearly the wrong choice.[18]

Granting competent patients a right to refuse is nevertheless justified for two reasons. First, such a right will promote respectful conversation between doctors and patients.[19] In the past, it was more expedient for doctors simply to use force. With a right to refuse, doctors will have an incentive to *talk* to patients. Talking to patients shows them respect as people.

More importantly, granting competent patients a right to refuse acknowledges that the patient himself is in the best position to judge what is right for him. Another's judgments about the medication may be wrong, or wrong for the patient. The patient may have perfectly rational reasons for coming to a different view of the medication in his case.[20] So long as the patient is competent, I trust him to make his own judgments about the costs and benefits of taking the medication as he perceives them. Although many involuntary patients are fairly disturbed, they may nevertheless retain the faculties to make competent choices.[21] If it is normatively desirable to allow competent

---

18. Although I believe that, properly used, psychotropic medication is often extremely effective at fairly low cost, medication practices at some state hospitals leave something to be desired. Among other things, doctors overmedicate patients, practice polypharmacy, do not properly monitor the tardive dyskinesia risk, and do not attend to ameliorating unpleasant side effects. *See, e.g.*, Rennie v. Klein, 476 F. Supp. 1294, 1299-1303 (D.N.J. 1979) (describing substandard medication practices in New Jersey state hospitals), *modified and remanded*, 653 F.2d 836 (3d Cir. 1981), *vacated*, 458 U.S. 1119 (1982); Steven J. Schwartz, *Damage Actions as a Strategy for Enhancing the Quality of Care of Persons with Mental Disabilities*, 17 N.Y.U. REV. L. & SOC. CHANGE 651 (1990). In such cases, refusal may well be the right choice—but for the sake of inducing one's doctors to take better care in managing the medication, not for the sake of avoiding the medication *per se*. *See infra* text accompanying note 19 (discussing the value of a right to refuse in promoting conversation and negotiation between doctors and patients).

19. *See, e.g.*, JAY KATZ, THE SILENT WORLD OF DOCTOR AND PATIENT 225-29 (1984); Alexander D. Brooks, *The Right to Refuse Antipsychotic Medications: Law and Policy*, 39 RUTGERS L. REV. 339, 369 (1987).

20. For example, the patient may find the drug-induced state so uncomfortable that he prefers the mentally ill state. Or the patient may so disvalue an outcome risked by the medication, such as disfiguring facial movements, that he prefers not to take the medication even though the risk is remote. Similarly, the patient may wish to survive by his own efforts. Acknowledging a need for medication creates a sense of shame and helplessness. Finally, the patient may prefer being mentally ill to being healthy, and, in some circumstances, such a choice may be rational. For example, the patient's life may be terribly bleak or painful, or he may derive important primary or secondary gains from his ill state of mind.

21. Courts and commentators recognize today that even being ill enough to be civilly committed is not equivalent to being incompetent. *See, e.g.*, Rennie v. Klein, 720 F.2d 266 (3d. Cir. 1983); Winters v. Miller, 446 F.2d 65, 68 (2d Cir.), *cert. denied*, 404 U.S. 985 (1971); Rogers v. Okin, 478 F. Supp. 1342, 1361-62 (D. Mass 1979), *aff'd in part and rev'd in part*, 634 F.2d 650 (1st Cir. 1980), *vacated sub nom.*, Mills v. Rogers, 457 U.S. 291 (1982); Rogers v. Comm'r of Dep't of Mental Health, 458 N.E.2d 308, 313-14 (Mass. 1983); S— v. S—, 490 S.W.2d 344, 351 (Mo. Ct. App. 1973); *see also* JAMES D. PAGE, PSYCHOPATHOLOGY, THE

patients to refuse medication, it becomes more urgent to decide what it is to be competent to refuse.

II. THE DIFFERENT PERSON TEST

A. *Introduction*

The different person theory rationalizes the intuition that the mentally ill are incompetent by focusing on the change in personality that mental illness brings about and the effects of that change on decisional capacity. The theory holds that a person is incompetent, not if what appear to be her values[22] and beliefs are unacceptable according to some external standard, but rather if they are not *her* values and beliefs, because mental illness has transformed her into a "different person." The person has lost touch with her own values and ways of looking at the world; she is simply not *herself*.[23]

The different person theory has had some currency in the law. For example, the testamentary capacity cases, with their symbolic notion of testation as representative of the testator's psychic will, at times use language suggestive of this theory, such as when they say that a will is not truly the testator's.[24] Similarly, on one reading, the

---

SCIENCE OF UNDERSTANDING DEVIANCE, 32-35 (1971); Note, *Developments in the Law: Civil Commitment of the Mentally Ill*, 87 HARV. L. REV. 1190, 1214-15 (1974).

22. I address no other competency theory that focuses on the role of one's *values* in decisionmaking, although decisionmaking clearly involves not only one's beliefs, but also one's values. A person values some goal, and, in light of her beliefs about the world, makes choices to serve that goal. Some commentators have attempted to specify irrational or otherwise impermissible values and preferences that vitiate consent. *See, e.g.*, CHARLES M. CULVER & BERNARD GERT, PHILOSOPHY IN MEDICINE: CONCEPTUAL AND ETHICAL ISSUES IN MEDICINE AND PSYCHIATRY 35-37 (1982); MICHAEL S. MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP 101-04 (1984). Yet the doctrine of competency accords people the freedom to identify the good for themselves—in essence, to choose to express their own values in decisions. The different person theory does not require the decisionmaker to hold any conventional values, but rather simply to be faithful to her own values.

23. For commentators in the treatment competency context who seem to support some kind of different person theory, see, e.g., 1 PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, MAKING HEALTH CARE DECISIONS 71 (1982) (explicates competency standard in "authenticity" language); Gerald Dworkin, *Moral Autonomy*, *in* MORALS, SCIENCE, AND SOCIETY (H.T. Englehardt & D. Callahan eds., 1987); Macklin, *supra* note 12; Miller, *supra* note 12; Anne-Marie Pavlo et al., *Christian Science and Competence to Make Treatment Choices: Clinical Challenges in Assessing Values*, 10 INT'L J.L. & PSYCHIATRY 395 (1987) (assumes that evaluating competency requires, at least in part, identifying patients' true values); Martha A. Matthews, Note, *Suicidal Competence and the Patient's Right To Refuse Lifesaving Treatment*, 75 CAL. L. REV. 707, 754-55 (1987) (proposes strict authenticity requirement for decisions to die). *But see* RUTH R. FADEN ET AL., A HISTORY AND THEORY OF INFORMED CONSENT 262-68 (1986) (considers and rejects idea of authenticity as requirement of competency).

24. This language tends to appear in the older testamentary capacity cases. *See, e.g.*, Scott

different person theory underlies the product of mental illness theory discussed in Part IV: the crime was a result of the mental illness "speaking," and not the defendant himself.

The different person theory has several virtues. First, it does not suppress unconventionality or find too many incompetent given the pervasive irrationality of mental life. That is, it sidesteps the "unconventionality" and "irrationality" criteria I identified in my last article: the decisionmaker's thinking and feeling are not forced into some conventional mold, and her irrationality is not germane. The different person theory requires only that the decisionmaker be true to herself.

A second virtue of the different person theory derives from its focus on the changes mental illness causes—and a person's consequent alienation from her true self—instead of on specific impairments in a person's decisionmaking capacity. One of the key problems in competency theory is the difficulty of linking up discrete conceptualizations of decisional impairments with the fullness and highly integrated experience of mental deterioration or mental illness. In essence, a problem exists in phenomenologically characterizing the state of incompetency—a problem that results in part from the poverty of language. This theory, however, does not require us to characterize discrete impairments, but merely to select those who are so mentally alienated from their true selves that it is not *they* who are speaking. In a sense, the theory permits us to focus on the phenomenological manifestations of mental illness, in all their richness, in order to *identify* those who are not fit to decide for their true selves. There is no need to identify which of these phenomenologically rich manifestations is critical to incapacity.

The different person theory plausibly answers some perplexing questions. For example, it explains what "different person-ness" has to do with incompetence. On the face of it, the "different person" can himself be quite capable—suffer no decisional impairments at all. But the "different person," because he is not the true self, does not know the true self's values and needs, and to that extent *is* incapable of making a decision that expresses those values and needs. Similarly, the different person theory responds to the concern that the evaluator may choose among the decisionmaker's competing values when they conflict, thus undermining one of the key the purposes of competency

---

v. Scott, 72 N.E. 708, 708 (Ill. 1904); Orchardson v. Cofield, 49 N.E. 197, 202 (Ill. 1897); O'Dell v. Goff, 112 N.W. 736, 738 (Mich. 1907); Irwin v. Lattin, 135 N.W. 759, 764 (S.D. 1912). The language also frequently appears in the "undue influence" cases. *See, e.g.*, Odorizzi v. Bloomfield Sch. Dist., 54 Cal. Rptr. 533, 539-40 (Cal. Ct. App. 1966). In the undue influence cases, the language is especially appropriate because it gives a sense of whose will (when not the maker's) the agreement does represent—namely, the undue influencer's.

doctrine. In reality it is the true self who makes the value choice, which the evaluator simply reflects in her decision. Finally, the theory solves the problem that people do not want evaluators to override their choices merely because they later wish they had chosen otherwise; they identify with the self thus thwarted. If the choice not honored is a different person's, the real self does not suffer the pain and indignity of having *her* choice overridden. As an ideal matter, then, the different person theory seems quite appealing.

Of course, the different person theory is an "as if" theory: we proceed *as if* the mentally healthy and ill selves were different people. While philosophers have argued that it is possible for one body to house more than one person, such as in the case of multiple personality patients[25] or some severely brain damaged people, the overwhelming majority of mentally ill people are not literally different people from their mentally healthy counterparts. Among other things, they do not suffer sufficient memory discontinuities to be literally different people.[26]

The virtues of the different person theory would be clearest if mentally healthy and ill selves were literally different people. One would simply identify the right self—and different selves are easily distinguishable from each other—and let that self decide for the person. These virtues, however, also apply to the more plausible "as if" version of the theory. Under this version, the mentally ill and healthy selves are sufficiently unrelated to each other psychologically that it is as if they were different people. The mentally ill choice[27] is not really

---

25. I have argued elsewhere that the alter personalities of multiple personality patients may be sufficiently distinct in the required ways that we should construe them as literally distinct people. *See* Elyn R. Saks, *Multiple Personality Disorder and Criminal Responsibility*, 25 U.C. DAVIS L. REV. 383 (1992). With these patients, it is not that the mentally ill person is a different person from his healthy counterpart, but that the mentally ill person contains within himself different persons. Thus, the different person theory does not exactly fit the case of multiples, but may do so with slight modifications. Moreover, the idea behind the theory fits the case of multiples quite nicely: none of the personalities is competent to decide for the others. In fact, I have some qualms about finding multiples per se incompetent on this basis. *Id.* at 459-60 n.174. Nevertheless, this theory may be appealing in the case of multiples and, therefore, may have a very limited application in the incompetency arena.

26. The theories of personal identity that do not turn on bodily identity are known as "psychological" theories of personal identity. All of these require memory continuities for personal identity. For some psychological theories of personal identity, see, e.g., JOHN LOCKE, AN ESSAY CONCERNING HUMAN UNDERSTANDING 456-57 (Alexander C. Fraser ed., Dover Publications, Inc. 1959 (1290)); DEREK PARFIT, REASONS AND PERSONS (1984); PETER K. UNGER, IDENTITY, CONSCIOUSNESS AND VALUE (1990); David S. Oderberg, *Johnston on Human Beings*, 86 J. PHIL. 137 (1989). For some recent collections of articles on personal identity, see, e.g., THE IDENTITIES OF PERSONS (Amélie O. Rorty ed., 1976); PERSONAL IDENTITY (John Perry ed., 1975).

27. Of course, strictly speaking, choices are not mentally ill—people are. I use the

the choice of the healthy self, but is, in some sense, an "inauthentic" choice.

The idea that mentally healthy and ill selves are "as if" different people is quite plausible. Mental illness is often described in a way that lends itself to such a theory. Thus, a mentally ill person is said to suffer "ego alien" impulses and thoughts,[28] and, on recovery, to return to his "premorbid personality."[29] For instance, a shy, restrained, frugal person may become outgoing, seductive, and profligate under the sway of a manic attack. Her values and behavior may completely change, and she may be most distressed about her behavior when she returns to her normal self. Some patients do not even remember their experiences when mentally ill,[30] and when they do, may not recognize themselves in them. That some patients indeed speak as if their mentally ill and healthy selves are different people is quite striking. Thus, the idea that mentally ill and healthy selves are so different that they are as if different people seems eminently plausible.

B. *A Map of Related Theories*

1. THE NECESSITY FOR THE DIFFERENT PERSON NOTION

Before considering problems of the different person theory, I should like to draw a kind of conceptual map of some related theories. In particular, the different person theory is closely related to a theory known as the "thank you" theory: the caregiver is justified in overriding a patient's choice if the patient will later consent to the caregiver's intervention—will "thank" him for it.[31] Although the thank you the-

---

locution "mentally ill choices" for convenience, to avoid more cumbersome constructions such as "choices that are a product of mental illness" or "choices that mentally ill people make that are affected by their mental illness." On the connection between the different person theory and the product of mental illness theory, see *supra* text accompanying note 24 and *infra* text accompanying note 150.

28. For a description of the confusion and terror accompanying the strange experiences in the early stages of a psychotic break, see, e.g., SILVANO ARIETI, INTERPRETATION OF SCHIZOPHRENIA, ch. 22 (2d ed. 1974); HAROLD I. KAPLAN & BENJAMIN J. SADOCK, COMPREHENSIVE TEXTBOOK OF PSYCHIATRY IV 684-85 (4th ed. 1985).

29. *See, e.g.*, THOMAS P. DETRE & HENRY JARECKI, MODERN PSYCHIATRIC TREATMENT 116, 119 (1971); KAPLAN & SADOCK, *supra* note 28, at 685; LAWRENCE KOLB & H. KEITH BRODIE, MODERN CLINICAL PSYCHIATRY 459 (10th ed. 1982).

30. Not remembering unpleasant or traumatic events is called "psychogenic amnesia." *See, e.g.*, DSM-III-R, *supra* note 14, at 273-75; THE NEW HARVARD GUIDE TO PSYCHIATRY 100 (Armand M. Nicholi, Jr. ed., 1988). Psychiatric patients are sometimes unable to remember their psychotic episodes. Personal communication with John L. Young, M.D., Associate Clinical Professor of Psychiatry, Yale University and Unit Chief, Whiting Forensic Institute, Middletown, Conn. (May 4, 1989).

31. *See* ALAN A. STONE, MENTAL HEALTH LAW: A SYSTEM IN TRANSITION 66-70 (1975). It was Stone, I believe, who coined the term the "thank you" theory. Although many

ory is indeed related to the different person theory, the latter is far more plausible than the former. All the accoutrements of the "different self" notion, it turns out, are necessary for the thank you theory to be plausible—at least on competency grounds. Thus, advocates of the thank you theory must commit to a kind of different person notion for their theories to be adequate competency theories.

To explain why this is so, I would like to explore the related notion of an inauthentic choice. An inauthentic choice is not merely a choice that is inconsistent with one's settled personality. That would be to forbid all changes in values, even good changes that the person comes to cherish. An inauthentic choice is also not a choice one would later regret. People change their minds about earlier choices all the time, but they are not thereby committing to the inauthenticity of those choices. They may wish they had chosen otherwise without denying that it was *they* who chose. Indeed, if regret were enough for inauthenticity, changes in values would once again be forbidden, for many such changes will call into question decisions based on the earlier values.

Perhaps we should understand an inauthentic choice to be a choice that one would repudiate retrospectively. To repudiate a choice retrospectively is as if to say, "I was not myself when I made the choice—it was not, essentially, *my* choice." There are reasons to think that retrospective repudiation tends to track regret, and perhaps even inconsistency with one's settled personality.[32] But they are not the same thing, and the former seems necessary if a choice is to be deemed inauthentic.

As we shall see, the notion of "retrospective repudiation" is not entirely problem-free. For now it is enough to say that the notion is meant to call forth the idea that the person disowns the former choice as his—proclaims it an inauthentic choice. In turn, the notion of an

---

have come to think of the theory as justifying interventions because of the prospect of future consent, Stone's theory is, in fact, a fairly well-developed paternalistic theory that also requires some incompetency on the patient's part before it is invoked. Stone asserts that the "only justification for abrogating procedural safeguards is the provision of benefits which ameliorate human suffering." *Id.* at 18-19. Yet, he also would civilly commit a patient only if the patient's refusal of treatment were incompetent according to Stone's criteria. *Id.* at 68. The "thank you" theory I criticize in the text is the more general theory that the prospect of future consent justifies interventions, not Stone's particular version of the theory.

32. One's settled personality is partly a function of choices that one has reflected on, and made, in the past, and so, in a sense, one has reflected on choices that are in character with one's personality. Because people generally accept themselves, choices that are in character are likelier to be accepted in retrospect. Choices that are out of character, on the other hand, may be made without sufficient thought, and are likelier to be rejected as foreign.

inauthentic choice is meant to call forth the idea that the choice was in essence a different person's.

Someone may now argue that we do not need all these difficult notions to justify overriding a person's choice. All we need is the related—but easier to manage—notion of the prospect of future consent to the caregiver's decision or, put differently, of later rejection of one's earlier decision. (I use the term "rejection" as a generic term that is broader than regret or repudiation.) Surely, it will be argued, the prospect of such future consent to our own intervention is sufficient to justify it.

A simple hypothetical, however, will show why this is not so. Imagine a manic-depressive patient who is doing very well on lithium. When his psychiatrist retires, he consults a new psychiatrist who has rather different views about the patient's illness and its proper treatment. This psychiatrist is enamored of theories about mental illness that were more popular in the sixties.[33] He tells the patient that "mania" is simply an alternative mode of being and that, far from being an illness, it is an exalted state in which great creativity is possible. To be authentic, the patient must not try to escape the state through drugs, but must embrace it in all its glory.

The patient is somewhat reluctant to discontinue his lithium and return to the manic state because he remembers the destructiveness of that state. He also remembers a great deal of suffering, although during the state itself he seemed to feel great. The psychiatrist informs the patient that these memories may not be accurate. They may be the result of his former psychiatrist's indoctrinating him into mainstream American psychiatry—and perhaps of the drugs that were supposed to cure him, but really just suppressed him. If he will only give the new regime a chance, he will attain a state of happiness and well-being that others will envy.

The patient cannot bring himself to discontinue the drugs, but the doctor denies him access to them, gently explaining that this is for his own good. Within a short time the patient is in a full-blown manic state. He seems again to feel on top of the world and has absolutely

33. For some works setting forth such theories, see, e.g., RONALD D. LAING, THE DIVIDED SELF (1960); RONALD D. LAING, THE POLITICS OF EXPERIENCE (1967); THOMAS S. SZASZ, THE ETHICS OF PSYCHOANALYSIS (1965); SZASZ, *supra* note 17; THOMAS SZASZ, THE MYTH OF MENTAL ILLNESS: FOUNDATIONS OF A THEORY OF PERSONAL CONDUCT (1961). These views have been fairly well discredited by contemporary theorists. *See, e.g.*, MOORE, *supra* note 22, at 180-81; Z.J. Lipowski, *The Integrative Approach to Psychiatry*, 24 AUSTL. & N.Z. J. PSYCHIATRY 470 (1990); Michael Moore, *Some Myths About Mental Illness*, 18 INQUIRY 233 (1975); Steven Reiss, *A Critique of Thomas S. Szasz's "Myth of Mental Illness,"* 128 AM. J. PSYCHIATRY 1081 (1972).

no interest in resuming his medication. While the patient thinks he feels good, he is, in fact, in a good deal of pain, and his work and relationships suffer greatly. The doctor, on the other hand, is so committed to "authentic experience" that he is pleased with the outcome and very hopeful of great creativity to come. He also has the satisfaction of the patient's gratitude, for the patient heartily thanks the doctor again and again for showing him the way.

The example suggests that the prospect of future consent alone cannot justify overriding a patient's choice and imposing our own. If anyone believes that this actually may be a good case for respecting the future choice—that maybe the patient *is* better off being "authentic"—imagine now a situation in which no sensible person could subscribe to the doctor's theories or be happy with the result. Perhaps the doctor is an evil scientist who enjoys seeing people suffer and collecting data on their suffering for his forthcoming treatise. Or perhaps the patient begins wildly killing others in his manic state, all the while thanking the evil doctor.

The point of the example is not that the thank you theory is wanting because there may be misguided, incompetent, or even evil doctors. Even if there are misguided doctors (I doubt there are many truly evil ones), they are probably so few in number that the theory will misfire only a trivial amount of the time. A preference for freedom also misfires on occasion, but it is not a good argument against freedom that some people misuse it and cause significant harm.

