THE OXFORD HANDBOOK OF

PHILOSOPHY OF CRIMINAL LAW

Edited by

JOHN DEIGH AND DAVID DOLINKO

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2011 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
The Oxford handbook of philosophy of the criminal law / edited by John Deigh and David Dolinko.
p. cm.
Includes index.
ISBN 978-0-19-531485-4 (alk. paper)
1. Criminal law—Philosophy. I. Deigh, John. II. Dolinko, David.
III. Title: Handbook of philosophy of the criminal law.
K5018.O96 2011
345'.001—dc22 2010023180

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

CHAPTER 12

INSANITY DEFENSES

WALTER SINNOTT-ARMSTRONG
AND KEN LEVY

In 1843, Daniel M'Naghten believed that the Tories, including prime minister Robert Peel, were plotting to destroy him and that the only way to defend himself was to kill the prime minister first. M'Naghten shot into the prime minister's carriage and killed the passenger, who turned out to be the prime minister's private secretary, Edward Drummond. M'Naghten's act clearly qualified as murder; he intentionally killed a human being with malice aforethought. M'Naghten did think that his act was justified as self-defense, but the threat was both imagined and not immediate, so he could not cite self-defense as a defense. Nonetheless, the jury did not find him guilty. They instead found him not guilty by reason of insanity. He was committed for the rest of his life to a mental institution, but he did not have to go to prison at all.[1]

This verdict outraged many people in 1843, just as happened when John Hinckley was found not guilty by reason of insanity after shooting President Reagan in 1981.[2] Such (in)famous cases create the impression that political assassins can get off just by pleading insanity. That impression, however, is inaccurate because many insanity pleas are not successful. In 1963, Jack Ruby pleaded insanity but was convicted of murdering Lee Harvey Oswald, the presumed assassin of John Fitzgerald Kennedy. In 1968, Sirhan Sirhan pleaded diminished responsibility but was convicted of murdering Robert Kennedy. In 1972, Arthur Bremer pleaded insanity but was convicted of attempting to assassinate Alabama governor George Wallace. So the insanity defense is hardly a free ticket home.

Another reason why the insanity defense is not a free ticket home is that people who are found not guilty by reason of insanity are not sent home. They are locked up in mental institutions, and they sometimes spend more time in these mental institutions than they would have spent in prison if they had been found guilty.

In 1983, the Supreme Court held that someone who was found not guilty by reason of insanity could be held indefinitely even though he had been charged only with attempted shoplifting, which called for a maximum sentence of one year.[3] Thus, the verdict "not guilty by reason of insanity" should not always be welcomed by defendants or feared by citizens.

Still, successful insanity defenses are often controversial. On the one hand, when a defendant is known to have intentionally caused a death, it strikes many as unjust to find that person not guilty. On the other hand, it strikes many as unfair to find someone guilty of committing an offense for which he was not morally responsible. This conflict between intuitions lies at the heart of the insanity defense.

Is Insanity Medical or Legal?

To understand this controversy, it is crucial to ask what kind of concept insanity is. It is common to think that insanity is a medical condition. Psychiatrists, however, almost never describe their patients as "insane" or "sane." It would not help them in diagnosis or treatment to employ this dichotomy. Instead, they use such diagnostic categories as schizophrenia, paranoid delusion, kleptomania, and borderline personality disorder to decide what is wrong with their patients and how to treat them. It is the *judges and lawyers* who have to decide who is insane and which mental conditions make someone insane. The law classifies some people as sane and others as insane in order to determine who should be held criminally responsible, competent to stand trial, capable of handling their own financial affairs, or capable of living without supervision. In this way, insanity is a legal concept.[4]

Compare vision.[5] Optometrists determine whether one's eyesight is 20/20, 20/40, 20/80, or 20/200 (corrected or uncorrected, in bright or low light); whether one is nearsighted or farsighted; and which colors one is able to detect. But even with the most detailed diagnosis of an individual, lawmakers still need to decide how good someone's vision needs to be in order to obtain a license to drive a car or a bus or to fly a plane or how bad vision must be in order to qualify for disability benefits or insurance payments—and so on for other legal purposes.