The problem with the thank you theory, however, is not that it will sometimes produce undesirable consequences, but that it authorizes (indeed, requires) us to characterize an intervention in a way that is plainly wanting as a theoretical matter. The force of the counterexample is to show that not all interventions are actually justified by the prospect of future consent. In the same way, a free action that was morally despicable would be a compelling counterexample to some claim that all free actions are good. It may remain equally true in both cases that a practice of permitting interventions that receive future consent, or self-regarding free actions, produces the best results over the long run.[34]

What more is needed to justify an intervention for which the patient gives future consent? Two possibilities exist: we approve of the choice itself on independent grounds; or we have some autonomy-based warrant for giving special deference to the later choice. The

34. I do not think that a thank you theory in fact produces the best results in the long run. I think a competency theory similar to the law's standard does.

first possibility converts the thank you theory into pure paternalism:[35] we are justified in imposing our choice inasmuch as it is a *good* choice—and we get to decide what is good.[36] Because doctors' interventions are generally presumed to be in their patients' interests, the driving force behind the thank you theory may be such a paternalistic desire to do patients good.

If the thank you theory so understood is essentially paternalistic, permitting interventions in people's best interests, it nevertheless does take a bow to the concerns animating competency doctrine by valuing the patient's later consent. But it can only *rest* on such consent at the cost of permitting interventions like that of the mad psychiatrist without any warrant other than the prospect of the patient's momentary ratification of the caregiver's choice.

What about the second possibility—that we have some autonomy-based warrant for respecting the patient's later consent? On this view, the question is not whether some choice is good, but whether the patient will ratify it. If he will, we have evidence that *he*—or at least some part of him—thinks the choice is good. Yet later consent alone is not enough; we need the notion of a different person. The driving force behind this theory is a concern with the values underlying competency doctrine.

The reason simple consent is not enough is threefold. First, once again, future consent theories alone would justify bad choices merely because of the prospect of a moment of consent against a background of conflict; the manic patient's decision not to resume medication

35. By "pure paternalism" I mean intervening in a person's life over his protest solely on the ground that this will be better for him. When one does so because one deems the person incompetent to look after his own interests, one's act is still paternalistic, but is justified on autonomy-type grounds: decisions of autonomous people deserve respect, but this person lacks characteristics central to autonomy. A different person theory is perhaps less respectful of autonomy than a cognitive theory like the law's. The different person theory, in the name of the more representative future person's autonomy, allows us to override the choices of the current person, even though this person appears to exhibit the ordinary characteristics of autonomy himself. I have suggested that perhaps he lacks important kinds of knowledge about the other self that does impair his decisionmaking ability, but finding him incompetent does require us to choose between selves in a way that may inevitably implicate our own values. Thus, theories that defer to the current self's choices unless it is grossly impaired seem more respectful of autonomy than theories that allow the overriding of choices in the name of future autonomy. Nevertheless, the different person theory escapes the charge of pure paternalism in a way that thank you theories that do not become versions of the different person theory do not.

36. Proponents of autonomy/competency theories need not deny that there are objectively good and bad choices. But they are skeptical that we can often identify these reliably, especially for other people. Thus competency doctrine reserves value judgments about a person's own self-interest to the person himself, provided he is competent. The person himself is the acknowledged authority on his own best interests.

would suffice. If we require that the *authentic self* makes the choice, we are likelier to achieve a good outcome: generally, people's real selves make the best choices for them. In our example, the patient's true self *wants* the medication.[37]

A second reason we need the notion of a different self is that if a true self returns to ratify the choice, then implementing the choice reflects the values of that self and not simply our *own*. In other words, rather than simply choosing between two options on the basis of our own judgment about which is right, we are allowing the true self to choose. As we shall see, we are still forced to choose between the *selves* which will choose, so this rationale is not completely satisfying; but it has some force.

Finally, to deny legal respect to a person's choice simply because her choice was later rejected is not to identify any impairments in *her*. So long as she had access to the relevant variables, there is no reason to think her incompetent. Intact people make bad choices all the time. If the person choosing, however, is not the true self, she will lack access to the true self's values and needs, and *will* suffer impairments relevant to competency. In short, the notion of different selves is necessary to explain the patient's impairments and thus to cast the inauthentic choice theory as a *competency* theory.

To be justified on competency-type grounds, then, the thank you theory must become a different person theory. One might think that the notion of an inauthentic choice—as opposed to a choice simply later rejected—would get us as far. This may be true, but that is only because the notion of an inauthentic choice implicitly refers to a different self.[38] An authentic choice is a choice that reflects the true self's values, while an inauthentic choice betrays those values, although the inauthentic choice might itself be the appropriate choice for a *different* self.[39] Without some reference to a true self that the

37. Even when this does not occur, we have a robust theory about why such choices deserve deference. The true self knows the self best and cares about it most. And honoring that self's choice shows it a respect that is likely to further overall autonomy. The point is that, even if the true self makes a choice we think bad, we are justified in upholding it because this self has special warrant to choose. (Of course, the more usual competency doctrine justifies our upholding even bad choices because of *its* more robust theory of why such choices deserve deference.)

38. Indeed, I earlier suggested that the notion of retrospective repudiation refers to an inauthentic choice and that the notion of an inauthentic choice refers to the notion of different selves. *See supra* text accompanying note 32.

39. One may suggest that betrayal is a coherent notion precisely because we assume that there is *not* a different person with different values but the same person betraying his own values. Perhaps then we need, not a different person notion, but a way to distinguish change from betrayal. This suggestion has some force, but it does not undermine my position. First, it remains true that the person's choice does not reflect *his* true values. Thus, in a sense, it

choice betrays, the notion of an inauthentic choice cannot become meaningful in a way that differentiates it from any simply bad choice.

2. AN ALTERNATIVE TO THE DIFFERENT PERSON READING

One may now argue that there *is* a reasonable interpretation of the thank you theory that does not rely on the notion of different selves. This theory, moreover, is a competency theory, and not just a form of paternalism.

According to this interpretation of the thank you theory, the patient's later thanks provides convincing evidence of past incompetency. One could interpret the patient's rejection of her earlier choice as evidence that a "different person" made the choice. But one could also take it as powerful evidence that something quite different, involving traditional impairments of decisional ability, was occurring. By rejecting her earlier choice, the patient is saying that she deems herself to have been unable to choose—not in a state of mind deserving of respect. The thank you theory offers good evidence that something was fundamentally wrong with the patient's earlier decision, however inadequate language is to neatly detect and package it. Of course, if later thanks is good evidence of incompetency, the prospect of such later thanks justifies making an incompetency finding now.

On this interpretation, however, the thank you theory is wanting. When a patient rejects an earlier choice, she seems to be saying, "I was not myself when I made the choice." This could mean, again, that she feels she was then as if *someone else* when she made the choice—the different person theory. But it could also mean that she feels she was in an impaired state of mind when she made the choice—the thank you theory, on this interpretation.[40] The problem is that the theory offers no *account* of the kind of impairment the patient is suffering from—of what her incompetency consists. The theory is not a theory of incompetency at all on this view. Rather, it is a theory of evidence of incompetency.

The theory of evidence of incompetency is itself problematic.

---

reflects *someone else's* values—a different person's. To some extent, then, the choice is as if that of a different person, and it may not matter if a strictly different person did not make the choice. Perhaps more important, it *is* plainly possible for one person to betray another. In any case, I have given other reasons for thinking the different person notion is necessary. *See supra.*

40. Of course, it is also possible that the patient has *simply* changed her mind—prefers now the choice that the doctor has imposed—and does not actually *repudiate* the past choice. Again, regret and repudiation are different concepts, and for the different person theory to be workable on any of its versions, many of its concepts must be further refined. *See infra* part II.D.

First, why should we expect people to be able to judge when they were incompetent, when society itself lacks a good sense of what incompetency is? It is surely unreasonable to expect individual patients to be better at determining the nature of incompetency than theoreticians, and better at determining its presence than trained evaluators. (By contrast, it may be more reasonable to think people are somewhat authoritative on the question of when they were not themselves.[41]) Second, the thank you theory, under this view, is completely untestable. There is no independent notion of incompetency to determine if the patient's later thanks *is* good evidence of this incompetency.

Even if these problems attending the test as an evidentiary test did not exist, the fact that it is an evidentiary test, and not a standard of incompetency, means that my basic claim about the thank you theory stands: to serve as an adequate *standard* of incompetency, the thank you theory must in essence become a different person theory.

Having taken a strong stand that a thank you theory must rely on the different person notion, I concede that the theory nevertheless holds *some* attractions as a competency theory, even apart from its linkage with the different person theory. These attractions derive from the virtue of providing a powerful empirical basis for identifying incompetency, thus obviating the need to rest on theoretically-derived notions that may not be completely satisfying. The patient *herself* is saying, in effect, that she is confident that she was not in a proper mental condition to make the decision. As a *research strategy*, attempting to predict who will thank caregivers for their interventions, and studying what their prior reasoning had in common, may help us to better define and refine the concept of incompetency. But it does not provide a concept of incompetency itself.

C. *Problems with the Different Person Theory*

Although the different person theory holds considerable attrac-

41. I do not want to make too much of this suggestion. First, whether a past decisionmaker was a different self (or as if a different self) is perhaps a philosophical question on which philosophers, not individual decisionmakers, are expert. Second, for a person to have a strong sense that she was as if a different person when she made a past choice may be just a way of her saying "I regret, and wish to distance myself from, the choice." This is not the same as being as if a different person. It is more like a wish. It is a way for a person to say that she does not want to see herself as the sort of person to make a certain choice, rather than her not *being* that sort of person. Nevertheless, although people may have motives for not wishing to see themselves as they really are, they are generally in a better position to know themselves than other people are to know them. By contrast, they are in no better position than others to judge when their reasoning was so deficient as to be incompetent. We need a *theory* of competency for that—which this version of the different person theory does not provide.

tions, it also faces three significant difficulties. First, how do we select the self whose choices we will respect? Second, since I suggest in response to the first problem that we should focus on the *enduring* self, what do we do about the chronic patient, whose mentally ill choices then become authentic? Third, can we even apply the test reliably enough that it is workable in the case of acute patients?

1. WHICH SELF DO WE PREFER?

The different person theory faces a serious problem: how do we justify preferring the self that endorses our choice to the self that repudiates it?[42] Given the conflict the patient evidences, it is not enough to say that he thanks us—he also damns us.[43] We need a reason for preferring one of the selves. That the self that thanks us may manifest itself later in time is surely not enough.[44] Similarly, preferring the self that thanks us because we deem its choice right is simply paternalism in another guise.

We could prefer the healthy self solely on the ground that it *is* healthy. The ill self is by definition disordered, and disorders are generally agreed to be "bad." On at least some views, however, assumptions of disorder are themselves nothing but value judgments,[45] and so to prefer the healthy self is impermissibly to enshrine *our* values about which selves are worthy. The patient may disagree. While we may, once again, be *right*, competency doctrine acknowledges the individ-

42. In this context, the self that endorses our choice occurs later. But the same problem arises in other competency or quasi-competency contexts in a different way. For example, in the situation of self-binding, we tend to want to honor the choice of the *earlier* self over the current self. Why should we prefer that self over this? Indeed, the seriousness of this problem becomes especially clear in the context of medication-refusal, where at any given time, T1, there is one future self, S1, who will later thank the caregiver for giving treatment, and another future self, S2, who will later thank the caregiver for withholding treatment. Unless the caregiver has a reason for preferring S1 over S2 or S2 over S1, the mere prospect of a future self's thanks does not enlighten the caregiver as to which course—to treat or not—to pursue.

43. Although I continue to use the "thank you" language at times for convenience, I mean for it to incorporate the whole of the different person notion.

44. It is true that the law sometimes affixes legal significance to the particular *time* of an event or transaction. For example, saying "I decline" may at certain times have legal consequences (e.g., after another has made one an offer), and at other times, not (e.g., after one has already entered a binding contract). But a pure "later self" rule would run into obvious problems here, such as that of the self who mentally deteriorates over time. It is also artificial in a way that seems more objectionable in this context than in, for example, the contracts context.

45. *See, e.g.*, Joseph Livermore et al., *On the Justifications for Civil Commitment*, 117 U. PA. L. REV. 75, 80 (1968); Michael L. Perlin, *Unpacking the Myths: The Symbolism Mythology of Insanity Defense Jurisprudence*, 40 CASE W. RES. L. REV. 599 (1989-90); Bruce J. Winick, *The Right to Refuse Mental Health Treatment: A First Amendment Perspective*, 44 U. MIAMI L. REV. 1, 47-49 (1989).

ual decisionmaker herself as the final authority on matters of this kind.

If competency doctrine reserves value judgments to the greatest extent possible to the decisionmaker herself, choosing among different selves on the basis of our value judgments about which self is most worthy is at odds with the doctrine's commitment. We can respect competency doctrine's commitment only by anchoring choices for a person to the "true self" identified in some value-neutral way. This is not to suggest that one could *never* arrive at a view of people's true selves that would meet this desideratum[46]—just that it is difficult.

Perhaps the best we can do is to say that the real self, the authentic self, is the enduring self. Thus, when bouts with mental illness transform a person's values and preferences, we are right to listen to the healthy, enduring personality. Its thanks truly do justify our interventions.

I can see many problems with the notion of the true self as the enduring self, and I suggest it simply as something to think about. My main goal is to emphasize that we need *some* such value-neutral notion of true and false selves if the different person theory is to be a satisfactory competency theory.

2. THE CHRONIC PATIENT

If the authentic self should be understood as the enduring self,

46. For example, suppose it became possible, at birth, to take an imprint, so to speak, of a person's brain chemistry, which would identify people in the way that fingerprints now identify people. Perhaps the brain imprints would identify the person's true self. If the brain imprints did not change with age, but *did* change as a result of mental illness, it would be easy to identify when mental illness rendered a person a different person, and there would be clear authority to override the inauthentic person's choices. Medication would presumably restore the person to his real self—which the restoration of his unique brain print would correctly signal—and then he would be permitted to make his own decisions.

This hypothetical presents an attractive case for a different person theory precisely because it seems to allow us to identify the true self in a value-neutral way. We prefer the healthy self not because it is healthy—i.e., not disordered—but because it is the original self. Problems, however, immediately arise. What if the birth self is disordered, and giving it medication would render it healthy, thus creating a new brain imprint that it would carry with it for the rest of its life? Indeed, preference for originalness itself depends on a value judgment. Being first seems, in principle, no more reason for preferential treatment than being last or in the middle.

Another possibility would be to abandon the quest for the "true self," but attempt to ensure the "thanker's" neutrality by asking him *at the end of his life* whether he approves of the doctor's earlier intervention. If a person in his last moments, reviewing his life as a whole, expresses thanks for the doctor's intervention, there is some assurance that he is making a considered judgment unaffected by impulses of the moment. This is an interesting idea, but is immediately confronted by the problem of the chronic patient, see *infra* part II.C.2, or indeed of anyone whose old age distorts his reasoning or preferences in any way. Not everyone achieves saintly wisdom in the last moments of life.

the different person theory faces immediate problems as applied to the cases of some mentally ill people. Put most generally, the problem is that the classic model of a patient's recovering from mental illness and repudiating his choices does not fit all, or even most, cases of mental illness and its sufferers' choices. Some patients never return to their "premorbid" personalities. Many psychiatric illnesses, such as schizophrenia, tend to follow a chronic course.[47] Then, the person's mentally ill choices become authentic. The different person theory is clearly inappropriate for chronic patients' choices. No self will come into being to retrospectively repudiate these choices.[48]

The idea of a counter-factual healthy self as the basis of authenticity judgments—of a "healthy" person who would choose differently—is unacceptable for two reasons. First, as discussed previously, supporting the decisions of the hypothetical healthy self over his unhealthy counterpart requires a choice based on society's values about which self is most worthy. Once again, the patient may disagree. The different person theory honors competency doctrine's commitment to respecting people's values only if it anchors choices for a person to his "true self" identified in as value-neutral a way as possible. The counter-factual healthy self fails this requirement.

The second reason not to invoke the idea of a counterfactual healthy self that would choose otherwise is that that self *is* counterfactual. Even if the choice of the healthy self is obvious—which is rarely the case—that person no longer exists.[49] He is a bare construct. If he is only a construct, what does the patient care what he would have chosen? Imposing his values is not much more palatable than imposing the values of a neighbor or a friend—or our own values.[50] By

47. *See, e.g.*, KOLB & BRODIE, *supra* note 29, at 344, 380; ANDREW E. SLABY ET AL., HANDBOOK OF PSYCHIATRIC EMERGENCIES 388 (1986). An intriguing problem of chronicity is presented by the manic-depressive patient who alternates, without respite, between mania and depression. Who is the real self here?

48. One could dispense with the requirement of retrospective repudiation in the case of the chronically ill. But this is either to embrace pure paternalism—one overrides the choice because treatment is *good* for the patient—or it is to claim that it is permissible to invoke the idea of a counterfactual healthy self. Although no one *in fact* retrospectively repudiates the choice, it can be presumed that the healthy self *would have*. I believe that this move essentially eviscerates the different person theory as an autonomy theory. Thus, dispensing with the requirement of retrospective repudiation would be to abandon an autonomy-based different person theory as grounds for justifying interventions.

49. One may object to locutions of this kind because they suggest that the "healthy person" who "no longer exists" is a literally different person from the unhealthy person. In fact, there is one and the same person who was healthy and now is ill and will never again be restored to health. What I mean by the "healthy person no longer exists" is that this person is not now and will never again be healthy.

50. This is not to say that if the patient is grossly incompetent by any standard of incompetency and we must make *some* decision for him, we should not prefer to choose for

contrast, the patient may be more receptive to the idea of our imposing values that he himself, in a different aspect, embraces. Even if he is not receptive, and suffers fully the indignity and pain of having his choice overridden, there is compensation in the form of the preferred self's feeling gratified by our respect and support. Finally, the preferred self is spared having to live with the consequences of the choices of his brother self. Because "hypothetical selves" do not derive the pleasure of being respected, nor suffer the pain of living with the consequences of unwise choices, the merely "hypothetical benefits" they experience simply do not compensate for the insults to the repudiated self.

The upshot is that the different person theory does not justify a finding of incompetency in the case of the chronically ill. In the example of the manic patient whose doctor deprived him of medication, one would have to support his later decision to refuse, on this theory, if the doctor's actions rendered him permanently manic. That is, while the different person theory is superior to a future consent theory in the case of the acutely ill, where the enduring, healthy self makes the right choice, it is no improvement in the case of the chronically ill. The theory either fails as a theory of incompetency in the case of these patients because it does not apply to them at all, or it requires a finding of competency, and so may condemn them to an undeserved fate. Of course, one could always supplement the theory

---

him what he would have chosen if competent over what a neighbor would choose or (more likely) what we think best in his circumstances. Commentators often recommend such a "substituted judgment" standard in such cases. For example, in right to die cases, substituted judgment is desirable where a terminally ill patient is wholly beyond reason but has clearly expressed the wish to die in the past should he ever be in his current condition. *See, e.g.*, Rebecca Morgan & Barbara Harty-Golder, *Constitutional Development of Judicial Criteria in Right-To-Die Cases: From Brain Dead to Persistent Vegetative State*, 23 WAKE FOREST L. REV. 721, 743 (1988); John W. Parry, *A Unified Theory of Substituted Consent: Incompetent Patients' Right to Individualized Health Care Decision-making*, 11 MENTAL & PHYSICAL DISABILITY L. RPTR. 378, 381-84 (1987); Elizabeth Shaver, *Do Not Resuscitate: The Failure to Protect the Incompetent Patient's Right of Self-Determination*, 75 CORNELL L. REV. 218, 219 (1989).

The two situations are somewhat different, however. First, in the competency context, the incompetency finding is *based* on the hypothetical self's view. It is not that the hypothetical self's view directs our course once the patient is found incompetent on other grounds. Second, in the usual case, the past, competent self whose views are implemented in the substituted judgment context really did once exist. By contrast, the hypothetical self in the context of chronic patients' future consent is *solely* hypothetical.

In any event, imposing a past self's views on the now incompetent patient is a second-best solution. People would generally prefer to choose for themselves, rather than to have choices imposed on them, even when they have made the choices for themselves in the past. Nevertheless, their preference against such imposition of choices will perhaps be somewhat muted if they believe they themselves made these choices in the past and/or will return to appreciate them in the future.

with some kind of cognitive test, but the need to do so shows how limited the different person theory is as a general theory of incompetency.[51]

3. THE ACUTE PATIENT

The different person theory is not so obviously inapplicable to the case of the acutely ill. Still, it may not be appropriate for all those suffering from acute illnesses. Rather, it may be appropriate only for those suffering from first illnesses, who we hope will thank us, and for those suffering from recurrent bouts of illness who have indicated, for example, in treatment wills,[52] that their true selves *do* thank us.

Yet even thus restricted, the theory faces severe problems in practice. The classic model of mental illness underlying the different person theory still fails to fit all or even most cases. Even when their illness is temporary, recovered patients may well not repudiate their past choices and thank their caregivers.[53] All of these patients might be unduly grudging. But it seems far more plausible to suppose that many of their choices are *misidentified* as "mentally ill" choices, or, even though truly "mentally ill," are not experienced as the choices of a different person.

The essential problem is the difficulty of reliably determining which choices are "mentally ill" and will later be repudiated. Consider the analogy of a person who uncharacteristically yells at a friend while suffering from a headache. The yelling may have been the head-

51. The different person theory does not purport to identify specific impairments in patients. If it did, supplementation by a test that identifies other impairments might be expected and reasonable. (Consider the volitional impairment test, which is expressly designed to be supplemented by a cognitive test.) Rather, this test purports to be a general theory of competency. That it fails in the case of the chronically ill is therefore a serious shortcoming.

52. On treatment wills in the psychiatric context, see, e.g., Rebecca Dresser, *Life, Death and Incompetent Patients: Conceptual Infirmities and Hidden Values in the Law*, 28 ARIZ. L. REV. 373 (1986); John W. Parry, *The Court's Role in Decisionmaking Involving Incompetent Refusals of Life-Sustaining Care and Psychiatric Medications*, 14 MENTAL & PHYSICAL DISABILITY L. RPTR. 468 (1990). On treatment wills in general, see, e.g., George J. Alexander, *Time For a New Law on Health Care Advance Directives*, 42 HASTINGS L.J. 755 (1991); Ben A. Rich, *The Values History: A New Standard of Care*, 40 EMORY L.J. 1109, 1112-19 (1991).

53. The large number of cases in which patients contest involuntary detention and treatment bears witness to this fact. Consider also the existence of ex-patient and advocacy groups with names that suggest a decided preference against forced treatment. *See, e.g.*, BRUCE J. ENNIS & RICHARD D. EMERY, THE RIGHTS OF MENTAL PATIENTS 215-20 (1978). Arguably, the relevant data are how many patients who were *properly* treated remain ungrateful for their treatment. But there is no reason to think that all or even many of the above cases involved patients who received improper treatment. Indeed, if many are receiving improper treatment, there is all the more reason to permit them to refuse. More important, it seems highly relevant to ask how many patients who are simply *involuntarily* treated resent their treatment. These are the patients to whom a different person theory will apply.

ache "speaking." But it may also have represented the person's settled values: "I've felt this way for a long time now, but the right moment for expressing it has just come along." Alternatively, the person may have judged that a change in values was appropriate for the occasion: "He provoked the anger, and I was justified in yelling." Similarly, he may have decided that a permanent change in values was warranted: "I've finally seen what he is like, and no longer feel well-disposed toward him." Finally, the person may identify the choice as his own even if the headache brought it on: "I shouldn't have done that, but it was my choice and up to me to make."

These problems can be expressed in the language of competency doctrine itself. The doctrine presupposes that the decisionmaker himself is in the best position to identify his values and the choices that express his values. If others were equally well-placed to do so, there would be far less reason to reserve the choice to the decisionmaker alone. Other people are in a relatively poor position to make choices for a person in light of his values, because values often conflict, or are too general to determine a particular choice under the press of individual circumstances.[54]

When a person's values appear to be in flux, as in the different person context, predictions are even more precarious.[55] One must then determine not only what choice is most consistent with the person's past values, but also whether the person will choose to accept new values on reflection, either for that circumstance or permanently. In other words, to predict a person's considered choice of values is no easier than predicting the ordinary choices, based on his values, that competency doctrine shields from our interference.

Yet, perhaps we are overestimating the difficulty of establishing "mentally ill" choices and underestimating the extent to which the mentally ill really are as if different people. The major mental illnesses are characterized by distinct symptom clusters. For example, part of being depressed is to feel worthless, part of being manic is to

54. There will always be some easy cases. It is fairly certain that a person who has never hesitated to take aspirin in the past would want aspirin on an occasion in which a doctor predicts with high confidence that the aspirin will prevent an imminent and agonizing death. Most choices, of course, are not this easy. Competency doctrine recognizes the general difficulty of identifying the choices another would make, and is not sanguine about our ability to distinguish between cases in which one *can* choose for another and cases in which one cannot. The difficulties involved in choosing for others are exacerbated under conditions of change. *See infra* text accompanying note 55.

55. It is often very difficult to predict for *oneself* what choice one would make in either a new situation, such as significant illness, or in a situation in which one's values are changing. It is, of course, much harder to make a choice for someone *else* in those circumstances.

feel important, and part of being paranoid is to feel suspicious.[56] The presence of such a manifestation of an illness, when a person has the illness, is very likely to be a symptom of the illness. And to identify symptomatic beliefs and values is to say that the mental illness, and not the person himself, is speaking.[57] To the extent that there are such belief and value clusters which are characteristic of particular mental illnesses, the different person theory seems quite plausible.

Two factors, however, make this suggestion less compelling than it seems. First, particular symptom clusters do not as regularly characterize the major mental illnesses as this suggestion implies. Consider, for instance, the category of "mood incongruent" psychotic features found in the American Psychiatric Association's manual for diagnosis, DSM-III-R.[58] Indeed, the very fact that there is fairly low reliability in making specific diagnoses of the major mental illnesses, while there is reasonably high reliability in more general diagnostic categories, such as psychosis,[59] suggests that the symptoms of mental illness are fairly protean. Thus, our ability to identify a belief or desire as a symptom of the patient's mental illness is not as reliable as it might seem.[60]

This consideration is perhaps not overwhelming, because it sug-

---

56. *See, e.g.*, DSM-III-R, *supra* note 14, at 200, 214-15, 218-19.

57. It is implicit that the beliefs and values must be characteristic of a particular disorder, and not simply a sign of general distress. For example, people with illnesses of all varieties become more dependent. *See, e.g.*, KATZ, *supra* note 19, at 210. Hence, expressions of dependency needs are not a reliable sign that the person has a mental illness which has transformed him into a "different person" who is improperly making the choice at issue.

58. *See* DSM-III-R, *supra* note 14, at 402.

59. *See, e.g.*, Hermann O. Schmidt & Charles P. Fonda, *The Reliability of Psychiatric Diagnosis: A New Look*, 52 J. ABNORMAL & SOC. PSYCHOL. 262 (1956). Many studies have shown that reliability with the best diagnostic instruments is especially low in the case of specific disorders, such as specific kinds of personality disorder. *See, e.g.*, Roger K. Blashfield & Martha J. Breen, *Face Validity of the DSM-III-R Personality Disorders*, 146 AM. J. PSYCHIATRY 1575 (1989) (use of DSM-III-R for personality disorders leads to inaccuracies as a result of definitional overlap); Jeffrey H. Newcorn & James Strain, *Adjustment Disorders in Children and Adolescents*, 31 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 318 (1992) (low reliability for adjustment disorders in children and adolescents); Gregg Gorton & Salman Akhtar, *The Literature on Personality Disorders, 1985-88: Trends, Issues, and Controversies*, 41 HOSP. & COMMUNITY PSYCHIATRY 39 (1990); John S. Werry, *Overanxious Disorder: A Review of its Taxonomic Properties*, 30 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 533 (1991) (variable reliability for overanxious disorder).

60. I do not want to overstate this point. There may be some mental illnesses, such as major depression, for which reliability is fairly high. Moreover, reliability seems to be improving with the advent of instruments such as the DSM-III-R. Wolfgang Hiller et al., *Development of Diagnostic Checklists for Use in Routine Clinical Care: A Guideline Designed to Assess DSM-III-R Diagnoses*, 47 ARCHIVES GEN. PSYCHIATRY 782, 782 (1990); *see also* Brian McConville & Paule Steichen-asch, *On the Usefulness of the DSM-III-R Versus the DSM-III for Child Psychiatrists*, 35 CAN. J. PSYCHIATRY 367 (1990); Fred R. Volkmar et al., *DSM-III and DSM-III-R Diagnoses of Autism*, 145 AM. J. PSYCHIATRY 1404 (1988). Even

gests only that one may not be able to identify which diagnostic category a person's symptoms represent, not whether they represent illness at all. It does, however, cast some doubt on the reliability of our judgments that someone's decision is the mental illness "speaking." The point is that some "free-floating" symptoms—symptoms that do not regularly characterize a particular mental illness—may not be symptoms at all, but simply manifestations of an individual person's idiosyncratic response to the world.

The second consideration is more troubling. The presence of even a characteristic manifestation of a particular mental illness may also be characteristic of a person's normal personality, and hence an expression of his characteristic personality style. Indeed, many people develop mental illnesses consistent with their premorbid personalities. For example, people with paranoid personalities go on to develop paranoid psychoses.[61] Thus, the ability to identify a belief or value as that of the "different," mentally ill person, rather than the person's characteristic self, is greatly reduced. Consequently, identifying "mentally ill" choices seems quite as difficult as I have suggested. If this is so, the different person theory may be so difficult to apply that it should be rejected even as applied to the acutely ill.

D. *A Suggested Response if Evaluators Could Reliably Predict Patients' Thanks*

For the theoretical reasons discussed above, the different person theory is likely to be extremely difficult to apply. Competency doctrine itself is premised on an inability to choose for another, and the usual case does not present the complication of the presence of change. Expectations based on these theoretical reasons, however, may prove wrong. What would happen if controlled studies demonstrated that psychiatrists *could* predict—perhaps as a result of a deepening understanding of the symptoms of mental illness—when recovered patients would thank their caregivers? Would the different person theory, operationalized in terms of patients' thanks, then be adequate?[62]

---

with categories such as major depression, however, diagnosis may be confounded by such features as mood-incongruent delusions.

61. *See, e.g.*, HAROLD I. KAPLAN & BENJAMIN J. SADOCK, SYNOPSIS OF PSYCHIATRY 527-28 (6th ed. 1991); KOLB & BRODIE, *supra* note 29, at 594.

62. Although I have argued that we need the entire different person notion to override a patient's choice in competency grounds on the basis of this theory, in practice, focusing on the person's choice—at least in the first instance—may make sense. The different person theory, on this view, rules out mentally ill choices that are later regretted. Providing separate criteria for mentally ill and healthy selves may take us no further than looking at the choices themselves, which are probably good proxies for the selves which make them. For the

Even in this unlikely event, more is needed before the different person theory is acceptable as an adequate theory of competency. First, simple data on who says "thank you" would be contaminated by the phenomenon of patients wishing to please their caregivers, who retain a good deal of power over them and whom the patients might wish to please for a wide variety of reasons. Additionally, some patients may be idiosyncratically grudging of thanks. The point is that the simple expression or withholding of thanks may not be a reliable indicator of which patients are truly grateful or resentful. Indeed, the patients may not be dissembling. They may not know themselves what they truly feel.

Second, the mere expression or withholding of thanks indicates, at most, regret toward or approval of the overridden choice. Operationalizing the different person standard as a thank you standard is clearly a matter of second-best. What we really want is an indication that the patient actually *repudiates* his earlier choice, declaring the choice as not truly *his*. In that case, the patient is saying that he was as if a different person when he made the earlier choice. Now the problem here is that we have no reliable way of distinguishing mere regret from active repudiation. More fundamentally, we have not adequately analyzed the concept of repudiation—we do not even know exactly what we are looking for. What we need are a better-analyzed notion of repudiation and an operationalized standard of when such repudiation is indeed occurring.

Even if we could overcome these two core problems, the different person theory would face additional difficulties. These difficulties derive from an initial willingness to gloss the standard in terms of a patient's repudiation of past choices, rather than insisting on a more careful explication and operationalization of the concept of a different self. As noted earlier, when people repudiate past choices they *may* merely be expressing regret and distancing themselves from the choice, rather than identifying themselves as truly altered when they made the choice.[63] Perhaps inauthentic choices are made by inauthentic choicemakers, but without a careful analysis of different selves, we shall never know. Even if these problems could be solved, the different person theory would face real problems in justifying its preference for healthy selves over unhealthy selves.

Perhaps this pessimism is too harsh. With operational definitions of repudiation and criteria for establishing its genuineness, the differ-

---

different person theory to be fully adequate, however, one would want to do the conceptual work linking these choices to a robust notion of different selves.

63. *See supra* note 41.

ent person theory may be testable, at least in its "retrospective repudiation" form. And if evaluators could predict the occurrence of genuine repudiation, *and* such repudiation could be linked with philosophically coherent criteria for different selves, we might have a workable, adequate standard of competency. Such predictions and linkages will likely be unreliable, and such conceptual clarification a long way off, but workers in the field might prove me wrong. Thus, although the different person theory is inadequate today, with future conceptual and empirical work, it might become more attractive.

E. *Conclusion*

The different person theory is cast in language with intoxicating implications. If mentally ill and healthy selves truly were different persons, one would need only to identify who was making a choice to know whether to honor it. The different self's choice would be no more likely to express the person's true values, and would have no more claim for his reflection, than if a neighbor or a friend had made it. In essence, there would be no attempt to second-guess a person's values if different persons were involved in the choice. One would merely identify persons and let them choose their *own* values. The language of the different person theory also is intoxicating in that different people in ordinary life are easy to distinguish from each other, and thus applying the theory promises to be simple and straightforward.

Of course, the "different person" language is only metaphorical. Mentally ill and healthy selves are not literally different people, so these virtues of the different person theory are less than fully realized. Moreover, operationalized as a retrospective repudiation theory, the different person theory utterly fails as applied to the chronically ill, and it is too unreliable as applied to the acutely ill.

The future, however, may prove this assessment wrong. If theorists can give content to the notion of retrospective repudiation as distinct from regret; if clinicians can operationalize the notion so that it is apparent when someone is actually and genuinely retrospectively repudiating a prior choice; and if evaluators can predict when such retrospective repudiation will occur, we will be well on our way to a practicable competency standard, at least as applied to the acutely ill. If theorists and practitioners can then link the concept of an inauthentic choice to an adequate concept of a different self lying behind the choice, and can overcome the problem of preferring some selves over others, this competency theory would become quite attractive. Until

that time, should it ever arise, the different person theory faces such acute problems that we should reject it.

III. THE VOLITIONAL IMPAIRMENT TEST

A. *Introduction*

The volitional impairment standard, known historically in the criminal law as the "irresistible impulse" standard, complements a cognitive standard. The resulting standard finds incompetent, in addition to those with cognitive impairments, those who are so overwhelmed by their mental states that they effectively lack the ability to exercise choice.[64] The idea that a person should be blameless for conduct if he effectively lacks choice seems immediately appealing. The analogous conclusion in the competency context seems equally appealing: if a person cannot help making a particular treatment decision, why respect that decision? The decision is not a product of his free exercise of judgment and will, but is, rather, forced on him. If the choice is, in effect, not a reflection of his personhood, failure to honor his choice does not deny him respect as a person. Indeed, to hold him to the consequences of his choice, which may be extremely destructive, seems to be misguided, if not inhumane.

As an ideal matter, then, it seems desirable to find incompetent people who have no control over their choices. But to speak simply of a person's not having control over his choices is very general—and rather conclusory. It would no doubt be an important contribution to scholarship in this area to describe different clinical entities in which

64. On the nature of the volitional impairment test in the criminal context, along with criticisms and responses to those criticisms, see ABRAHAM S. GOLDSTEIN, THE INSANITY DEFENSE 67-79 (1967); Richard J. Bonnie, *The Moral Basis of the Insanity Defense*, A.B.A. J. Feb. 1983, at 194; Joseph E. DiGenova & Victoria Toensing, *The Federal Insanity Defense: A Time for Change in the Post-Hinckley Era*, 24 S. TEX. L.J. 721 (1983); Jodie English, *The Light Between Twilight and Dusk: Federal Criminal Law and the Volitional Insanity Defense*, 40 HASTINGS L.J. 1 (1988); Kathryn Fritz, *The Proposed Federal Insanity Defense: Should the Quality of Mercy Suffer for the Sake of Safety?*, 22 AM. CRIM. L. REV. 49 (1984); Jerome Hall, *Psychiatry and Criminal Responsibility*, 65 YALE L.J. 761 (1956); Insanity Defense Work Group, *American Psychiatric Association Statement on the Insanity Defense*, 140 AM. J. PSYCHIATRY 681 (1983); Chet Kaufman, *Should Florida Follow the Federal Insanity Defense?*, 15 FLA. ST. U. L. REV. 793 (1987); Edwin R. Keedy, *Irresistible Impulse as a Defense in the Criminal Law*, 100 U. PA. L. REV. 956 (1952); Joseph M. Livermore & Paul E. Meehl, *The Virtues of* M'Naghten, 51 MINN. L. REV. 789 (1967); Henry T. Miller, *Recent Changes in Criminal Law: The Federal Insanity Defense*, 46 LA. L. REV. 337 (1985); Judith A. Morse & Gregory K. Thoreson, *Criminal Law*—United States v. Lyons: *Abolishing the Volitional Prong of the Insanity Defense*, 60 NOTRE DAME L. REV. 177 (1984); Harry J. Philips, Jr., *The Insanity Defense: Should Louisiana Change the Rules?*, 44 LA. L. REV. 165 (1983); Francis V. Raab, *A Moralist Looks at the* Durham *and* M'Naghten *Rules*, 46 MINN. L. REV. 327 (1961); Sobeloff, *supra* note 9; John B. Waite, *Irresistible Impulse and Criminal Liability*, 23 MICH. L. REV. 443 (1925).

behavioral controls are impaired and the ways in which they are impaired. We would then have a richer, more nuanced understanding of why we should not legally respect the decisions of those who are volitionally impaired in these different ways.

A number of different clinical entities seem candidates for a study of this kind: impulse control disorders,[65] affective disorders,[66] addiction disorders,[67] and certain personality disorders.[68] Moreover, these different disorders seem to have quite different impacts on behavioral controls. For example, people who suffer impulse control disorders may either experience extremely strong impulses or have very weak ego strength. By contrast, mood-disordered people may face choices that are too hard; their mood may create an optimism or pessimism that suffuses their perception of alternatives (without producing actual delusions) in a way that constrains choices.

The impediment to free choice seems quite different in the two cases. In the first, internal forces buffet the person about in the way that external forces do when one person uses another's body to his own ends. In the second, the person faces unacceptable alternatives, in the way that a person does when another person points a gun to his head and demands money. The first case produces a kind of physical necessity to act, while the second, a kind of moral necessity. In any event, carefully considering the different kinds of disorders presenting volitional impairments would no doubt further understanding and evaluation of the defense.

Although this project is worth undertaking, this Article rests content with our inchoate understanding of volitional impairments, and asks whether, given the paradigm cases, there are reasons specific to the treatment competency context for rejecting a volitional standard.

B. *Standard Criticisms of the Test in the Criminal Law*

Although the irresistible impulse test has a long history in the criminal law's insanity doctrine,[69] for unclear reasons, it has never had equal currency in the civil law's competency doctrines. This is somewhat surprising, because compelled choices, as we have seen, do not seem deserving of respect. In the competency context, a compul-

65. *See, e.g.*, DSM-III-R, *supra* note 14, at 321-28.

66. *Id.* at 213.

67. *Id.* at 245.

68. *Id.* at 335-59.

69. For some early cases, see Parsons v. State, 2 So. 854 (Ala. 1886); Commonwealth v. Rogers, 7 Mass. 500 (1 Met.) (1844); Commonwealth v. Mosler, 4 Pa. 264 (1846).

sive pill-popper or doctor-disobeyer, if there are such patients, would seem incapable of free choice. Yet, although the irresistible impulse doctrine does not have a firm place in civil law, it is not entirely absent there either,[70] and commentators have approved of its use in the civil law.[71]

The criticisms of the irresistible impulse doctrine in the criminal law have taken various forms. The first two suggest that the test, as stated, is too narrow. The third argues that it is unnecessary, and the fourth contends that it is too difficult to apply. These criticisms are on different levels conceptually. The first two object solely to the *form* of the test, while the second two object to the test altogether. The criticisms are also of unequal force, and more or less difficult to accommodate when persuasive.

The first criticism is that the notion of an irresistible impulse is too narrow because it seems to rule out inner compulsions that do not arise suddenly.[72] It seems as well to admit that persistent cravings or overwhelming moods or emotional states can impair people as surely as, and in a similar way to, sudden impulses. Thus, the notion of an irresistible impulse may also be misleading in suggesting a kind of internal, mental force, impelling action in the way that an external physical force might.[73] Although one might experience a strong impulse in this way, one might also experience an overwhelming emo-

70. For example, § 15 of the Restatement of Contracts includes a volitional prong: "(1) A person incurs only voidable contractual duties by entering into a transaction if by reason of mental illness or defect . . . (b) he is unable to act in a reasonable manner in relation to the transaction and the other party has reason to know of his condition." RESTATEMENT (SECOND) OF CONTRACTS § 15 (1981). One notable use of the volitional impairment doctrine is New York's law of contractual capacity. *See, e.g.*, Ortelere v. Teacher's Retirement Bd., 250 N.E.2d 460 (N.Y. 1969); Faber v. Sweet Style Mfg., 242 N.Y.S.2d 763 (N.Y. Sup. Ct. 1963); In re Estate of Gebauer, 361 N.Y.S.2d 539 (N.Y. Sup. Ct. 1974). At least one other jurisdictions has followed suit. *See* Nohra v. Evans, 509 S.W.2d 648 (Tex. Civ. App. 1974) (reversible error not to instruct that "mental capacity" referred to a person's ability to exercise his will in addition to unimpaired cognition); *see also* Krasner v. Berk, 319 N.E.2d 897, 900-01 (Mass. 1974) (citing *Ortelere*; plaintiff found incompetent on cognitive grounds); Gore v. Gadd, 522 P.2d 212, 214 (Or. 1974) (court held that "[a]ssuming, without deciding, that the [*Ortelere*] test of competency should be extended to include affective disorder, we find the evidence insufficient to establish that plaintiff was incompetent"). At least one court discussed the *Faber-Ortelere* approach but interpreted its legislature as mandating a cognitive approach only. *See* Smalley v. Baker, 69 Cal. Rptr. 521 (Cal. Ct. App. 1968).