Similarly, although psychiatrists are best qualified to determine a person's mental condition, lawmakers still need to decide whether that mental condition removes legal responsibility or some other legal status. Where the law draws the line between sanity and insanity depends on particular contexts and purposes. The law might draw one line between those who are competent to stand trial and those who are not and a different line between those who need to be civilly committed to mental institutions and those who do not.

Because the insanity defense is concerned with criminal responsibility and punishment, the crucial question here is which kinds of mental conditions do or

should remove criminal responsibility. Different formulations of the insanity defense specify different mental conditions for this role.

The M'Naghten Rule

"Madness" became a complete defense to criminal charges as early as the time of Edward III in fourteenth-century England.[6] In the sixteenth century, a prominent legal treatise specified "knowledge of good or evil" as the test of sanity and responsibility.[7] Courts proceeded to adopt a variety of insanity defenses until most courts settled on the rule in M'Naghten's case in 1843:

> To establish a defense on the grounds of insanity, it must be clearly proved that, at the time of committing the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know that what he was doing was wrong.[8]

This rule places a heavy burden of proof on the defense, distinguishes between two kinds of knowledge, and requires a mental disease to cause a defect of reason.

Although this rule refers to knowledge, what really matters here seems to be true belief rather than knowledge. If knowledge is defined as justified true belief plus some Gettier condition,[9] then agents can lack knowledge by (1) lacking any belief, (2) having false beliefs, (3) lacking justification, or (4) failing a Gettier condition. It would be inappropriate to excuse someone with true beliefs simply because those beliefs are not justified or fail some Gettier condition. Thus, it is (1) and (2) rather than (3) or (4) that make defendants lack knowledge in the way that removes responsibility under the M'Naghten rule.

The M'Naghten rule's distinction between knowing the nature and quality of the act and knowing that it is wrong also needs to be clarified. A classic example of not knowing "the nature and quality of the act" is a husband's strangling his wife while believing that he is squeezing juice from a lemon. This agent does not know that he is killing anyone. Other examples are harder to classify. Consider Joy Baker, who shot and killed her aunt while deluded into thinking that her aunt was the devil and had come to hurt her.[10] Baker seems to have known that she was shooting a person,[11] but she did not know that her act was not necessary to prevent harm to herself.[12] Thus, Baker knew part, but not all, of "the nature and quality of her act."

Of course, no agent knows every quality of any act, so courts need to decide which qualities of an act are essential. In order to limit which qualities of an act must be known for responsibility and therefore which delusions excuse, the judges in M'Naghten's case proposed a counterfactual test:

> [If] he labours under such partial delusion only, and is not in other respects insane, we think he must be considered in the same situation as to responsibility

> as if the facts with respect to which the delusion exists were real. For example, if under the influence of his delusion, he supposes another man to be in the act of attempting to take away his life, and he kills that man, as he supposes, in self-defense, he would be exempt from punishment. If his delusion was that the deceased had inflicted a serious injury to his character and fortune, and he killed him in revenge for such supposed injury, he would be liable to punishment.[13]

In the former case, the defendant is honestly motivated by self-preservation, the motivation that is required for self-defense. In the latter case, the defendant's motivation is not self-preservation but revenge, which has never been a recognized defense. So even if he were not deluded but correct in his beliefs, his act still would not be justified.

This counterfactual test, however, might seem to rule out too much. Consider Howard Barton Unruh, who killed thirteen people in Camden, New Jersey, on September 6, 1949, because he was convinced that his neighbors "ha[d] been making derogatory remarks about [his] character." He even imagined that actress Barbara Stanwyck was one of his hated neighbors. What finally set him off was his discovery that someone had stolen the gate to his fence. Unruh was never tried for the massacre. Instead, he was diagnosed as a paranoid schizophrenic, pronounced insane, and put in a unit for the criminally insane. While Unruh's delusion seems to be the very kind of mental illness that is supposed to remove responsibility, at least according to many people at the time and many proponents of the insanity defense, he still seems to be responsible under the counterfactual test because his acts would not have been justified even if his beliefs about his neighbors had been true. In this case, then, that added clause seems to yield the wrong verdict.[14] Thus, it remains controversial and difficult to specify precisely which parts of the nature and quality of the act the defendant must know in order to remain responsible for her actions under the M'Naghten rule.