71. *See, e.g.*, Meiklejohn, *supra* note 12; Weihofen, *supra* note 12.

72. *See, e.g.*, United States v. Freeman, 357 F.2d 606, 620-21 (2d Cir. 1966); United States v. Durham, 214 F.2d 862, 874 (D.C. Cir. 1954); MODEL PENAL CODE § 4.01 cmt. 3 (Tent. Draft No. 4, 1955); ROYAL COMMISSION ON CAPITAL PUNISHMENT 1949-53 REPORT (¶ 314, at 110) (1953); RUDOLPH J. GERBER, THE INSANITY DEFENSE 39 (1984); Sobeloff, *supra* note 9.

73. *See, e.g.*, Dejarnette v. Commonwealth, 75 Va. 867, 878-79 (1881); English, *supra* note 64, at 17-18.

tion or mood as demanding a particular choice, in the sense of making all other alternatives seem unacceptable. For example, profound hopelessness might make efforts to spare others from imagined mental suffering seem the only conceivable course, just as extreme elation might make sharing one's bounty with others.

Both of these concerns about the irresistible impulse doctrine seem valid. It does not make sense to legally disable those with sudden impulses, but not those with other, longer-standing compulsions. Similarly, it does not make sense to find incompetent or insane only those whose mental states impel in the way that an external force might; other kinds of inner compulsion may be equally disabling. The criminal law has adequately responded to these criticisms by reformulating the doctrine in terms that do not suggest such a narrow focus. Thus, courts and commentators now read the test to require an "inability to conform to the law,"[74] and refer to it as the "volitional insanity defense."[75]

Another criticism of the volitional test—that requiring total impairment, or complete irresistibility, may be too harsh[76]—meets a similar fate. In addition, requiring truly irresistible impulses may place an imposing demand on evaluators in light of the seemingly impossible task of reliably distinguishing truly irresistible impulses from impulses that were simply extremely hard to resist.[77] The criticism seems valid, and the law's response satisfactory. The law has eliminated the requirement of complete inability to control. Thus, those lacking simply "substantial capacity" to control their behavior now receive the benefit of the defense.[78]

The third argument against a volitional test is less persuasive.

---

74. *See, e.g.*, U.S. v. Torniero, 735 F.2d 725, 729 (2d Cir. 1984), *cert. denied*, 469 U.S. 1110 (1985); *see* MODEL PENAL CODE § 4.01(1) (Proposed Official Draft 1962).

75. *See, e.g.*, English, *supra* note 64; Christopher Slobogin, *The Guilty But Mentally Ill Verdict: An Idea Whose Time Should Not Have Come*, 53 GEO. WASH. L. REV. 494 (1985).

In the sections following this historical review, I avoid the term "irresistible impulse" and speak of strong, overpowering, compulsive, or irresistible wishes or desires, of hard choices, of an inability to control one's behavior, and other such variants. Unlike impulses, wishes and desires need not be sudden. I recognize that sometimes one is unable to modulate one's behavior for reasons other than being under the sway of strong desires, but, at times, it is the case of being under the sway of strong desire that is precisely my focus.

76. *See, e.g.*, WAYNE R. LAFAVE & AUSTIN W. SCOTT, CRIMINAL LAW § 37, at 284-85 (1972); ROYAL COMMISSION ON CAPITAL PUNISHMENT, *supra* note 72, at 94-95, paragraphs 264-65); GERBER, *supra* note 72, at 38-39.

77. *See, e.g.*, American Psychiatric Association, *Statement on the Insanity Defense*, *in* ISSUES IN FORENSIC PSYCHIATRY 11 (1984); Bonnie, *supra* note 64; English, *supra* note 64.

78. *See* MODEL PENAL CODE § 4.01. Although in subsequent sections of this Article I continue to speak of "irresistible" wishes or desires, I wish to be understood not to require *complete* irresistibility. I continue to use the term for the sake of convenience.

This argument has two steps: first, only those who are so impaired that they suffer from severe cognitive distortions, and thus pass a cognitive test, cannot control their behavior;[79] and second, because a volitional test is therefore unnecessary, it should be abandoned.[80] The first step argues that choices are never too hard unless the person's cognitive functioning is so impaired that she suffers from patent delusions and thus passes a cognitive test. According to this view, reason is at the helm, so that one remains essentially free unless one's reason is impaired. One's choices are sufficiently constrained only if one's mood disorder is so profound that it overwhelms reason with irrational thoughts, or if the irrational thoughts themselves constrain the choices.

But why should we think that only those who are either disorganized or suffering from delusions that constrain their choices cannot control their behavior? It seems possible that a person may be cognitively intact but simply lack the ego strength to resist an overpowering desire. Commentators in the insanity context have discredited the view that only the cognitively impaired cannot control their behavior, claiming that it relies on an outmoded psychology that sees delusions as *the* symptom of mental disorder and equates mental illness with cognitive disarray.[81] Psychiatrists and psychologists today appreciate

---

79. *See, e.g.*, INSANITY DEFENSE IN FEDERAL COURTS: HEARINGS ON H.R. 6783 BEFORE THE SUBCOMM. ON CRIMINAL JUSTICE OF THE HOUSE COMM. ON THE JUDICIARY, 97th Cong., 2d Sess. 231-32 (1982) (statement of Stephen Morse, Professor, University of Southern California Law Center, on behalf of the Association for the Advancement of Psychology).

80. *See, e.g.*, Insanity Defense Work Group, *supra* note 64, at 685; Judith Morse & Gregory K. Thoreson, United States v. Lyons: *Abolishing the Volitional Prong of the Insanity Defense*, 60 NOTRE DAME L. REV. 177, 185-87 (1984).

Richard Bonnie makes a related argument in response to the claim of some commentators that we need a volitional test to find incompetent some mood-disordered people who are not sufficiently impaired cognitively to meet a cognitive test. *See* Bonnie, *supra* note 64, at 197. For example, a severely depressed mother may kill her child so that the child may avoid a life of suffering. This mother would fail some versions of the *M'Naughten* test if she knew that murder was illegal. Bonnie correctly responds that these commentators advocating the volitional test are concerned about cognitive impairment, and the obvious move is to broaden the cognitive test by using, for example, an "appreciation" standard to capture these cases. Bonnie's conclusion is the same as the argument I address in the text: there is no need for a volitional test. But the structure and rationale, as well as the persuasiveness, of the arguments are very different.

I have some concerns about applying a broad cognitive standard rather than a volitional standard to cases such as the depressed mother. I have argued previously that a narrow cognitive test is more appropriate in the treatment competency context. *See* Saks, *supra* note 6. Were I persuaded that the depressed mother should be excused from criminal sanctions, I might prefer to apply a volitional test to her case to avoid confusion about the scope of cognitive standards in contexts such as treatment competency.

81. *See, e.g.*, Durham v. United States, 214 F.2d 862, 870-74 (D.C. Cir. 1954) (tracing the

that a breakdown of reason does not play the leading role in mental illness, and that pure mood disorders can be as disabling as any thought disorder.[82] And so modern science sustains the common sense intuition that mood-disordered people may be compelled to act in ways consistent with their mood, quite apart from any delusions they may have.

The mute, motionless depressive is a case in point. This person suffers a depressed mood as profound as depression can get, and his "choice" not to move or speak seems classically "hard." Yet there is no evidence that such depressives always suffer from delusions. To insist that they do, and that the somewhat less depressed person who can tell us he has no delusions does not face hard enough choices, may be to cling to the outmoded psychology of earlier generations without clear justification.

The second step of this argument is also suspect. Even if it were true that we should excuse only the volitionally impaired who also meet a cognitive test, that is not an argument against a volitional test. The test is perhaps unnecessary in such cases, but there may be symbolic value in having an additional test, or a test that focuses on impairments in will. There may be real practical advantages as well. Suppose it is relatively easy to prove the presence of certain volitional impairments that tend to accompany the required cognitive impairments but extremely onerous to prove the latter impairments themselves. In this case, providing a properly narrowed volitional test would ease the defendant's burden of proof. It would establish, in effect, an irrebuttable presumption that those who meet the volitional test also meet the cognitive. If most defendants would meet the cognitive test anyway, if they could only prove it, allowing defendants an additional means to establish their innocence would tend to serve justice without compromising any other values. Does not justice then *require* a volitional insanity defense?

The answer is to be found in a missing premise: evaluators will make too many mistakes using a volitional test, with the result that people who do not in fact meet the test will appear to meet it. The third criticism, then, relies on a premise that is the core of the fourth:

---

decline of the "right-and-wrong" test of criminal responsibility); GERBER, *supra* note 72, at 30-33; SHELDON GLUECK, LAW AND PSYCHIATRY 47-48 (1962).

82. *See, e.g.*, JOHN A. TALBOTT ET AL., THE AMERICAN PSYCHIATRIC PRESS, TEXTBOOK OF PSYCHIATRY 403 (1988) ("Mood disorders span a wide spectrum of conditions, ranging from reactions to loss and other negative life experiences to severe, recurrent, debilitating illnesses"). Outcome in schizophrenia, the most severe of the thought disorders, is generally poor; nevertheless, a significant proportion of patients are symptom-free on long-term follow-up. *Id.* at 369.

we should reject a volitional test because of the risk of "moral mistakes."[83] This fear of moral mistakes animates commentators espousing a wide variety of views on how much volitional impairment, and accompanying cognitive disarray, the test should require.

The fear of moral mistakes seems well-founded. Psychiatrists themselves lend it support. Thus, speaking for the American Psychiatric Association, a respected panel of psychiatrists has said that "[t]he line between an irresistible impulse and an impulse not resisted is probably no sharper than that between twilight and dusk."[84] Requiring merely "substantial incapacity to resist" does not help much. The line between such "substantial incapacity" and slightly less incapacity is no clearer.

Indeed, it perhaps understates the problem to speak of a line-drawing problem. The problem is not that once psychiatrists have gauged the strength of impulses, they do not know where to locate the appropriate cut-off point. Rather, the problem is that psychiatrists have no good way to gauge the strength of impulses or desires at all.

It is true that psychiatrists can no more peer into a person's mind to gain access to her thoughts than to gain access to her desires. Thus, to some extent, they must rely on the person's account of her state of mind in both cases, which may or may not be reliable. Yet psychiatrists need not rely *totally* on such self-reports, and then identifying the content of a person's thoughts seems far easier than identifying the strength of her desires.

Thus, the content of a person's thoughts may be evident in her behavior. Imagine a patient who cowers in the corner whenever her psychiatrist enters the room. Although a person's desires may also be evident in her behavior—for example, a patient who tries to go AWOL seems clearly to wish to leave the hospital—the *strength* of those desires is surely not evident. Psychiatrists simply observe a patient acting in a certain way. Although they can thus presume that on some level she desires to so act, they can infer nothing from the action itself about the strength of her desires.

I do not want to overstate this point. There may be good evidence, at times, that a person is overcome by powerful desires. For example, the person's general behavior may show signs of being impulse-ridden—the behavior may be so obviously pleasure-seeking (in the short term) in such a disorganized and haphazard way that we

83. *See, e.g.*, Bonnie, *supra* note 64, at 195-96; Fritz, *supra* note 64, at 64; Miller, *supra* note 64, at 344. *But see, e.g.*, English, *supra* note 64, at 49-50; Morse & Thoreson, *supra* note 64, at 185.

84. *See* Insanity Defense Work Group, *supra* note 64, at 685.

presume higher-level governance to be lacking. Even a person who is not out of control in this way may show behavioral signs of struggle against an unwanted impulse. Nevertheless, the evidence of strong desires underlying a particular action seems less available, in general, than the evidence for disordered thinking.

Finally, even when there is every reason to trust the patient's veracity, her self-reports about the strength of her desires are likely to be less reliable than her self-reports about the content of her beliefs. Reporting the content of a belief requires no comparative judgments, while reporting the strength of a desire obviously does. But quantitative judgments of this kind, to be informative to others, require a common scale, which is arguably lacking. Although a person can confidently say that her desire today for a candy bar is strong *for her*, she has no idea if it is strong compared to others' desires. In the same way, a person cannot know whether she feels pain more or less than others. Unless the competency evaluator has a sense that the patient is using the same scale as the evaluator—indeed, unless the evaluator has a sense that her own scale is accurate—she must be wary of competency judgments based on the patient's self-reports in this regard. In short, without some shared, public scale, self-reports of the strength of one's desires simply are a poor basis for competency judgments.

Thus, the charge that evaluators cannot reliably apply a volitional standard appears to have merit. Many commentators and lawmakers have therefore recommended abandoning the volitional standard in the criminal law,[85] as many jurisdictions have done.[86]

---

85. For example, the American Bar Association and the American Psychiatric Association suggest that the irresistible impulse test is so problematic it should not be used. *See* AMERICAN BAR ASSOCIATION STANDARDS FOR CRIMINAL JUSTICE, PROPOSED CRIMINAL JUSTICE MENTAL HEALTH STANDARDS 323, 329-32 (Official Draft 1984); Insanity Defense Work Group, *supra* note 64, at 685; *cf.* Bonnie, *supra* note 64, at 196; Livermore & Meehl, *supra* note 64, at 833; Stephen J. Morse, *Crazy Behavior, Morals and Science: An Analysis of Mental Health Law*, 51 S. CAL. L. REV. 527, 588-89 (1978); Barbara Wootton, *The Insanity Defense*, 77 YALE L.J. 1019, 1027 (1968) (book review).

86. For example, the federal "not guilty by reason of insanity" statute largely adopted the Model Penal Code test but deleted its volitional prong. *Compare* 18 U.S.C. § 17(a) (1988) *with* MODEL PENAL CODE § 4.01(1) (Proposed Official Draft 1962). *See generally* Robert F. Schopp, Note, *Returning to* M'Naghten *to Avoid Moral Mistakes: One Step Forward, or Two Steps Backward for the Insanity Defense?*, 30 ARIZ. L. REV. 135 (1988). In the 1970s, more than half the states had adopted the Model Penal Code test, including its volitional prong. Christopher Slobogin, Criminal Law and the Mentally Disabled: Reconsidering Excuse Doctrine 6 (Apr. 1, 1992) (unpublished manuscript, on file with the author). Other states used the traditional "irresistible impulse" test. Today, only approximately 20 states still use some form of volitional test. For examples of jurisdictions that previously used some form of volitional standard but have now abandoned such a standard, see United States v. Rosenheimer, 807 F.2d 107, 111 (7th Cir. 1986); United States v. Lyons, 731 F.2d 243, 248

That recommendation does not seem obvious to me, for reasons that will become clear. But while one should perhaps retain a volitional test in the criminal context, we should reject it in the treatment context: concerns about reliable application combine with other considerations to suggest that a volitional standard is not appropriate in that context.

C. *The Legitimacy of Importing a Test Appropriate to Criminal Responsibility to the Arena of Treatment Competency*

A standard's suitability to measure criminal responsibility may not be a reasonable basis for thinking it appropriate to measure treatment competency. And the objections and responses to use of the standard in the criminal context may not be pertinent to the treatment context. Both the volitional impairment standard and the product of mental illness standard derive from the criminal law. The strength of my analysis, therefore, turns, in part, on the legitimacy of borrowing in this way from the criminal law.

Concerns about moving freely between the criminal and civil contexts are only partly warranted. Criminal responsibility is best understood as competency to make the choice that leads to the criminal act. In all competency contexts, the general question is whether the actor is sufficiently intact that we should permit him to make choices that pose him a risk of suffering unhappy consequences. In the criminal context, those consequences include blame and punishment imposed by the state. Yet, the consequences of treatment choices may be equally weighty. In either case, the capacities that make one a suitable decisionmaker should, in principle, be largely the same, or at least overlapping.

Nevertheless, there are important differences between the contexts. First, criminal responsibility involves a kind of moral accountability that may be unnecessary in the competency context. To be morally accountable—in part, to suffer the particular consequence of being blamed—may require fulfillment of conditions not particularly relevant to the competency context. For example, some argue that the ascription of blameworthiness requires that one's act reflect on one's character.[87] Being a moral agent may require something more or different than being a competent decisionmaker. To this extent, criminal responsibility and competency may not be entirely of a piece.

(5th Cir. 1984); Hart v. State, 702 P.2d 651, 653 n.1 (Alaska Ct. App. 1985); People v. Skinner, 704 P.2d 752, 759 (Cal. 1985); Sanders v. State, 585 A.2d 117, 124 (Del. 1990).

87. For an account of the character-based theory, together with citations to its main proponents, see Saks, *supra* note 25, at 418-24.

Second, the same arguments in the criminal and treatment contexts may warrant different conclusions given second-order and normative commitments. In particular, the tests often suffer serious second-order application problems. Using the tests creates a risk of too many "moral mistakes." This kind of "risk of error" reasoning may warrant different conclusions depending on which kind of error one fears most—which may differ in the two contexts.

The kind of error to which insanity and competency tests are most liable is that of "false positives," that is, finding people who are responsible/competent to be irresponsible/incompetent. In the insanity context, I worry most about depriving of freedom those who are morally blameless, and thus I would endorse a test that errs on the side of false positives in that context. In the treatment competency context, by contrast, I worry most about depriving of freedom and dignity those who are capable of choice, and thus would reject such a test in that context. In the context of our current discussion, this means that we might retain a volitional impairment standard in the arena of responsiblity, but reject it in the arena of incompetency. Others, of course, may have different normative commitments leading them to different conclusions.[88]

The idea that a normative preference for liberty should lead to liberal insanity tests and strict competency tests is strengthened by the fact that the subjects of insanity and incompetency inquiries are relatively powerless. The context of criminal responsibility pits the individual against the state, authorizing the state to visit suffering on the individual itself. Because of gross power imbalances between the individual and the state, we may wish to lighten the individual's burden so that he may more easily avoid further entanglement with the state. In particular, we may wish to provide the individual with more liberal insanity standards in order to give him broader means to establish his unsuitability for punishment. The same argument leads to the opposite conclusion in the treatment competency context. There, we may wish to protect the powerless patient's liberty by withholding a liberal incompetency test from the state.

Although there clearly are important differences between the criminal and treatment contexts, there are important similarities as well. So long as we keep the differences in mind, responsibility stan-

88. For example, in the criminal context, one might worry most about false positives because of the ideal of equal application of the law combined with the fact that criminal law imposes only minimum standards of conduct. By contrast, in the treatment context, one might worry most about false negatives, because one subscribes to a form of paternalism in this context, wishing most to help people in need of help.

dards may prove quite useful to an analysis of competency standards. Yet we must also be mindful of characteristics specific to our particular context. I turn now to such characteristics in the treatment context.

D. *Arguments Specific to the Competency to Refuse Context*

There are good reasons to expect that few desires are irresistible in the context of psychotropic treatment decisions.[89] First, this section suggests general reasons why desires for psychotropic medication decisions are much less likely to be irresistible than desires for actions in the criminal law context. I then discuss some formal features of the context in which a person makes a treatment choice, and argue that these features help to ensure that there will be few desires irresistibly leading to enactment here. Because the cases of irresistible desires leading to treatment refusal may be so numerically insignificant, and because the volitional test is troublesome for other reasons, I argue against its adoption. Little is lost by ruling out the test because we avoid an undue number of false positives without thereby permitting too many false negatives.

1. A REASON FOR THINKING IMPULSES TO REFUSE ARE UNLIKELY TO BE IRRESISTIBLE

Desires underlying treatment choices are not as likely to be irresistible as desires underlying crimes for several reasons. The drives most likely to move a person to act against his will—to ruthlessly satisfy a hunger or dominate a threatening aggressor[90]—are no doubt the more basic drives, libidinal and aggressive.[91] These drives may, at

89. Perhaps I should make the weaker claim that fewer people are apt to have irresistible desires in the treatment context. Yet it strikes me that the more absolute claim—*few* people are likely to have such desires—is justified, given the nature of irresistible desires and the formal features of the choice here. If my stronger claim fails to be persuasive, at least the weaker may be justified by the considerations raised in this section.

90. The concept of an irresistible drive may also seem more at home in the realm of action than choice, because acting on such primal drives makes good survival sense. There is no surer way to satisfy a drive than to act oneself to satisfy it.

91. Dividing drives into libidinal and aggressive drives has clear support in psychoanalytic thinking over the last century. *See, e.g.*, OTTO FENICHEL, THE PSYCHOANALYTIC THEORY OF NEUROSIS 57-58 (1945); MELANIE KLEIN, ENVY, GRATITUDE AND OTHER WORKS 271-72 (1975); RENE SPITZ, THE FIRST YEAR OF LIFE 167-68 (1965). Nevertheless, although the *division* seems appropriate, the tendency today is to talk of "drive-derivatives" rather than "drives." The latter language is thought more appropriate to physiological discourse than to psychological. *See, e.g.*, ANNA FREUD, NORMALITY AND PATHOLOGY IN CHILDHOOD: ASSESSMENTS OF DEVELOPMENT 14-15 (1965); Morton Shane & Estelle Shane, *Change and Integration in Psychoanalytic Developmental Theory*, *in* CALVIN SETTLAGE & REED BROCKBAND, NEW IDEAS IN PSYCHOANALYSIS 69-70 (Calvin F. Settlage & Reed Brockband eds., 1985). Moreover, object relations-theorists (as opposed to the more classic "drive-

times, be directly implicated in actions governed by the criminal law. For example, a person may have a strong physical craving or addiction that impels theft to satisfy the urge. Or a person may feel such extreme, though unwarranted, fear that he strikes out to subdue someone he perceives as a threatening aggressor. Similarly, a person's strong maternal instincts may lead her to kill a child whom she expects to lead a life of suffering.[92] In short, criminal acts are extremely varied, and may occur in extremely varied circumstances, with some implicating the more primordial drives.

Psychiatric treatment choices, by contrast, seem unlikely to implicate the most basic desires. Taking or not taking psychotropic medicine does not seem the kind of event one would have overwhelming feelings about. The patient may want to get better, fear certain side-effects, or prefer to be self-sustaining. But with relatively harmless agents that are not likely to cause or prevent death—in which case a primordial drive might well be behind a decision to accept or refuse—people's desires are not likely to be so strong as to compel their choices.

The difference between the criminal law context and the psychiatric treatment context should be clear. Because the criminal law covers many harmful acts in very many imaginable contexts, and thus involves the whole gamut of human motivations, the more basic, primitive motivations may sometimes be at work. Psychiatric medication decisions, by comparison, involve a relatively narrow range of issues, and do not usually involve fundamental needs and desires, such as survival.

Two considerations may call this conclusion into question. First, taking medication may implicate a loss of self, which is a fundamental threat. And second, it may simply be wrong that only primordial desires give rise to significant volitional impairment.

---

theorists") are much more visible today, and greatly de-emphasize the role of the drives or drive-derivatives in psychoanalytic theory and practice. Some influential object relations theories are those of Klein, Kohut, and Kernberg. *See generally* OTTO F. KERNBERG, OBJECT-RELATIONS THEORY AND CLINICAL PSYCHOANALYSIS (1976); OTTO F. KERNBERG, SEVERE PERSONALITY DISORDERS: PSYCHOTHERAPEUTIC STRATEGIES (1984); MELANIE KLEIN, THE PSYCHO-ANALYSIS OF CHILDREN (Ernest Jones ed. & Alix Strachey trans., 3d ed. 1949); MELANIE KLEIN, OUR ADULT WORLD AND OTHER ESSAYS (1963); HEINZ KOHUT, THE ANALYSIS OF THE SELF: A SYSTEMATIC APPROACH TO THE PSYCHOANALYTIC TREATMENT OF NARCISSISTIC PERSONALITY DISORDERS (1971); HEINZ KOHUT, SELF PSYCHOLOGY AND THE HUMANITIES: REFLECTIONS ON A NEW PSYCHOANALYTIC APPROACH (Charles B. Strozier ed., 1985).

92. Suppose in both of the last two cases that the actors have no frank delusions—that neither is grossly out of touch with reality. Rather, the first person feels fearful and the second, pessimistic, and it is these feelings that impel their actions.

a. Psychotropic Medication as Threatening a Loss of Self

Psychotropic medication alters one's mind, and a person who needs psychotropic medication may have an overwhelming fear of such alteration.[93] She may fear not only the psychic changes themselves, but even more, that she will be unable to assess these changes by virtue of her very loss of self. Loss of self would seem as significant a threat as some of the other threats I have termed "primordial." Overwhelming fears of loss of self may give rise to truly irresistible desires to refuse the substance that poses the threat.

Now one might suggest that overwhelming fears of radical and permanent personality change as a result of drugs with psychological effects are uncommon. Consider, for example, the prevalence of alcohol and drug use in our society. Alcohol consumption is fairly common and perhaps experimentation is almost universal.[94] Yet fears—and underlying notions of impairment of identity—may be culture-bound. Although our culture does not perceive use of alcohol as a threat to identity, it might fear, for instance, use of psychedelic mushrooms as impairing identity. For the Navaho Indians, by contrast, taking these mushrooms is part of the fabric of everyday life.[95] Much,

---

93. *See, e.g.*, EDWARD M. PODVOLL, THE SEDUCTION OF MADNESS 227-29 (1990). There are other reasons that a patient may have an extremely strong desire to refuse. For example, the patient may be a celebrity who derives such satisfaction from his appearance or thinks his career so depends on it that he is terrified of tardive dyskinesia (TD), a neurological side-effect associated with prolonged use of antipsychotic agents that is evidenced by tic-like movements. Even the slight movements which are most common would disturb him, and severe movements would so horrify him that he is unwilling to take even the remotest risk of developing them. *See generally* AMERICAN PSYCHIATRIC ASS'N, TASK FORCE REPORT, TARDIVE DYSKINESIA 18 (1979); Daniel E. Casey, *Tardive Dyskinesia*, 153 W. J. MED. 535 (1990); J. Gerlach & Daniel E. Casey, *Tardive Dyskinesia*, 77 ACTA PSYCHIATRICA SCANDINAVICA 369 (1988).

The same seems true in the ordinary medical treatment context. Generally, only when life is threatened will a treatment choice involve irresistible desires, but sometimes they will be present even in less important contexts. Imagine a person who strongly values mobility—an athlete, for example—who learns that he desperately needs an angiogram to rule out a life-threatening aneurism, and that the angiogram has a one percent risk of causing paralysis. This person may have an overwhelming desire not to have the angiogram.

I do not deny that sometimes situations such as these will produce very strong desires to refuse treatment. In such cases, there will also be countervailing considerations that may well attenuate or cancel out the force of the desire to refuse. Even though such strong desires are possible, I still believe that they are uncommon. Most people do not have overwhelming desires to refuse essential procedures because of very small risks of less-than-tragic disfavored outcomes.

94. *See, e.g.*, David C. Rowe & Joseph L. Rodgers, *Adolescent Smoking and Drinking: Are They Epidemics?*, J. OF STUD. ON ALCOHOL 110, 116 (1991); National Ass'n of State Alcohol and Drug Abuse Directors, ALCOHOL USE AND ALCOHOLISM, *in* ADDICTIVE BEHAVIOR: DRUG AND ALCOHOL ABUSE 131 (Morton Publishing Co. ed., 1985).

95. Members of the Native American Church take Peyote as part of religious worship in ceremonies from sundown Saturday to sunrise on Sunday. *See* People v. Woody, 394 P.2d

then, turns on people's *perception* of the effects of taking the drugs. Thus, the readiness of many people to try alcohol does not demonstrate that significant fears of psychotropic drugs must be rare.

Even if our society does perceive psychotropic drugs, unlike alcohol, as mind-altering in a deep and troubling way—as causing a true loss of self—this perception is inaccurate. Education can teach patients about the true effects of the drugs, especially that they do not reach deep into the personality.[96] Indeed, in one sense, the drugs cause less significant changes than alcohol. They are profoundly *normalizing*.[97] Thus, being intoxicated is a much more significant departure from normality than being medicated. If the medications *restore* one to one's normal state—if, on medication, one is just as one remembers oneself before becoming ill—then being on psychotropic medication is more like being sober than like being intoxicated.

813, 817 (Cal. 1964). Members typically take "quantities sufficient to produce an hallucinatory state." *Id.* According to the theology of the church, peyote "enables the participant to experience the Deity." *Id.* at 817-18. These "rituals are an integral part of the life process." *See* Employment Div., Or. Dep't of Human Resources v. Smith, 494 U.S. 872, 919 (1990) (Blackmun, J., dissenting) (quoting Brief for Association on American Indian Affairs as Amicus Curiae). The mushrooms, then, are fairly central to these Native Americans' way of life. *See, e.g.*, John Rhodes, *An American Tradition: The Religious Persecution of Native Americans*, 52 MONT. L. REV. 13 (1991); Debra A. Mermann, Note, *Free Exercise: A "Hollow Promise" for the Native American in* Employment Division, Department of Human Resources of Oregon v. Smith, 42 MERCER L. REV. 1597 (1991). Although they may see the mushrooms as exerting a fairly powerful mental influence during the ceremonies, members of the Native American Church do not fear them as radically and permanently altering their identities.

96. One might argue that I am begging the question: to say that education will be able to overcome these fears is to suppose that these patients are not suffering from irresistible desires, which is the very issue in question.

But this argument misconstrues the structure of mine. My argument takes the following form. Irresistible desires are likely only when the interests implicated by a decision are important. Patients may think that taking psychotropic medication implicates important interests, inasmuch as it threatens them with loss of self. But once they learn, through education, that they are wrong in this assumption, they will cease having any irresistible desires to refuse the medication, because the basis for their having such desires—the misperception that the medication implicates important interests—will have vanished.

If a patient's strong desire has come about for another reason, then he will continue to have the desire despite the education. Or if he has some deep need to think he is refusing because of fears of the changes, he will be uneducable in this regard. This scenario merely presents the case in which my diagnosis of the reason for the irresistible desire is mistaken. In cases in which the diagnosis is sound, my argument is not at all circular.

97. *See, e.g.*, Thomas G. Gutheil & Paul S. Appelbaum, *"Mind Control," "Synthetic Sanity," "Artificial Competence," and Genuine Confusion: Legally Relevant Effects of Antipsychotic Medication*, 12 HOFSTRA L. REV. 77, 101 (1983). By "normalizing" and "normal," I mean essentially "healthy." Of course, a patient may be ill for such a long time that the ill state has become normal (i.e., customary) for her. Then, changes that restore her to her prior, healthy state may appear extremely threatening. For this case, see my discussion of the phenomenon of resistance at *infra* text accompanying notes 108-15.

Indeed, whatever changes these drugs cause, they are clearly not permanent. But then it may not even matter how significant the changes are. The patient may give the drugs a chance and see if she approves of the changes. If the patient is fearful that the drugs will rob her of the ability to judge the changes, perhaps even change her very preference to avoid taking the drugs, she can always self-bind to a future "holiday" from taking the medication.[98]

If caregivers are unable to convince patients that they risk no loss of self by taking the medication, it is possible that the patients are suffering significant cognitive distortions, perhaps on an unconscious level. Although risk of loss of self is a fundamental threat that might well give rise to strong desires to avoid the threat, risk of changes that are less personality-altering than those that alcohol causes is not such a threat and should not give rise to such desires. If the patient persists in thinking that the medication poses a real and radical threat to his personhood, he holds a somewhat distorted belief in the face of evidence to the contrary. Does he think that his doctors have malevolently misled him? That they are good-willed but grossly incompetent? That the medications will have a unique effect on him, changing him from a Dr. Jekyll to a Mr. Hyde? Of course, a patient need not believe all that his doctors tell him in order to be sufficiently cognitively intact to make a competent treatment decision. But it is possible that a person who, after education to the contrary, continues to fear that the drugs will cause radical and permanent personality alteration is suffering from serious cognitive distortions. Most patients, by contrast, should lose any undue fears of loss of self once reassured about the true properties of the medication.

Thus, the first objection to my claim that people are unlikely to have overwhelming desires to refuse psychotropic medication does not seem compelling. With education, people should cease to see the medication as threatening a significant loss, and their powerful desires to refuse should cease.[99]

---

98. Again, psychotropic drugs do not cause radical changes in personality, and thus are unlikely to change a person's evaluation of any changes that do occur, or her preferences regarding drug-taking. Once beginning a course of medication, many patients are happy to be taking the medication, despite initial reluctance. This is generally because they truly *feel* better, or have lost delusions that contraindicated drug-taking. The drugs do not change preferences by changing the patient's personality, but rather by giving her reasons, such as feeling better, for altering her views. The fact that many patients in jurisdictions that deny them a right to refuse persist in objecting even after taking the drugs for some time suggests that psychotropic drugs do not render one helpless to resist their imagined effects.

99. Other related fears, however, may give rise to overwhelming desires to refuse medication. For example, a patient may fear that the medication will change a quality that is essential to his conception of himself. *E.g.*, OLIVER SACKS, THE MAN WHO MISTOOK HIS

b. Are Desires Really Overpowering Only When the Stakes Are High?

Yet it may not be entirely accurate to claim that people experience overpowering desires only when the stakes in a decision are high. Consider that kleptomaniacs often steal trivial items for which they have no use.[100] What primordial interests and drives are implicated here? Even if irresistible wishes may motivate some acts that do not

---

WIFE FOR A HAT 92-96 (1970) (case of "Witty Ticcy Ray," a Tourette's patient who decided not to take his antipsychotic medication during weekends because he felt it robbed him of the ability to play his jazz drums in an inspired and creative way). Similarly, a patient may feel that taking the medication will produce a lessening of self-esteem, inasmuch as, in his view, he is forced to rely on a substance to "prop himself up." The patient does not fear the pharmacologic effects of the drugs, but the psychological effects of what taking the drugs represents to him. Finally, the patient may fear a loss of control over his mental processes.

The first case differs from that in the text inasmuch as the patient may be *right* that the drugs will change a quality that he deems central to his self-concept. For example, the drugs may make him less lively. But few patients are likely to find relatively insignificant changes such as these to profoundly alter their very identity, such that they will develop an overwhelming fear of the changes and of the substance that causes them. (For the response to the argument that patients may have irresistible wishes to refuse even if they do not find the changes significant, see *infra* text accompanying notes 100-15.) Moreover, because the changes these drugs cause are not permanent, patients can discontinue the drugs if they are unhappy with the results, as Sacks' patient did during weekends, or work on accepting their new selves in therapy.

A patient may also fear loss of self-esteem. In taking the medication, she may no longer see herself as self-sustaining. Many may hesitate to take drugs for this reason. Whether loss of some self-esteem is a significant threat that might prompt an irresistible desire depends on the kind and amount of impairment of self-esteem and how the patient conceptualizes it. The need for drugs is likely to cause an impairment of self-esteem that the patient views as a significant threat when she considers self-sufficiency central to her selfhood. This case is then analogous to that of the person who feels liveliness central to his selfhood, and the same analysis seems to apply.

Some patients may fear loss of control over their mental processes as a result of taking the drugs. Here, the analysis differs from the cases above because this fear may have less basis in fact. Patients may fear loss of control in the sense that they fear their thinking will become out-of-control (in the way that alcohol causes one's thinking to feel out-of-control when one is intoxicated). But the drugs do not produce this effect. On the other hand, a patient will lose some control in the sense that, if he is on anti-depressants, he will simply find it much harder to feel abnormally depressed. But the person who is *not* on anti-depressants probably finds it even harder *not* to feel abnormally depressed. He has no more or less control over his mental processes on these drugs than off them. Finally, fear of loss of control may simply be shorthand for fear of uncontrollable changes in one's thoughts and preferences—in one's personality. This case then becomes the case discussed in the text. As I argued there, the fear is groundless. Education should permit patients to overcome all of these fears of loss of control over their mental processes. Reassurance that they can refuse the drugs later if they feel insufficiently in control should further reduce such fear.

100. DSM-III-R, *supra* note 14, at 322-23; *see also* David A. Fishbain, *Kleptomania as Risk Taking Behavior In Response to Depression*, 41 AM. J. PSYCHOTHERAPY 598, 602 (1987); Marcus J. Goldman, *Kleptomania: Making Sense of the Nonsensical*, 148 AM. J. PSYCHIATRY 986, 986-87, 994 (1991); Marcus J. Goldman, *Kleptomania: An Overview*, 22 PSYCHIATRIC ANNALS 68 (1992).

evidently implicate significant interests, however, I suspect that in many of these cases the actors have unconscious fantasies about the significance of their acts that *do* implicate important interests. For example, a person who steals trivial items may unconsciously equate the items with mother's milk. Her strong urge to steal may represent a strong urge to be fed, which *is* fairly basic.

The problem with this response is that, in the psychiatric context, such unconscious fantasies may also be at work. For example, a patient may have a strong urge to refuse medication because of an unconscious fantasy that he will merge with his doctor if he complies. In the general medical context, a person may have a strong urge to refuse a surgical procedure because he unconsciously equates the surgeon's cutting with castration.

Yet all of these cases seem to involve gross unconscious distortions. If so, arguably all provide a basis for incompetency findings quite apart from any compulsive desires that may be at work: one of the premises of the practical syllogism motivating the act is patently irrational. In fact, however, it is very unclear how competency theory should deal with unconscious delusions that motivate decisions.[101] Still, *if* we should deem such unconsciously delusional decisionmakers incompetent, then a volitional impairment test would not be necessary in this kind of case.[102]

On the other hand, if a cognitive test is not available to find unconsciously delusional people who therefore have overpowering wishes incompetent, a volitional test may be necessary. It may also be necessary if I am simply wrong about when irresistible desires occur. Perhaps they *do* often occur in routine situations even absent unconscious fantasies.

But I am going to challenge now the idea that it may be reasonable to expect significant numbers of overpowering wishes in such cases. Whether accompanied by unconscious fantasies or not, compulsive desires in more routine situations that do not involve the more significant interests seem to fall into a few discrete categories. Klepto-

101. *See* Elyn R. Saks, Should Unconscious Delusions Vitiate Treatment Competency? (May 1992) (unpublished manuscript on file with the author). Although my intuition is that unconscious delusions should not vitiate competency, I have not found any arguments that persuasively explain why.

102. My claim here is similar to but more plausible than the claim that others have made in the criminal law context that a volitional test is unnecessary because the required volitional impairment is coextensive with cognitive impairment. I have suggested that the two types of impairment may not be coextensive. *See supra* text accompanying notes 79-83. But pointing to cognitively intact volitionally disturbed patients does not rule out the possibility that they have *unconscious* cognitive impairments.

mania,[103] compulsive gambling,[104] and obsessive-compulsive disorder[105] are all recognized syndromes that give rise to strong urges in situations that do not involve significant interests.[106]

Is there such a recognized syndrome in the psychiatric treatment area? Psychiatrists have described a disorder in which persons voluntarily produce symptoms in order to obtain medical treatment.[107] But in our context, recall, it is incompetent refusals that are problematic. More important, we should deny these patients treatment, not because they are incompetent, but because they do not need it.

Psychiatrist have also described a phenomenon called "resistance," which is central to the psychoanalytic enterprise.[108] Here, the patient clings to her familiar constellation of defenses and symptoms, and "resists" the cure, because of fear of what change might bring. The patient has used her symptoms, in a sense, to identify herself ("I am my symptoms"), and thus fears loss of these symptoms. She also fears the unknown new self and what life as that self might be like.[109]

---

103. *See* DSM-III-R, *supra* note 14, at 322-23.

104. *Id.* at 324-25.

105. *Id.* at 245-47.

106. I include the addictions, because they involve strong physical urges, as similar to hunger, and therefore of the "high-voltage" kind.

107. This disorder is commonly known as Munchausen Syndrome, and is called "Factitious Disorder" by the American Psychiatric Association. *See* DSM-III-R, *supra* note 14, at 315-20. *See generally* Anne Cremona-Barbaro, *The Munchausen Syndrome and its Symbolic Significance: An In-depth Case Analysis*, 151 BRIT. J. PSYCHIATRY 76 (1987); P.M. Higgins, *Temporary Munchausen Syndrome*, 157 BRIT. J. PSYCHIATRY 613 (1990); James P. Mayo & John J. Haggerty, *Long-term Psychotherapy of Munchausen Syndrome*, 38 AM. J. PSYCHOTHERAPY 571 (1984); Kenneth Sinanan, *Evolution of Variants of the Munchausen Syndrome*, 148 BRIT. J. PSYCHIATRY 465 (1986). Patients with this disorder voluntarily produce symptoms of illness, apparently solely in order to receive treatment. For example, a patient may swallow iodine so that test results will suggest she has a serious illness. Often patients go from doctor to doctor seeking and obtaining treatment. Often the treatments are quite intrusive and involve considerable danger, such as surgery. This disorder is not well understood, but it may be helpful to conceptualize Munchausen patients as experiencing "irresistible wishes" to seek and consent to treatment. Nevertheless, a volitional impairment competency test is not necessary to deal with these patients' choices.

108. *See, e.g.*, ANTON KRIS, FREE ASSOCIATION 33-35 (1982); Stephen A. Morganstern, *Psychoanalytic Process and Transference, with Notes on the Need for Re-Analysis*, 59 PSYCHOANALYTIC Q. 712, 733 (1990); David Rosenfeld, *The Handling of Resistances in Adult Patients*, 61 INT'L J. PSYCHOANALYSIS 71 (1980).

109. This phenomenon ties in with the idea that patients will fear the psychological changes wrought by the drugs. *See supra* text accompanying notes 93-99. In a sense, it is a more subtle version of the same thesis: the patient fears not just radical and permanent personality changes, but any personality changes. Thus, reassuring the patient that the medication will not produce such radical changes that she will not recognize herself will not help, nor will reassuring her that the drugs will be normalizing. The point of resistance is that the patient clings to the abnormal, unhealthy state. The disordered state, in a sense, has become normal for the patient. Thus the patient fears not just changes that make her more abnormal, but any psychological changes that the treatment might bring about.

I do not doubt that resistance to psychoanalytic treatment is pervasive and powerful. But I question whether it is truly compulsive. Evidence suggests that it operates outside of patients' awareness. For example, a patient may, in all sincerity, cite many reasons for persistent lateness to sessions, and only later realize that her unconscious "arranged" for her lateness precisely to avoid the treatment. But the fact that an influence on our behavior is unconscious does not mean that it exerts *compulsion*. Most behavior has unconscious sources, yet most behavior is not compelled. Similarly, people often seem able to resist unconscious influences, perhaps unconsciously, so that their presence is not enough for compulsion.