When we turn our attention to the other kind of knowledge mentioned in the M'Naghten rule—knowledge that the act was wrong—it becomes crucial to determine which kind of wrongness the agent needs to know in order to be criminally responsible under the M'Naghten rule. There are four main possibilities.

The first possibility is that a responsible agent needs to know that the act is illegal or criminal—that is, *legally* wrong. To know this is to know something about the laws in the particular jurisdiction and how they apply to the act.

The second possibility is that a responsible agent needs to know that the act is contrary to the moral beliefs of most people in the particular society—that is, *socially wrong*. To call an act socially wrong in this sense is to refer not merely to custom or etiquette but, instead, to moral beliefs and principles generally accepted in that community. In order for a defendant to know that an act is socially wrong, then, she must know something about what people in a given society generally believe about morality.

A third possibility is that a responsible agent needs to know that the act violates that particular agent's own moral principles or moral beliefs—that is, that it is *personally wrong*. In order for a defendant to know what is personally wrong, she must be aware of her own moral beliefs and how to apply them.

Finally, a responsible agent might need to know that the act is just plain *morally* wrong. For a defendant to know this is not for the defendant to know what other people do or would say or believe about the act or about its moral status.[15] Instead, it is to know something about the act itself—namely, that there is at least one property of the act that gives it the moral status of being wrong. For example, a defendant cannot know that rape is just plain morally wrong unless the defendant knows that rape has some property that makes it morally wrong—for example, that it causes severe and unjustifiable harm to the victim. If the defendant believes that rape is wrong only because it happens to violate laws in a certain legal system or only because other people believe that it is wrong, then that defendant does not really understand what makes rape morally wrong and therefore does not know that rape is just plain morally wrong.

Knowing that an act is morally wrong in this fourth sense does not require knowing the law or the moral beliefs of society or of the agent. These beliefs are all distinct from the moral status of an act because the law, society, and the agent can all be incorrect. That is shown by cases like slavery, which can be seen as morally wrong even at times when it did not violate the law or the moral beliefs of society or of many individuals. Although many acts that are morally wrong will also be legally, socially, and personally wrong, these notions of wrongness come apart when individuals or societies hold deviant moral beliefs.

This fourth notion of moral wrongness might seem inappropriate for a legal test of insanity because it cannot be applied without making or assuming a moral judgment. To apply the first three tests, judges and juries need only do empirical research to determine what the law is or which moral standards are accepted by society or by this defendant. They do not need to make any normative judgments about whether or not those laws or standards are justified or correct. In contrast, in order to apply the fourth test to a particular defendant, juries or judges would have to ask (1) whether or not the defendant's act was immoral, (2) what makes it immoral, and (3) whether or not the defendant knew (or was able to know) that the act was immoral on that basis. These questions are normative, but the answer is usually not controversial, at least in cases involving core offenses, such as murder, rape, kidnapping, and theft. In any case, the judge or jury's decision will rest on its own moral principles rather than on its understanding of the moral principles of society or of the defendant.

Opponents might object that it would be unfair for a judge or jury to condemn a defendant merely because the defendant did not happen to share the judge's or jury's own beliefs about what is morally wrong. But if the law explicitly or implicitly directs the jury or judge to determine whether or not the defendant knows that an act is morally wrong, then, because knowledge implies truth,[16] the jury or judge cannot apply the law as stated without determining whether or not the act is *in fact* morally wrong. The law, unless redefined explicitly, then seems to authorize the jury or judge to apply her own views about what counts as moral knowledge and truth.[17]

M'Naghten jurisdictions do not agree about which kind of wrongness must be known in order for an agent to be responsible. Most seem to have remained silent,

and at least two have explicitly refrained from adopting a position, on this issue.[18] Regarding the jurisdictions that have taken a position, some of them maintain that defendants may generally be found not guilty by reason of insanity only if, as a result of mental illness, they did not know that their acts were *legally* wrong.[19] Other jurisdictions explicitly specify that legal knowledge is not enough for responsibility; that even if defendants knew that their acts were illegal, they might still be eligible for a verdict of not guilty by reason of insanity if they did not know that their acts were *socially wrong*.[20] No jurisdiction seems to accept the view that a defendant may be found not guilty by reason of insanity simply because he failed to know that his act violated his own personal moral beliefs.[21] In contrast, some jurisdictions that allow the insanity defense when the defendant did not know that the act was wrong do specify that they mean moral wrongness as opposed to legal wrongness and then do not say that they mean social wrongness or social morality. Since they refer to moral wrongness and do not redefine moral wrongness as social wrongness, it seems reasonable to interpret their test as requiring knowledge of plain moral wrongness.[22]