Perhaps resistance is a compulsive phenomenon in the sense that it occurs despite the patient's best intentions and efforts. How else can we explain its occurence? But other explanations are also possible: the patient's intentions are ambivalent and her efforts halfhearted—all, again, outside of her conscious awareness. If so, the patient is not overcome by powerful forces that she is struggling against, but, rather, is not truly struggling.

Indeed, there are reasons for thinking resistance is *not* truly compulsive. First, if it were, most psychoanalyses would likely fail miserably. That many analyses *are* successful suggests that resistance is not truly compulsive—or at least that it can be overcome by means other than force.[110]

Second, if resistance were truly compulsive, we would have to conclude that many high-functioning neurotic outpatients are incompetent. If compulsive resistance is not limited to psychoanalysis, but affects all treatment, then the numbers of mentally healthy people who are not competent to decide on treatment would be staggering.

If true, this would radically affect people's perception of competency—indeed, of their very *selves*. If competency continued to play the role it does in the treatment context, pervasive incompetency would also produce severe practical strain. For example, our methods of administering care would need to change radically. Of course, these conceptual and practical repercussions of pervasive treatment incompetency, so construed, do not demonstrate that it does not exist. But they may well provide reasons to prefer a competency standard that refers to more pronounced deficiencies that only relatively few suffer.[111]

---

110. This may suggest that psychiatrists should talk to apparently compulsive resisters of psychotropic medication rather than forcing treatment on them.

111. In a previous article, one of the criteria I set out for judging the adequacy of a treatment competency standard was that the standard not render too many incompetent in the

I am unable here to explore fully how pervasive resistance is, and whether it is truly compulsive. At least we need more research. One wonders, for example, whether resistance is as common in treatment contexts other than the psychoanalytic. Psychoanalysis is very unstructured. Patients may not have a good sense of what they are supposed to be doing. Thus, the notions of compliance and noncompliance are much fuzzier here than in the ordinary treatment context. Moreover, psychoanalysis is just plain *hard* in a way that most treatment is not, because it involves much work on the patient's part, and pain over a prolonged period.[112] For these reasons, one might expect more resistance in the psychoanalytic context than in other treatment contexts.

More empirical work is needed, then, to determine the pervasiveness of resistance not only in psychoanalytic, but also in non-psychoanalytic contexts.[113] The question is not solely one of refusal rates. Research must consider whether the refusal is a function of resistance or of other things. To consider *all* noncompliance a matter of resistance is plainly wrong.[114] Finally, the rate of resistance-based refusal that is *compulsive* must be determined. Although more work is clearly needed to answer this question—both empirical and concep-

---

face of the pervasive influence of the irrational and the unconscious. *See* Saks, *supra* note 6, at 950.

112. The implicit suggestion here is that resistance may not be entirely a matter of fear of psychological change. It may also be a matter of avoiding work and pain.

113. Evidence suggests that patients of all kinds are fairly noncompliant with medication recommendations. *See, e.g.*, Steven Axelrod & Scott Wetzler, *Factors Associated With Better Compliance With Psychiatric Aftercare*, 40 HOSP. & COMMUNITY PSYCHIATRY 397 (1989); Jenny L. Donovan & David R. Blake, *Patient Non-Compliance: Deviance or Reasoned Decision-Making?*, 34 SOC. SCI. & MED. 507 (1992); Linda M. Hunt et al., *Compliance and the Patient's Perspective: Controlling Symptoms in Everyday Life*, 13 CULTURE MED. & PSYCHIATRY 315 (1989). On the other hand, the rate of *compliance* with medication regimes also is fairly high. *See, e.g.*, Norman S. Harvey & Malcolm Peet, *Lithium Maintenance: 2. Effects of Personality and Attitude on Health Information Acquisition and Compliance*, 158 BRIT. J. PSYCHIATRY 200 (1991) (controlled education program improved patients' medication compliance); E.D. Myers & Alan Branthwaite, *Out-patient Compliance With Antidepressant Medication*, 160 BRIT J. PSYCHIATRY 83 (1992) (patient antidepressant medication compliance improved when patient allowed to choose dosage regime). Indeed, in the treatment competency context, most patients eventually accept their doctors' recommendation to take psychotropic medication. *See infra* note 115. In the general medical context, patients are extraordinarily compliant. For example, patients rarely refuse exigent procedures such as surgery. *See* 2 PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, MAKING HEALTH CARE DECISIONS 417 (1982).

114. Patients may not take prescribed medication, for example, because they do not feel sick, do not trust their doctors' judgment, or fear unknown risks. Indeed, the very fact that patients are highly compliant with recommendations for urgent treatment suggests that, in nonurgent situations, they may simply be evaluating the risks and benefits of treatment differently than their doctors, not that they are under the sway of resistance.

tual—my intuition is that patients are rarely *compelled* to resist treatment.

In conclusion, the phenomenon of resistance does not undermine my claim that few irresistible desires are implicated in the treatment context. Few people deciding on whether to take psychotropic medication are likely to be truly compulsive resisters. Some, of course, will be. But research is necessary to provide reasons to think they are compulsive resisters, and ways of identifying these patients. Now that medication-refusal is permitted in some places, physicians may begin to describe and provide criteria for identifying a subclass of resisters whose behavior truly is compulsive.[115] Until they do, we should not expect many irresistible impulses in this context: the decisions do not implicate such significant interests as to stimulate truly primordial drives, and they are not in the category of recognized exceptions.

2. A FURTHER REASON: FORMAL FEATURES OF IMPULSES FOR TREATMENT CHOICES

Even if one rejects my claim that, because the interests involved in psychotropic medication decisions are relatively unimportant, few irresistible desires are implicated, other more abstract reasons lead to the same conclusion. These derive from formal features of the context in which a patient makes a treatment decision. Specifically, the patient has a strong desire here, not for an act, but for a *choice* that someone *else* must enact or permit to be enacted. This formal feature has implications that will lead to very few problematic irresistible desires in this context. Some of these implications obtain whenever a competency finding is *prospective*. But prospectivity does not exhaust the implications which, in any case, lead to the same result, namely, few compulsive desires.[116]

---

115. For empirical research on medication refusal in jurisdictions in which there is a right to refuse, see, e.g., Paul S. Appelbaum & Steven K. Hoge, *Empirical Research on the Effects of Legal Policy on the Right to Refuse Treatment, in* THE RIGHT TO REFUSE ANTIPSYCHOTIC MEDICATION 87-89 (American Bar Ass'n Comm. on the Mentally Disabled ed., 1986). The authors' review of published and unpublished studies of antipsychotic medication refusal by psychiatric inpatients does not confirm psychiatrists' initial fears that an epidemic of refusals would sweep psychiatric facilities. Studies using data from formal procedures for evaluating refusing patients or requiring refusals to persist for at least 24 hours before recording them found low incidences of treatment refusal. *Id. See generally* Paul S. Appelbaum & Thomas G. Gutheil, *Drug Refusal: A Study of Psychiatric Inpatients*, 137 AM. J. PSYCHIATRY 340 (1980) (study suggesting that, when permitted to refuse medication, a substantial percentage of psychiatric patients will do so, but generally only one to five percent will do so consistently). Further study of those patients who refuse persistently may reveal whether their refusal is compulsive or in some other way problematic.

116. One might have an opposite concern about transporting the idea of an irresistible desire to act beyond treatment choices to the sphere of choice in general; *much* at the level of

The differences between the criminal law and treatment contexts are again relevant in this regard. In the criminal law context, one has a strong desire to perform an act oneself; in the treatment context, by contrast, one has a strong desire to make a *choice*, namely, to accept or refuse treatment. Some desires for choice *reduce* to desires to perform an act oneself,[117] but those in the treatment context generally do not. When one chooses to accept treatment, the *doctor* must implement one's choice. When one chooses to refuse treatment, the doctor must refrain from imposing it. Thus, in the first case one may have an irresistible desire for the doctor to *implement* treatment; in the latter, one may have an irresistible desire to get her to *refrain* from it. While

---

thought seems compelled compared to the level of action. Consider, for example, that if one sees a fly ball that tens of thousands of others are also seeing, one can hardly doubt that it is there; one "cannot help" believing in its existence. One may similarly be unable to help feeling grateful toward someone who has spared one pain. Undoubtedly, the experience of having thoughts and wishes come unwanted to mind is more widespread than the feeling of being unable to prevent an action. Although it is undesirable for disabling consequences to follow such widespread, innocuous occurrences, applying the concept of irresistible desires to choice may force us to excuse many people from responsibility and to deprive many people of choice.

This concern underscores the necessity for a careful analysis of the concept of a volitional impairment. For example, saying a thought or feeling is compelled may imply, on some level, a genuine desire that it not occur. Thus, when someone acts on a strong wish congruent with his normal personality, for instance, to help someone, no one would say that he yielded to an irresistible desire, even if not helping would be inconceivable to him. Similarly, a feeling of gratitude may be said to be compelled only if repudiated. And a compelled belief should be construed, not as a belief compelled by the evidence (which is a belief one would *want* to believe), but a belief one cannot help adopting despite the evidence. If this is right, many of the innocuous occurrences described above are not properly said to be compelled.

But perhaps this is not right. Consider that a profoundly depressed mother who kills her infant may have no desire not to kill her. She may be so in the grips of her illness that the killing seems absolutely right. Do we want to say was she not compelled to act? Perhaps on some level she would wish not to do the act. Or perhaps we should analyze actions differently from beliefs and feelings. The point remains: a more refined analysis of the idea of a volitional impairment is needed.

Even if it is incorrect to say that a compelled thought or feeling requires a strong desire that it not occur, the choices supported by the thought or feeling may remain uncompelled. For instance, one may be tormented by a classic obsessive desire to injure a loved one, yet such desires are classically *not* acted out, and thus remain unconnected to compelled choices. When they *are* connected to choice and action, as when they prompt one to seek help, the choices and actions are often themselves not compelled. *See* HUMBERTO NAGERA, OBSESSIONAL NEUROSIS: DEVELOPMENTAL PSYCHOPATHOLOGY 37-38 (1983); John C. Nemiah, *Obsessive-Compulsive Disorder, in* 2 COMPREHENSIVE TEXTBOOK OF PSYCHIATRY/III 1504 (Harold I. Kaplan et al. eds., 1980). Indeed, it is possible to argue both that most choices are not compelled and that many thoughts and feelings are, precisely because of the weak connection between the two. The possibly greater frequency of compelled mental than physical events does not lead to the unacceptable conclusion that all are incompetent, and therefore does not justify rejection of the idea of compelled choices.

117. For example, to feel compelled to choose to smoke is simply to have an overwhelming desire to smoke that causes one to engage in the act of smoking. When a "desire for a choice" reduces to "a desire for an act," the former makes as much sense as and is no less likely to be irresistible than the latter. The two locutions are just different ways of saying the same thing.

some desires for choice in the treatment context are not desires that someone else act or refrain, most are.

In the more common case of a desire for a choice that someone else must enact, the desire will not as irresistibly lead to enactment as an equally strong desire in the criminal law context for two reasons.[118] First, one's irresistible desire that someone else act is not irresistible to that actor. When enacting a choice requires someone else's cooperation, the path from one's desire to the act is less direct that when the act is one's own—and may, in fact, be impossible to cross. Thus, the doctor may refrain from medicating, and the patient may lack means, reasonable or otherwise, to cause the doctor to act.

Second, when one's desire is to make a choice that someone else must enact, one can always *protest* against the choice, and the actor may protect the decisionmaker's true preference. Because saying what one wants seems to be easier than doing it, protest should often be possible.[119] The formal features of the treatment context also help ensure that protest will be effective because, unlike the case of simple actions, there will necessarily be another actor here. Moreover, refusing to honor a choice is easier than physically interfering with an action, and thus, it will be easier for the auditor to respond.

When a person expresses one choice and then protests it, the classic situation calling for a competency evaluation on volitional impairment grounds is not present. If the doctor chooses to respect the protest as the patient's genuine choice, the patient will escape scrutiny on competency grounds altogether; no one inquires into competency when a patient makes the right choice. If the doctor is unable to determine which choice is truly the *patient's*, she may conclude that the patient is incompetent on the non-volitional ground that he lacks the ability unambiguously to express a preference. But these

118. The concern is only with irresistible desires that lead to enactment. Otherwise, a volitional impairment standard is not necessary to prevent them from being enacted. I argue below that there are means of preventing the enactment of irresistible desires in the treatment context. In doing so, I mean to exclude the case in which the examiner finds the patient incompetent *because* of the irresistible desire. In that case, a volitional impairment standard is needed to prevent the enactment.

119. Consider the case of the agitated psychiatric patient yelling out to staff that he is getting out of control. The patient cannot restrain himself, but he *can* ask for help restraining himself.

One might argue, however, that it is a mistake to think that asking for help to resist a strong desire is always easier than resisting it oneself. In the former case, one must ask for help, and that is hard for many people. Moreover, one will already be publicly committed to the opposite choice, and it may be embarrassing to reverse oneself in this way. But it is only if the wish not to publicly present oneself in a bad light is overwhelmingly powerful that a person will have a harder time asking for help to resist his other powerful wish than to resist it himself. I doubt that the wishes follow this pattern in very many people.

cases are not troubling, because in each the outcome means that a volitional competency standard is not needed to deny legal respect to the choices of people with overpowering desires.

If this is correct, the formal features of desires for choice should result in few desires irresistibly leading to enactment in the treatment competency context. Some of these features may occur whenever a competency evaluation, whether of an action or choice, is *prospective*. For example, when an examiner prospectively evaluates a person's capacity to perform an act, the examiner's acquiescence in the act is necessary, and he may not comply, for example, on the grounds of some non-volitional impairment. Moreover, the examiner's very presence permits the actor to protest the act he feels irresistibly driven to perform.

Both of these characteristics, however, are at least less robust than in the case of a choice that someone else act. In that context, the person whose cooperation is necessary may refuse to enact the choice for reasons other than the decisionmaker's incompetency. For example, a surgeon might deem the patient's choice medically unwise. Similarly, responding to the protest in that context requires the examiner only to abstain from acting, rather than to interfere physically with another's powerfully motivated act.[120] Although these differences are perhaps less pronounced in the medication-refusal context than in the context of accepting treatment,[121] the prospectivity of a competency evaluation still fails to exhaust the reasons that desires for choices that another person act are easier to contain than desires to act oneself.

Perhaps more important, the hypothetical of prospective evaluations of *acts*, rather than choices that another enacts, may be fundamentally flawed. Prospective evaluations of acts (e.g., crimes) is problematic for a number of obvious reasons, including undue interference in people's lives and lack of resources. A less obvious reason demonstrates why the analysis should focus on *choices that another*

120. The chronology may be such that the actor, irresistibly impelled to act, must be restrained *before* the examiner performs the competency evaluation, rather than as a result of his protest. Still, once restrained, he may then protest, so that the examiner does not need a volitional competency test to prevent the irresistible enactment of the desire. Whether restraint occurs before or after the evaluation, physical restraint may be necessary in the case of desires to act but not in the case of desires for choices that another enacts.

121. For example, a doctor may not cooperate with a choice for treatment for a variety of reasons. But in the medication-refusal context, the examiner *should* cooperate with the patient unless the patient is incompetent, or is in a condition that would justify noncooperation in the criminal context as well (for example, he presents an imminent danger to others). Similarly, imposing treatment on a protesting patient may well require the doctor physically to interfere with the patient's vigorous efforts to resist the treatment.

enacts rather than prospectivity. Prospective evaluations of acts are not generally performed, because if a person has the means within himself to perform an act he is irresistibly moved to perform, he will perform it and not wait for an evaluation. By contrast, if cooperation is necessary for his desire to be enacted, he has no choice but to wait. As a result, few people irresistibly driven to act will perform acts that can be prospectively evaluated.

Irresistible desires in the treatment context are less likely to be irresistibly enacted than those in the criminal law context whether the reason is the nature of the desire's object or the prospectivity of the evaluation. The formal features of the treatment competency context help ensure that irresistible desires are an insignificant concern there. Thus, rejecting a volitional impairment test should result in few false negatives in this context.

Although overpowering desires for treatment choices often conform to the pattern I described, some follow a different pattern. In this type of "irresistible desire to choose," one feels compelled to choose solely for the sake of *being seen to choose.* Thus, one may be compelled to choose an act without having any strong desires about the act itself. For example, a person may feel compelled to choose to go to X-rated movies because it bolsters his image as macho, or compelled to choose not to go to such movies because it undermines his image as sensitive, while his feelings about whether he actually goes to the movies are indifferent.

Being compelled to choose in this sense means feeling an overwhelming desire to represent oneself as wanting the action chosen. Declining treatment may be compelled in this sense when the patient has a strong desire to see himself as healthy and independent, just as accepting treatment may be compelled in this sense when the patient has a strong desire to see himself as good and compliant.

This case does not share the formal features of the first pattern described because no one else's action or forbearance is necessary to enact the desire to *choose*. Nevertheless, the focus is not on preventing the enactment of desires to choose as such but on preventing the *actions chosen*. The question then becomes whether *this* choicemaker can protest his choice so that the doctor will not honor it.

In the case of a person desiring to represent himself as wanting the action, protesting against the choice seems as difficult as actually resisting an impelled action. To protest is precisely to resist the desire rather than to enlist the aid of someone else in resisting the desire. It must surely be easier to enlist someone's help in resisting a desire than to resist it oneself. Thus, when a person is not able to stop himself

from acting, he may be able to ask someone to help him. But in this case, asking for help *is* to stop oneself from acting because the acting is representing oneself as wanting to go forward. The desire is for an appearance, and the protest belies that appearance.

In the case of irresistible desires of this kind, there may be an even better case than usual for overriding the choice. If the person does not much care about the act but only about being seen to reject it, we should have few scruples about mandating the act. We can help the patient without thwarting any of her actual wishes. Yet, if it is hard to determine the strength of a person's desires generally, determining when a person's desire is simply to be seen to choose, and not to have her choice enacted, seems to be nearly impossible. More important, there is little reason to think decisions of this kind are common. People are bound to care more about enacting their choices than merely evidencing them. Thus, the numbers here are probably insignificant.[122] If this is so, accepting a volitional impairment standard because of the possibility of such desires seems misguided.

To reiterate, the formal features of desires for treatment choices seem to suggest that few desires will irresistibly lead to enactment in the treatment context. Most desires for treatment choices require someone else's cooperation. The other person may be unwilling to cooperate, or the patient may protest the choice, thereby securing the other person's non-cooperation. Although other desires for treatment choices do not require another's cooperation and cannot be protested, because they are really desires to be seen to choose, most people are likely to care more about what they choose than about being seen to choose. Thus, dispensing with a volitional impairment standard is likely to result in few false negatives in the treatment competency context.

E. *Conclusion*

A volitional impairment standard is unnecessary in the treatment competency context, because few overpowering desires are likely to arise here. My methodology somewhat limits my conclusion, because many of my claims are empirical and more evidence is needed to support them.

In particular, I claim that irresistible desires are likely only when the stakes in a decision are high, thus implicating primordial interests.[123] This claim seems justified as a conceptual matter, in that it is

122. This claim and the reasons supporting it derive from my general claim in part III.D.1. that irresistible desires typically arise only when the stakes in a decision are truly high.

123. Other empirical claims that need testing are: (1) that fear of radical and permanent

only things that people *care* about a lot that are likely to move them overpoweringly to act. Yet, although there are theoretical reasons to adopt my claim as a reasonable research proposal, research is needed to confirm it. Reality may defy logic, and overwhelming desires may turn out to be common in trivial situations.

Moreover, I rely on common-sense ideas of which interests are important and trivial. These may not match those of individual patients who may have idiosyncratic views of what is important. If many do, or if the ideas that seem common sense to me are themselves idiosyncratic, irresistible desires to refuse psychotropic medication may turn out to be more common than I suppose, even if the logic of my position is sound.

Arguments based on such empirical claims seem acceptable if there is a basis for the empirical claims. Moreover, empirical research in this area may be hampered by the very things that prevent evaluators from reliably determining when irresistible desires are present. If we cannot determine when irresistible wishes are occurring, it is difficult to establish whether they occur generally in situations that implicate important interests. At this point, logic, not research, may be the best course available.

Another prominent empirical assumption I make is that overpowering desires for a choice that someone else enacts can be protested, and thus will not be as irresistible as overpowering desires to act oneself. But perhaps many people feel such strong desires that they lose sight of their true goals, or perhaps many people with less strong desires are nonetheless self-deceived about their true goals.[124] Individuals are unlikely to protest things that they believe, without conflict, that they want.

In cases that conform to the pattern I have described, desires for choices will be less irresistible than desires for acts. Therefore, the problem of false negatives is at least smaller in this context. But empirical research is needed to determine how many cases conform to my pattern and how many are examples of the "no apparent conflict" pattern that I have suggested. Without such research, we will be unable to ascertain the magnitude of the problem of false negatives in the treatment context. Although problems with doing empirical

---

personality change as a result of psychotropic medication is not common or invulnerable to education; (2) that resistance is unlikely often to be compulsive; (3) that it is easier to protest the enactment of an irresistible desire than to resist enacting it oneself; and (4) that most people are likely to care more about the enactment of their choices than simply evidencing them.

124. Irresistible desires may be "smart." They may know that protest will defeat their purpose, and thus they may deceive their possessor about what she truly wants.

research in this area exist, some work is clearly needed to get a better sense of how far my empirical claims reach.

A second methodological limitation of my argument is that I focus principally on one prominent kind of volitional impairment: very strong desires to act or choose. In particular, I suggest that strong desires to make treatment choices may not as irresistibly lead to enactment as strong desires to commit crimes, because of the presence of another actor who must cooperate with one's choice. I direct many of my arguments, then, to a particular kind of volitional impairment; and they may not be applicable to other kinds. Indeed, the idea that a very manic or depressed patient cannot help acting as he does seems quite compelling, but capturing what precisely is wrong with his action may be quite difficult. A different formulation of his impairment may stand up well as a specific kind of volitional competency test. We need more work to give content to the concept of volitional impairment, and we shall have to evaluate new formulations as they become available.

Despite these limitations, I believe I have provided at least some reasons to reject the concept of an irresistible desire in the psychiatric treatment context. In combination with those offered in the criminal context, these reasons justify rejection of the volitional impairment test.

IV. THE PRODUCT OF MENTAL ILLNESS TEST

The product of mental illness test accords a more important role to mental illness per se than do any of the other tests in the areas of insanity and incompetency. Made famous by the case of *Durham v. United States*,[125] the product test says that a person is not guilty by reason of insanity if his crime was a product of mental illness.[126] Because the sense that mental illness makes a difference seems fairly compelling, the test appears very promising. Indeed, the presence of mental illness has lurked in the background of all the tests considered in this Article—it may stand behind and give force to them. Our very

125. 214 F.2d 862 (D.C. Cir. 1954). For articles on *Durham* and its progeny, see David C. Acheson, McDonald v. United States: *The* Durham *Rule Redefined*, 51 GEO. L.J. 580 (1963); John R. Cavanagh, *Problems of a Psychiatrist in Operating Under the* M'Naghten, Durham *and Model Penal Code Rules*, 45 MARQ. L. REV. 478 (1962); William O. Douglas, *The* Durham *Rule: A Meeting Ground for Lawyers and Psychiatrists*, 41 IOWA L. REV. 485 (1956); Abe Krash, *The* Durham *Rule and Judicial Administration of the Insanity Defense in the District of Columbia*, 70 YALE L.J. 905 (1961); Philip Q. Roche, Durham *and the Problem of Communication*, 29 TEMP. L.Q. 264 (1956); *Symposium on Insanity*, 22 U. CHI. L. REV. 317 (1955); *Symposium:* United States v. Brawner, 1973 WASH. U. L.Q. 17.

126. *Durham*, 214 F.2d at 874-75.

project was to think harder about mental illness and all the impairments it may cause.

A. *Criticisms of the Product Test in the Criminal Context*

Commentators criticized *Durham*'s product test,[127] and the court finally abandoned it, because it granted experts enormous discretion to condemn or exonerate by defining mental illness and its relation to the crime as they wished. A number of discrete problems plagued the test. First, the test failed to give guidance on the nature of exculpatory mental illness.[128] Thus a famous weekend vote at St. Elizabeth's Hospital resulted in experts thenceforth counting personality disorders as mental illnesses for purposes of the test.[129] The District of Columbia Court of Appeals attempted to rectify this problem by providing, in *McDonald v. United States*,[130] a legal definition of mental illness.[131]

Second, the *Durham* court did not provide a clear explanation of the "product" relationship. If "but for" causality was at issue, most

---

127. *See, e.g.*, Alan M. Dershowitz, *Psychiatry in the Legal Process: A Knife That Cuts Both Ways*, 51 JUDICATURE 370, 371 (1968); Krash, *supra* note 125, at 929-32; Henry Weihofen, *Detruding The Experts*, 1973 WASH. U. L.Q. 38, 47-49.

128. *See, e.g.*, GLUECK, *supra* note 81, at 95-98; GOLDSTEIN, *supra* note 64, at 84-86; Herbert Wechsler, *The Criteria of Criminal Responsibility*, 22 U. CHI. L. REV. 367, 368-69, 373 (1955).

129. *See* RALPH REISNER & CHRISTOPHER SLOBOGIN, LAW AND THE MENTAL HEALTH SYSTEM: CIVIL AND CRIMINAL ASPECTS 499 (2d ed., 1990).

130. 312 F.2d 847, 851 (D.C. Cir. 1962) (en banc).

131. *Id.* at 851. Some might think that providing a legal definition of mental illness, thereby taking the question away from the experts, is not the right course. Indeed, in certain contexts, some people trust experts to make complicated judgments more than they trust courts or juries. *E.g.*, REISNER & SLOBOGIN , *supra* note 129, at 402 (suggesting the use of experts in antitrust cases is more prevalent than in cases involving the insanity defense).

Most of the literature in the context of psychiatric treatment decisionmaking, however, argues on normative grounds that judges and juries are the better decisionmakers. *See, e.g.*, Robert S. Berger, *The Psychiatric Expert as Due Process Decisionmaker*, 33 BUFF. L. REV. 681, 681-85 (1984); Bruce J. Ennis & Thomas R. Litwack, *Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom*, 62 CAL. L. REV. 693, 694-96 (1974); Lois Forer, *Law and the Unreasonable Person*, 36 EMORY L.J. 181, 189-92 (1987). The issues in these cases are moral issues on which experts, in fact, are not particularly expert. *E.g.*, KARL MENNINGER, THE CRIME OF PUNISHMENT 139 (1968); Stephen Morse, *Failed Explanations and Criminal Responsibility: Experts and the Unconscious*, 68 VA. L. REV. 971, 1083-84 (1982); Morse, *supra* note 85, at 625-26. Experts also may have conflicts of interest that prevent the appropriate impartiality. *See, e.g.*, Bernard L. Diamond & David W. Louisell, *The Psychiatrist As An Expert Witness: Some Ruminations and Speculations*, 63 MICH. L. REV. 1335, 1344-45 (1965) (suggesting psychiatric experts in criminal cases often have strong relationships with attorneys that diminish the impartiality of psychiatric evidence); Daniel D. Pugh, *The Insanity Defense In Operation: A Practicing Psychiatrist Views* Durham *and* Brawner, 1973 WASH. U. L.Q. 87, 95 (new physicians initially find all felons insane until they experience managing an unselected group of felons in the hospital after criminal commitment); Weihofen, *supra* note 127, at 53 (suggesting hospital physicians fear that diagnosing a patient

mentally ill defendants would be insane. Experts could rarely testify that a crime would certainly have occurred even if the defendant had not been mentally ill.[132] If "but for" causality was not at issue, what relationship was?[133] Experts simply supplied their own concept of a "product," and provided conclusory testimony that a crime was or was not such a "product." The court prohibited experts from testifying on this ultimate issue in *Washington v. United States*.[134]

Durham's central shortcoming, however, was that it failed to offer guidance as to *how* mental illness was to compromise the defendant's action. Once again, experts supplied their own views.[135] Given this fundamental lack of guidance, *Durham* was bound to fail. Its mission was to permit experts scope to testify in terms meaningful to them, and thus to the jury.[136] But instead of providing detailed descriptions of the origin and nature of the defendant's mental functioning, experts usurped the jury's role by making moral judgments about who should and should not be exonerated,[137] testifying in con-

---

as insane may result in difficult patients being returned to the hospital after their civil commitment).

The Supreme Court, by contrast, seems increasingly willing to grant psychiatrists enormous authority over their patients. *See, e.g.*, Washington v. Harper, 494 U.S. 210, 236 (1990); Youngberg v. Romeo, 457 U.S. 307, 332-34 (1982); Parham v. J.R., 442 U.S. 584, 606-07 (1979). For example, the Court allows psychiatrists to make the findings required by a particular test without any review. *See Parham*, 442 U.S. at 606-07. The Court also seems to allow psychiatrists to set their own test in the exercise of "professional judgment." *See Youngberg*, 457 U.S. at 323-24. The Court has expressed the judgment that experts are better able to appreciate the issues and to make the most satisfactory accommodation of the interests involved. *See, e.g., Harper*, 494 U.S. at 229-31; *Youngberg*, 457 U.S. at 322-24; *Parham*, 442 U.S. at 607-12. I find the commentators' position more persuasive than the Court's.

132. *See, e.g.*, Edward de Grazia, *The Distinction of Being Mad*, 22 U. CHI. L. REV. 339, 343 (1955); Warren P. Hill, *The Psychological Realism of Thurman Arnold*, 22 U. CHI. L. REV. 377, 393-94 (1955); Pugh, *supra* note 131, at 95.

133. Some commentators have speculated that the court considered mental illness causality no differently than causality in other legal contexts. *E.g.*, Henry Weihofen, *The Flowering of New Hampshire*, 22 U. CHI. L. REV. 356, 359-60 (1955). Other commentators suggest that the court moved toward an heuristic test of causality. *See, e.g.*, Philip Q. Roche, *Criminality and Mental Illness—Two Faces of the Same Coin*, 22 U. CHI. L. REV. 320, 322-23 (1955). Other commentators suggest that "product test" causality was meant to ease the burdens found in prior insanity tests. *See, e.g.*, Herbert Wechsler, *The Criteria of Criminal Responsibility*, 22 U. CHI. L. REV. 367, 371 (1955).

134. 390 F.2d 444 (D.C. Cir. 1967).

135. *See, e.g.*, GOLDSTEIN, *supra* note 64, at 84-85. Indeed, the test did not even specify how much mental illness had to affect the act. *See, e.g.*, Krash, *supra* note 125, at 930-31 (citing the holding in Carter v. United States, 252 F.2d 608, 617 (D.C. Cir. 1957), which required that the illness "critically" or "decisively" affect the defendant's behavior, as evidence of *Durham*'s failure to define when mental illness should exonerate the defendant).

136. *See* de Grazia, *supra* note 132, at 342; Douglas, *supra* note 125, at 489; Manfred S. Guttmacher, *The Psychiatrist As An Expert Witness*, 22 U. CHI. L. REV. 325, 329 (1955); Krash, *supra* note 124, at 928.

137. That moral judgments would infect expert judgments under *Durham* is not surprising

clusory language tracking the terms of the test. *United States v. Brawner*[138] abandoned the *Durham* test and adopted the Model Penal Code test in its place.[139]

B. *Importing the Notion to the Treatment Competency Context*

The history of *Durham* should give pause to those considering a product of mental illness test of incompetency. But the test actually may be more acceptable in the competency context. First, expert dominance is less a concern when a jury is not the decisionmaker. Judges and hearing officers are perhaps more skeptical of experts. They are also likely to be more sensible of the legal nature of questions such as a patient's competency.[140] To the extent that we allow the experts themselves to make the decisions, they are simply performing their designated role when they apply the test as they see fit. They are guilty of usurpation then. Indeed, some may think that these *are* properly expert questions in this context.

Perhaps more important, we may make a *Durham*-type test more acceptable by remedying from the start its most glaring problems. If the core problem was the test's failure to offer the factfinder guidance as to *how* mental illness was to compromise the defendant's action,[141]

---

given the openness of the test. *See, e.g.*, Stephen J. Morse, *Undiminished Confusion in Diminished Capacity*, 75 J. CRIM. L. & CRIMINOLOGY 1, 36-38 (1984). Conclusory testimony is problematic precisely because it tends to usurp the role of the jury. *See, e.g.*, Weihofen, *supra* note 127, at 38-39; *see also* R.E. Schulman, *To Be Or Not To Be An Expert*, 1973 WASH. U. L.Q. 57, 64-65 (suggesting that courts abdicated their responsibility to properly define the decisionmaker by failing to define criminal responsibility).

138. 471 F.2d 969 (D.C. Cir. 1972) (en banc).

139. *Id.* at 973, 981-83.

140. But both judges and juries seem to defer fairly often to psychiatrists' judgments. *E.g.*, David B. Wexler & Stanley E. Scoville, *The Administration of Psychiatric Justice: Theory and Practice in Arizona*, 13 ARIZ. L. REV. 1, 60 (1971) (study found that 97.9% of judges concurred with the psychiatric expert in commitment hearings). Indeed, a number of commentators criticize *Durham* on the ground that it creates an environment in which triers of fact accept psychiatric expert opinion as definitive on issues of insanity. *See, e.g.*, Allen A. Bartholemew & Kerry L. Milte, *The Reliability and Validity of Psychiatric Diagnoses in Courts of Law*, 50 AUSTL. L.J. 450, 450, 454-55 (1976); Weihofen, *supra* note 127, at 43. If factfinders defer fairly often to psychiatrists, that calls into question the practical urgency of determining who is the better decisionmaker, and of removing the question from the experts if they are not. *See supra* note 131. Nevertheless, judges and juries sometimes refuse to listen to the experts and, to that extent, may protect some patients. Consider that in contested cases the opinions of one party's experts are not followed; and that some patients are not civilly committed despite expert testimony on the state's side and none on their own. Whether judges or juries are more skeptical of psychiatrists is hard to determine. But it seems likely that judges will have a better grasp than juries of the concept of a *legal* issue on which experts are not the final authority. *See supra* note 131.

141. By contrast, the cognitive and volitional tests of insanity specify the relevant effects of the mental illness on the act. Commentators were quick to point out that the *McDonald* court's definition of "mental illness" incorporated the cognitive and volitional impairments of

we may simply specify the necessary relationship between the illness and its product: a person is incompetent if his mental illness results in distorted beliefs or desires that are themselves the basis for his act or decision. I focus on distorted beliefs, because the case of distorted desires raises unnecessary complications.

This solution, of course, simply converts the *Durham* test into a cognitive test. Indeed, on any sensible view, the *Durham* test must become some other form of competency test. One might read it, for example, as a compulsion test. I have chosen a cognitive test because I find such tests the most plausible. Yet viewing *Durham* as an essentially cognitive test does not mean that we should simply dismiss it. Focusing on the role of mental illness in producing the cognitive distortions may be more productive than focusing on the distortions alone.

One might immediately object to this project: do not all cognitive competency tests cast mental illness in a causal role? The answer is no. Although formulations of the insanity test always require the presence of mental illness, some formulations of competency tests do not.[142] Perhaps the reference to mental illness is implicit. But it is striking that the language is often not present in competency tests.

Even if all such tests implicitly refer to mental illness, however, thinking about the role of mental illness in producing cognitive distortions promises to be worthwhile. If the product test, as I have construed it, is different from ordinary cognitive tests, then mental illness must play a different or more central role there. Establishing its unhelpfulness would then discredit the product test so understood. If, by contrast, the product test is little different from the typical cognitive test, then my argument that referring to mental illness is not helpful would apply to all such tests. We learn that competency tests that include a reference to mental illness are no improvement over those that do not. Indeed, our very project in this Article to consider competency standards that give a more central role to mental illness, and

---

the traditional tests. McDonald v. United States, 312 F.2d 847, 851 (D.C. Cir. 1962) (mental illness "includes any abnormal condition of the mind which substantially affects mental or emotional processes and substantially impairs behavior controls"); *see, e.g.*, GOLDSTEIN, supra note 64, at 86; REISNER & SLOBOGIN, *supra* note 129, at 499.

142. For testamentary capacity standards, see, for example, IDAHO CODE § 15-2-501 (1992); ILL. ANN. STAT. ch. 110 1/2, para. 4-1 (Smith-Hurd 1992); KY. REV. STAT. ANN. § 394.020 (Michie/Bobbs-Merrill 1992). For competency to contract standards, see, for example, Gulf Life Ins. Co. v. Wilson, 181 S.E.2d 914, 916 (Ga. Ct. App. 1971); Bach v. Hudson, 596 S.W.2d 673, 675-76 (Tex. Civ. App. 1980). On capacity to contract, see generally Alexander M. Meiklejohn, *Contractual and Donative Capacity*, 39 CASE W. RES. L. REV. 307 (1988-89).

all its disabling consequences, comes to seem questionable, even if taken on for perfectly understandable reasons.

Before considering the role of mental illness in producing cognitive distortions, we must attend to some preliminary problems with the proposal. Like the *Durham* rule itself, a rule that a person is incompetent if mental illness produces distorted beliefs that affect his decision or act makes too much turn on psychiatrists' conceptions of mental illness—with unfortunate consequences. The test must specify the nature of mental illness required, as the court saw clearly in *McDonald*. In this context, one prominent set of competency cases, the wills cases, classically cites people who have become perverse, mean, or suspicious with age as those whose freedom competency law is designed to protect.[143] Yet, contemporary psychiatry would probably say that many of these people suffer from some form of mental disorder. Some may be suffering from early-stage organic illnesses, such as Alzheimer's, while others may have crossed the line from "personality quirks" to a recognized disorder.[144] Indeed, the American Psychiatric Association labels such common occurrences as "caffeine intoxication" mental disorders.[145]

A competency test that uses psychiatry's concept of mental illness thus departs from our prereflective notions of competency. It is clearly necessary to limit the concept of mental illness used in the test. Limiting it to that of serious mental illness, such as psychosis, would seem a sensible solution.

Additionally, this test, unlike the law's, does not require serious distortions before it finds incompetency. Thus, it finds people incompetent even when there is some evidence for their belief, so long as the belief is a product of mental illness. For example, this test, but not the law's, would find incompetent a depressed person who felt that he was bad and deserved to suffer.[146] This view is problematic, however, because it trenches on people's freedom to arrive at the truth according to their own views, and thus violates the "unconventionality criterion" (standards ought to protect unconventional values and beliefs).[147]

---

143. *See* Saks, *supra* note 6, at 976.

144. For instance, a depressive person may develop true clinical depression, or a paranoid person may develop a true paranoid psychosis. *See supra* text accompanying note 61.

145. DSM-III-R, *supra* note 14, at 138-39.

146. This example is of a belief for which there is only slight evidence. But this rule arguably would find incompetent even a person whose belief had quite a bit of evidence, so long as psychiatrists were willing to say that the belief was in fact a product of mental illness.

147. *See* Saks, *supra* note 6, at 950. Later, I suggest that product theorists profess to know what impairs the person's ability to assess evidence—confusing inner reality with outer—and thus when the abilities criterion has been breached. See *infra* text accompanying notes 151-

I will not rehash the arguments supporting an "unconventionality criterion" here. If they are sound, we can simply read the product view to find people incompetent whose decisions are based on severely distorted beliefs that are a product of serious mental illness. The question is whether this test has any advantages over a severe cognitive distortion test that does not give mental illness so central a role.

C. *The Role of Mental Illness*

1. DOES THE PRESENCE OF MENTAL ILLNESS ADD ANYTHING TO THE COMPETENCY TEST?

Several theories suggest that the role of mental illness in producing distorted beliefs helps explain why we should deny legal respect to a person's choice. The most obvious theory is also the most implausible: because the mental illness *caused* the belief, one is not accountable for holding it, and should be protected from any harm it may cause. At least according to the soft determinists, however, the mere fact that mental illness caused the belief does not make one unaccountable.[148] If antecedent causation *is* sufficient for unaccountability,[149] as hard determinists hold, then arguably no one would be accountable for anything. Even if some events *are* uncaused, it is not clear why a person should be found incompetent when mental illness causes her distorted belief but not when a sleepless night or a gratuitous dressing-down by her boss does. Mental illness' causal role in producing beliefs contributes little to our understanding of incompetency.

Another view perhaps has more promise: when a person's belief

153. But invoking this theory is insufficient to justify finding incompetent those with mild distortions. Although the product theorist has a theory of why beliefs are false, he has no better way to tell when they are false than does a theorist who follows the law's standard. Consider an analogous competency theory: a person who holds beliefs due to inattentiveness because of a sleepless night is incompetent. The theory offers an explanation of why beliefs are distorted. But it provides no way to reliably determine when beliefs are distorted. This is precisely the situation in which an unconventionality criterion is necessary.

148. The soft determinist position maintains that "[w]e are fully responsible no matter what causes may exist for our behavior . . . ." MOORE, *supra* note 22, at 360-61.

149. *See, e.g.*, Roderick M. Chisolm, *J.L. Austin's Philosophical Papers*, *in* FREE WILL AND DETERMINISM 339-45 (Bernard Berofsky ed., 1966); John Hospers, *Free Will and Psychoanalysis*, *in* REASON AND RESPONSIBILITY: READINGS IN SOME BASIC PROBLEMS IN PHILOSOPHY 355-64 (Joel Feinberg ed., 4th ed. 1978). For articles discussing the free will-determinism problem, see generally Michael Corrado, *Automatism and the Theory of Action*, 39 EMORY L.J. 1191 (1990); Michael S. Moore, *Causation and the Excuses*, 73 CAL. L. REV. 1091 (1985); Benjamin B. Sendor, *Crime as Communication: An Interpretive Theory of the Insanity Defense and the Mental Elements of Crime*, 74 GEO. L.J. 1371 (1986); John L. Hill, Note, *Freedom, Determinism, and the Externalization of Responsibility in the Law: A Philosophical Analysis*, 76 GEO. L.J. 2045 (1988).

is a "product" of mental illness, he has no *choice* in holding it, and thus should not be held to its consequences. This view is not simply a restatement of the last: that mental illness *causes* the belief. Rather, it says that mental illness *compels* the belief. This view amounts to an irresistible desire test with regard to the decisionmaker's beliefs (rather than her acts or decisions). The test's lack of success in the context of acts and decisions suggests a similar outcome in the context of beliefs. The notion of a compelled belief seems very difficult to analyze,[150] and application problems are likely to be equally as severe.