The differences among these interpretations are illuminated by deific-decree delusions. A classic example is *Hadfield*.[23] Hadfield reportedly believed himself to be a modern Christ sent by God to be crucified in order to save the world. In order to get the crucifixion process going, he shot in the king's direction. Hadfield seems to have believed (1) that his act was illegal, (2) that society would judge his act morally wrong if they heard about it, given their current beliefs, yet (3) that his act was not in fact morally wrong. Suppose, then, that Hadfield were a modern-day defendant in a M'Naghten jurisdiction. Given (1)–(3), Hadfield is eligible for the insanity defense under the M'Naghten rule if that rule is interpreted so that it applies to those who did not know that their acts were personally wrong or morally wrong (the third and fourth interpretations), but not if that rule is interpreted so that it applies only to those who did not know that their acts were legally or socially wrong (the first and second interpretations).

In contrast, imagine that, while Hadfield still believed (1) and (3), he was deluded about society's beliefs, so he thought that society shared his beliefs that he was Christ, that his shooting would save the world, and therefore that it was not morally wrong. In this case, Hadfield would have been eligible for the insanity defense under the social, personal, and moral wrongness interpretations but still not under the legal wrongness interpretation of the M'Naghten rule.

When combined with M'Naghten's counterfactual test,[24] delusions like these reveal the need for further distinctions. Contrast two variations on Hadfield's story. In both variations, Hadfield knows that society does not actually share his delusions and does actually judge his act to be morally wrong. These cases vary in Hadfield's beliefs about what society counterfactually would believe if his society did share his beliefs that he is Christ and that his act would save the world. "Hadfield I" believes correctly that, if society shared these beliefs, then it would not judge his act to be morally wrong. So Hadfield I knows that his act is contrary to society's actual moral beliefs not because they subscribe to different moral principles but because they

apply the same moral principles to what they take to be a very different factual situation. Hadfield I, then, knows that his act is socially wrong on the basis of a factual disagreement rather than on the basis of a disagreement about moral principles.

In contrast, "Hadfield II" believes correctly that his society would still judge his act to be morally wrong even if they came to share his beliefs that he is Christ and that his act would save the world (perhaps because they hold absolute principles against ever endangering innocent lives, as Hadfield did). Hadfield II, then, knows that his act is socially wrong not so much on the basis of a factual disagreement as on the basis of a disagreement about moral principles.

Given this distinction, the M'Naghten rule could be interpreted as excusing those who do not know that their acts are socially wrong on the basis of a factual disagreement or as excusing those who do not know that their acts are socially wrong on the basis of a disagreement about moral principles. Hadfield II would be excused under both interpretations, but Hadfield I would be excused under only the former and not the latter interpretation. Applying these interpretations to cases would require determinations not only of what the defendant believes about actual society but also of what the defendant believes about what society would conclude in counterfactual situations.

A parallel distinction can be drawn for the moral wrongness interpretation. Let's assume that Hadfield's act would not have been morally wrong if it really would save the world, as he believed. In contrast, Unruh's killing would still have been morally wrong, even if all of Unruh's beliefs were correct about his neighbor's derogatory comments. Let's further assume (going beyond the reports) that Unruh believed that his act was not morally wrong because he believed that it is morally permissible to kill people for making derogatory remarks. If so, Unruh lacked a kind of moral knowledge that Hadfield did not lack. Hadfield might have held defensible moral principles (such as that saving the world is an adequate justification for shooting in the king's direction) even though he misapplied those principles because he wrongly believed that he was Christ and that his plan would help to save the world. In contrast, Unruh (under our supposition) did not know either the situation (because he was not in fact being insulted) or the appropriate moral principles (because derogatory comments do not justify killing). In other words, Unruh (under our supposition) lacked moral knowledge of both the facts and the relevant moral principles, whereas Hadfield lacked only knowledge of the facts and not of the relevant moral principles. If the M'Naghten rule were interpreted as excusing only those who are ignorant of the relevant moral principles, then only Unruh, not Hadfield, would be excused on this basis. But if the M'Naghten rule were interpreted as excusing those who are ignorant of the facts and therefore of the morally correct course of action, then both Unruh and Hadfield would be excused.