A third view of the significance of mental illness' producing the belief is that a distorted belief that mental illness produces is not the *person's own* true belief. This is just the different person test applied to the decisionmaker's beliefs, rather than her choices. The test fails here as surely as it failed there, and for the same reason: evaluators will not be able to reliably apply the test.

A fourth view is somewhat different: diseased beliefs are significant precisely *because* they are diseased. Diseases are seriously substandard, undesirable by definition, and in need of correction. Yet, although one may feel an impulse to correct a diseased belief as much as, for example, diseased breathing, the criteria for when to offer help and when to interfere with a patient's choice are surely not the same. Interference on the basis of disease, without more, could reflect mere prejudice—similar to excluding from a race a runner with a congenitally lame leg, but not a runner whose overexertion produced a similar strain. One might offer the two different treatment, but nothing else would be warranted. The mere fact that disease is undesirable does not justify finding incompetent a person with a diseased belief.

The last theory explaining the significance of mental illness in producing the belief is the most plausible: it is not that diseases are substandard and undesirable, but that their influence on a belief signals that it has been arrived at in the "wrong" way. A belief produced by mental illness is infected by internal needs and wants. It is not a pure response to the evidence. Consider that psychiatrists understand paranoid beliefs to result from the person's projecting his

150. As discussed previously, one view holds that a belief is compelled if one cannot *help* holding it, and, on this view, many run-of-the-mill beliefs of the non-ill are compelled. *See supra* note 116. Another view suggests that a belief is compelled only when inner or outer pressures force one's attention away from the external evidence. *Id.* Yet, although many delusional beliefs are exactly of this kind, some are experienced as temptations to believe, rather than as beliefs one cannot help believing. *See, e.g.*, THE NEW HARVARD GUIDE TO PSYCHIATRY, *supra* note 30, at 268. If the whole topic of compulsion is unclear, the topic of compelled beliefs is even more so.

inner hostility outward.[151] The person is unconsciously angry at his mother, but consciously perceives that his mother is trying to hurt him. Consider, too, the phenomenon of denial: a person has a psychological need to feel invulnerable and in control, and so he denies that he has a serious illness.[152]

In short, people with delusional beliefs are not responding solely to the evidence presented to their senses and reason. Their beliefs are also a function of inner needs, wishes, and fears, which, by defensive maneuvers of the mind, are expressed in a disguised form. The person simply confuses inner with external reality.

At this point, however, it becomes clear that considering the role of mental illness in producing the belief does not further the inquiry.[153] Three of the theories of the significance of mental illness track the terms of *other* competency tests, but simply apply them to distorted beliefs, instead of decisions or acts. The "cause" view is overbroad, because it applies to all mental phenomena, and the "disease" view focuses on something irrelevant to the inquiry.

The view I find most plausible—that the presence of disease means that the decisionmaker has arrived at the belief in the wrong way—seems not to advance the inquiry. The rationale for denying people's choices legal respect remains the same: that their ability to assess evidence is impaired. Mental illness, on this view, may explain *how* one's ability to assess evidence is impaired, but it is *that* one's ability is impaired that justifies the finding of incompetency. In short, mental illness has some explanatory value on this view, but not much.

2. DOES MENTAL ILLNESS HAVE A TRUE EXPLANATORY ROLE IN CONCEPTUALIZING IMPAIRED ABILITY TO ASSESS EVIDENCE?

Nevertheless, mental illness' explanatory role in conceptualizing impaired ability to assess evidence is not *nothing*. To that extent, the product standard may improve on the law's. But I shall now suggest

151. *See* Sigmund Freud, *Psycho-analytic Notes on an Autobiographical Account of a Case of Paranoia (Dementia Paranoides)*, *in* 12 THE STANDARD EDITION OF THE COMPLETE WORKS OF SIGMUND FREUD (James Strachey trans., 1958); *see also* TALBOTT ET AL., *supra* note 82, at 137 (describing mechanism of projection).

152. *See, e.g.*, TALBOTT ET AL., *supra* note 82, at 137-38 (describing mechanism of denial).

153. Other commentators have argued, as noted, that the product test offers no guidance as to how an act's production by mental illness impairs the act and that the test must therefore be interpreted as another competency test. The arguments in the text regarding mental illness' significance in producing the distorted belief could apply equally to its significance in producing the choice or act—and commentators have in fact *made* these arguments in that context. Yet, responding to this problem by specifying that the mental illness produces a distorted belief simply recreates the problem, and is no less resistant to a solution.

that mental illness' explanatory role in this theory is illusory. The idea that distorted beliefs are infected by inner needs and wants, thereby confusing inner reality with outer, must in practice reduce to something like the law's view: that distorted beliefs are beliefs insufficiently supported by the evidence.[154]

Is there in fact any way to describe severely distorted beliefs that does not advert to their lack of support in the evidence? The idea that inner needs and fears infect the beliefs seems to refer implicitly to the external evidence. How can one identify what is inner without identifying what is outer? Indeed, one must refer to the external evidence not only to establish what is inner, but also to establish that what is inner played an impermissible role in producing the belief. If the external evidence supports the belief, the inner needs and wants are simply irrelevant—they have not produced any distortions. The point is that we can determine whether they have only by looking at the external evidence.

Perhaps if the inner needs and fears that infect one's beliefs are themselves archaic or primitive, as unconscious phenomena often are, we can discredit the belief without having to look at its support in the external evidence. It appears that we can—but only because the archaic, primitive unconscious beliefs are themselves distorted. Yet one can determine *this* only by seeing how well the evidence supports these unconscious beliefs. Archaic, primitive beliefs are precisely those that depart wildly from the evidence. Thus, invoking their influence simply transfers the inquiry, in the first instance, to the unconscious, but still asks the same question: whether the evidence supports the beliefs.

If this is correct, the role of mental illness in producing distorted beliefs contributes little to an understanding of incompetency. In theory, it is the distortions that count. In practice, identifying distortions must proceed by looking at the external evidence, as the law's test requires. The idea that inner needs and fears divert one from the external evidence provides an explanatory framework, but the influence of those needs and fears can be identified only by looking again at the external evidence.

The explanatory and practical limitations of the mental illness account are purely contingent. In the future, psychiatrists and psychologists may be able to provide a mental illness-based means of identifying and explaining distorted beliefs. For example, sophisticated physiological tests may reveal when mental illness is working

154. *See* Saks, *supra* note 6, at 966-77.

its distorting effects. Even if so, it may still be easier to look at the evidence than to try to plumb the depths of a person's unconscious mind to find these distorting influences. If scientists *can* someday develop efficient and reliable tests to identify problematic beliefs, a suitably refined product test may become more attractive than the law's. But that day does not seem close at hand.

3. IS MENTAL ILLNESS REALLY IRRELEVANT?

Despite my arguments, one may retain a strong intuition that mental illness *should* play an important role in our competency judgments. Take the following case: two people have the same distorted belief. Person A's belief is a result of his mental illness, over which he has no control; Person B's belief is a result of his adherence to a particular religious faith. Society tends to want to hold Person A nonaccountable for acts or decisions based on his belief, but not Person B.

Consider an even clearer case: Person A's belief is a result of his mental illness, over which he has no control; Person C's belief is a result of his intoxicated state. If Person C voluntarily became intoxicated, society wants to hold him responsible for any crimes he commits, even if his distorted belief contributed to them.[155] The point is that he was responsible for becoming impaired, even if his impairment then attenuated his responsibility for the acts it underlay.

These examples, especially the second, may appear to highlight a role for mental illness in an insanity or incompetency standard: it signals that the agent is in no way responsible for his distorted belief. By contrast, the religious adherent could have refrained from pursuing the fanatical religion, and the drunkard could have refrained from drinking. Arguably they had some control over the ultimate development of their distorted belief. The mentally ill person had none.

Now in some sense, I think, this merely reintroduces a causal or compulsive role for mental illness. Once one becomes committed to a fanatical religion, or intoxicated, one's current mental state may as clearly cause or compel one's distorted belief as mental illness does. But one could have done something in the past to avoid getting in the state. This is not true for the mentally ill person, who at no point escapes a circle of compulsion or causation.

This view is problematic, because the distinction in terms of past control between mental illness on the one hand, and adherence to a religion or being intoxicated on the other, is hardly sharp. For instance, a person may end up indoctrinated into fanatical religious

155. This is certainly the law's position. *See, e.g.*, Foster v. State, 374 S.E.2d 188 (Ga. 1988); *In re* Matherly, 354 S.E.2d 603 (W. Va. 1987).

belief for reasons out of her control. Perhaps her parents subjected her to intense training from a young age or she joined a religious group without being aware of its techniques of indoctrination or the true content of its faith. Similarly, a person may lose the ability to refrain from becoming intoxicated via a route for which society does not assign culpability. By contrast, to the extent one believes causation or compulsion escapable, some mentally ill people may voluntarily take steps that exacerbate their mental condition.

Of course, many may believe causation is not escapable: thus I do not need to invoke special examples of this kind. Pushing the inquiry back to a time in the past when a person chose to drink no more allows one to distinguish between cases that permit control and those that do not than looking at the immediate antecedents of the belief themselves. The person lacked real power to choose whether to drink even at that point. On this view, it is a mistake to consider mental illness different from any other cause of a belief or action.

To see where this view of the role of mental illness leads, I will assume the common-sense distinction between a choice one makes because one is intoxicated and a choice one makes because one is mentally ill, even if both people are equally impaired at the time of the choice. The distinction again is in terms of the person's ability to control his state at some time in the past. The issue is whether this factor is relevant to the concept of *treatment competency*.

It may not be. In cases of retrospective evaluation of a person's responsibility or accountability, his past control over his current state is clearly relevant. He is blameworthy if he failed to take steps to protect against this kind of harm, and the prospect of future sanctions may provide him with incentives to take these steps on other occasions. By contrast, if we are *prospectively* evaluating a choice, we have much less reason to care about what steps a person could or could not have taken in the past. The person's impairment is reason enough to proscribe his choice or action. Even in the case of retrospective evaluations, if the issue is competency rather than responsibility, we may not care whether the person had control over the impairment that compromises his choice. For example, a will written during a drunken stupor may be void.[156]

Thus, mental illness may play a role in responsibility evaluations, although even in that case, its role is somewhat problematic. But the role of mental illness seems irrelevant, at least in this way, in treatment competency evaluations. I do not deny that the presence of

156. *E.g.*, *In re* Estate of Fleege, 230 N.W.2d 230 (S.D. 1975); *In re* Estate of Rhodes, 436 S.W.2d 429 (Tenn. 1968).

mental illness may serve an evidentiary role.[157] Nor do I deny that we have a deep conviction of its significance as a non-evidentiary matter. It simply seems very uncertain *why*.

D. *Conclusion*

The product test's focusing on the role of mental illness does not seem to advance the inquiry into competency. Speaking of the illness producing the crime or treatment choice is too vague and, in fact, amounts to a non-test. Focusing on the illness producing a distorted belief is an improvement. Yet the effort to specify more clearly *how* it helps leads nowhere. Talk of pure causation is not to the point because all beliefs are caused by *something*. Talk of the diseased nature of the belief is unhelpful because the presence of a disease as such is not relevant to a decision to deny someone respect. Talk of compulsion, different person-ness, or distorted beliefs simply reintroduces into the product standard tests considered more directly elsewhere. Indeed, to talk in these terms is to refer to particular impairments that the illness causes, in which case, we seem not to need the illness itself. Perhaps the presence or absence of mental illness is important in retrospective evaluations of responsibility, although even then one might refer to impaired or intact control in the past. But it does not seem important to a concept of treatment competency.

The best of these theories, in my view, is the theory that the illness explains *how* distortion occurs. But the idea of the illness' role in producing the distortion failed to make any real advance. The law's test, which refers to the external evidence, appears superior.

Thus, a product of mental illness standard, to the extent it is different from the law's, does not seem an advance over the law's. Indeed, it may raise significant problems of its own. Although psychiatrists might prefer a product standard because it speaks a language that they know—the language of mental illness and its effects—the standard may be subject to greater abuse than the law's. It may be both too easy to claim that mental illness produced a belief and too hard to dispute such a claim.

It is important to stress, however, that the product standard and the law's standard are actually very close. In fact, the product standard, so understood, is essentially a cognitive standard in disguise.[158]

157. Because of an illness' evidentiary role, it may be advisable for competency tests to include a reference to mental illness, even if only for this limited purpose.

158. Indeed, it is the "gross inability to assess evidence standard" that I defended in my last competency article. *See* Saks, *supra* note 6.

It is possible that a different interpretation of the product standard, such as one that casts it in terms of distorted *desires* caused by mental illness, would be more successful. But this interpretation is not successful. Most importantly, the considerations that justify a cognitive reading of the product standard are all "inability to assess evidence" considerations. Giving a more central role to mental illness, therefore, does not further the analysis, and may present unnecessary complications. A detour into the language of mental illness seems unnecessary and unwise.

V. CONCLUSION

The law's treatment competency standard has survived comparison with three standards that give a more central role to mental illness and all of its disabling effects. A different person standard seems attractive as an ideal matter because, if the mentally ill self is as if a different person from his healthy counterpart, he is not in an adequate position to decide for that self. Nevertheless, the theory fails as applied to the chronically ill, and is too difficult to apply to the acutely ill, because mentally ill choices are too difficult to identify reliably. Even if evaluators came to be able to reliably apply the test, we would need more conceptual clarification of the notion of retrospective repudiation, as well as linkages of that notion with different person-ness, in order for the different person theory to be completely satisfactory.

A volitional impairment standard, also attractive as an ideal matter, is problematic in application, and unlikely to apply to many psychiatric treatment choices in any case. Most people will not have irresistible desires to refuse medication because the interests involved in the decision are not that weighty. In addition, the formal features of treatment choices are such that wishes irresistibly leading to enactment are unlikely.

A product of mental illness standard must be so qualified that it amounts to little more than a "patent inability to assess evidence" standard. The test's focus on the role of mental illness in no way advances our understanding of competency, and may be more subject to abuse than the law's. Thus, none of these standards is an improvement over the law's.

One interesting result of my analysis is that it appears to reach beyond the area of treatment competency. My analysis of the volitional impairment test, it is true, turns on issues quite specific to that context—the features of medication choices that make irresistible desires unlikely there and the formal features of the treatment-choice

context. Similarly, mental illness may play a role in the context of criminal responsibility that it does not play in the competency context. But for the most part, my analysis of the different person test and the product of mental illness test does not depend on specific features of the treatment context. It appears that the criticisms of the product and different person tests are general, warranting the conclusion that neither test is appropriate to any competency area.

A remaining issue is whether it even matters which competency standard is used. In the criminal context, some evidence suggests that the standard does not greatly affect outcomes.[159] For example, the rate of insanity acquittals did not significantly rise after the *Durham* test was adopted.[160] If the competency standard used does not affect outcomes, comparing different competency standards may seem a particularly arid exercise.

But to those small number of people acquitted under the new criminal law test, adopting the test mattered considerably. In addition, there is good reason to expect that more liberal tests would matter more in the treatment competency context than in the criminal. In the latter context, checks against high rates of insanity exist in people's conflicts over such findings. Criminal behavior produces fear and anger. Even when there is evidence that the criminal was impaired, some may want to exact a penalty of him—may want the perpetrator of harm in effect to suffer harm himself. If so, factfinders may be likely to have strong resistances to finding insanity. Indeed, evaluators themselves may be conflicted about such findings for many of the same reasons. Human nature provides some natural checks against high insanity rates.

No such checks would seem to exist in the treatment competency

159. Studies conducted with mock jurors asked to render a verdict under different insanity tests suggest that the test may not be dispositive. *See, e.g.*, RITA J. SIMON, THE JURY AND THE DEFENSE OF INSANITY 213-15 (1967) (no significant difference in jury verdicts between *M'Naughten* and *Durham* tests); Norman J. Finkel et al., *Insanity Defenses: From the Jurors' Perspective*, 9 LAW & PSYCHOL. REV. 77 (1985) (jurors' decisions concerning the insanity defense were not significantly different among six insanity tests).

160. *See* Krash, *supra* note 125, at 948-50 (acquittal rates in the District of Columbia five years after *Durham* rose from .023% to 1.33%, or from three of 2103 criminal defendants to 17 of 1714 criminal defendants). One may think that the acquittal rate is not that important because different insanity tests may create different incentives for prosecutors in their charging and prosecution methods. For example, if, under a new test, the acquittal rate remains steady but significantly fewer charges are brought, then the new test will have had quite a big effect. But it seems unlikely that a liberal test would have this kind of effect. Such a test is likely to dispose prosecutors to acquiesce more readily in an insanity disposition rather than contest the case, not to simply drop it, and uncontested insanity cases are counted as part of the acquittal rate. Still, unexpected effects on prosecutor behavior may diminish the significance of the relatively steady acquittal rate.

context. Most psychiatrists seem relatively unconflicted about treating mentally ill patients whom they perceive to need treatment, even if this means forcing treatment on them. Studies show that, when physicians in general err, they almost always err on the side of making a diagnosis when none is warranted. The theory is that it is better to treat someone who is not sick than to fail to treat someone who is.[161] As a result, physicians also tend to err on the side of recommending unnecessary or unhelpful treatment—of overtreating.[162]

Diagnosing and recommending treatment is one thing, while forcing treatment is quite another. But that professional psychiatric organizations, when they intervene in court cases, always argue against a right of competent patients to refuse treatment only underscores the profession's commitment to treatment even in the face of reasoned objection.[163] In short, psychiatrists are trained to be healers, and they seem relatively unconflicted about fulfilling their calling.

Psychiatrists play a major role in competency determinations. In some places, they make the findings themselves. There are then no jurors who might be skeptical of psychiatry or able to imagine themselves in the shoes of the patient to consider the psychiatrists' decisions.[164] In other contexts, lay factfinders make the ultimate decision.[165] But such factfinders, generally deferential to experts anyway, may be even more so in the treatment context. If this is so, standards giving psychiatrists more scope to find incompetency, as the standards considered in this Article plainly do, are likely to result in more findings of incompetency. Checks of the kind that exist in the criminal context simply will not operate to limit the numbers in the

---

161. *See, e.g.*, James W. Ellis, *Volunteering Children: Parental Commitment of Minors to Mental Institutions*, 62 CAL. L. REV. 840, 865-67 (1974); Morse, *supra* note 85, at 556.

162. *See* THOMAS S. SZASZ, PSYCHIATRIC JUSTICE 72-74 (1965). Unnecessary treatment may also result from fear of medical malpractice. *E.g.*, Robert C. Macaulay, Jr., *Health Care Containment and Medical Malpractice: On A Collision Course*, 21 SUFFOLK U. L. REV. 91, 91 (1986). Specialized practice groups in medicine may also increase the likelihood that patients become overtreated, at least to the extent that these experts disagree as to the appropriate treatment and patients choose multiple routes to find a cure. *See, e.g.*, KATZ, *supra* note 19, at 189.

163. *See, e.g.*, Brief for the American Psychiatric Association and the Washington State Psychiatric Association as Amici Curiae at 5, Harper v. State, 759 P.2d 358 (Wash. 1988) (No. 88-599), *rev'd sub nom.* Washington v. Harper, 494 U.S. 210 (1990) (American Psychiatric Association "has participated in almost every federal appellate case presenting similar questions"—namely, the question of when objecting patients may be administered antipsychotic drugs).

164. *E.g.*, Williams v. Wilzack, 573 A.2d 809, 813 (Md. 1990). Experts are also the decisionmakers in other contexts, although the substantive standard is not one of competency. *E.g.*, Rennie v. Klein, 653 F.2d 836, 848-51 (3d Cir. 1981).

165. *See, e.g.*, Rogers v. Commissioner of Dep't of Mental Health, 458 N.E.2d 308, 313-14 (Mass. 1983); Rivers v. Katz, 495 N.E.2d 337, 343-44 (N.Y. 1986).

competency context.[166]

If this is so, the treatment competency standard adopted should have an important effect on outcomes. Considering different competency standards may also further theory in important ways. There is value in understanding, as a conceptual matter, the virtues and vices of different standards. For example, studying different standards may force us to consider the purposes of competency doctrine, as well as the presuppositions about human nature that underlie it. Similarly, it may encourage us to think more clearly about just what abilities are necessary for a decision to deserve deference. It may also lead us to think about practical problems, such as those of application, and thus present a clearer picture of what makes a test good. Finally, it may raise related issues that deserve more study, such as the role of mental illness in any competency standard. My project confirms the attractiveness of the law's standard, as well as achieving some clarity on the conceptual matters that a study of competency raises.

166. I am not suggesting that psychiatrists will *lie* in their incompetency findings in order to achieve the desired outcome of treatment. But because the tests permit more incompetency findings, or are so difficult to apply that they do not rule them out, the psychiatrists desiring the outcomes will not hesitate to make the findings. The result may be that more patients who are in fact not incompetent will be barred from making their own treatment choices.