Courts do not draw all of these fine distinctions when they apply the M'Naghten rule, but they do express various attitudes toward cases like these. Some jurisdictions hold that deific delusions do not excuse and therefore that defendants who claim that they were following God's commands should still be found guilty. This reaction might be motivated either by skepticism that the defendant really did

believe in a divine command or by fears of practical problems if all such defendants were found not guilty.[25]

Other jurisdictions, however, reach the opposite conclusion: that such deific delusions do remove responsibility when defendants thought that following God's commands was the morally right thing to do.[26] Some of these jurisdictions excuse only those defendants who believed that society would not have morally condemned their acts if society had shared their factual beliefs, as with Hadfield I.[27] On such tests, responsibility depends on knowledge of what can be called counterfactual social wrongness. Other jurisdictions excuse defendants if their acts would not have been morally wrong if their delusional beliefs had been true—a kind of counterfactual moral wrongness test.[28]

Whichever kind of knowledge is required for responsibility, the M'Naghten rule requires that lack of knowledge to be related to "a defect of reason, from disease of the mind." Bare lack of knowledge does not remove responsibility, especially if the agent should know what he is doing. Someone who does not know that there is poison in a drink that he serves still might be responsible if he was negligent—that is, if he should have known or should have checked to see whether there was poison in the drink. In contrast, it seems inappropriate to say that the agent ought to have known better when his lack of knowledge was due to "a defect of reason, from disease of the mind."

It is still not clear what counts either as a disease of the mind or as a defect of reason. One plausible interpretation is that a lack of knowledge results from a defect of reason when it does not respond to reasons to the contrary. This absence of reasons-responsiveness constitutes (or is evidence of) an inability to know what one is doing or that it is wrong.[29] M'Naghten's belief that the prime minister was out to get him was probably resistant to reason in this way. No matter how much one argued with him, he would still believe this falsehood. People who suffer from paranoia are often like that. They are not able to correct their beliefs in light of evidence or reason. That might explain why they are said to labor under "a defect of reason."

Irresistible Impulse and Loss of Control

The exclusive focus on knowledge in the M'Naghten rule motivated revisions in many jurisdictions. The original formulation covers only mental diseases that affect cognition, but it fails to cover other mental diseases, including volitional or emotional diseases that disrupt or undermine people's abilities to choose and act in certain ways.[30] For example, Lorena Bobbitt was found not guilty by reason of (temporary) insanity after she cut off her husband's penis with a kitchen knife while he slept on June 23, 1993. "The defense argued—and the jury, after slightly more than six hours of deliberation, apparently agreed—that Mrs. Bobbitt, flooded with nightmarish images of her husband's abuse and suffering from a variety of mental illnesses, snapped psychologically after her husband raped her, and yielded to an

'irresistible impulse' to strike back."[31] Mrs. Bobbitt knew what she was doing and presumably knew that it was wrong, since her husband was no threat to her while he was sleeping, and her act would not, and was not intended, to make her safe. So she would seem to have no excuse under the purely cognitive M'Naghten rule. Yet she still might seem not guilty by reason of insanity for attacking her husband.

Consider also kleptomaniacs who steal items even though they know that they could easily afford the items, there is a significant danger that they will be caught, and the punishment that they are likely to receive if caught is many times worse than any pleasure or profit that they might gain from stealing.[32] Why, then, do they continue to steal even if it is not in their self-interest? They claim that they just can't help themselves. If so, their mental diseases are much like being addicted to a drug, except that they usually did nothing to create their mental diseases.

This kind of disease can be seen as a kind of "defect of reason," but here the defect is in what is called practical reason rather than theoretical reason. The judges might have intended their reference to a defect of reason in the M'Naghten rule to cover this kind of practical or volitional defect.[33] Later generations, however, saw reason as more narrowly cognitive. Given this restrictive interpretation and our advancing knowledge of mental diseases that are not purely cognitive, many experts thought that some special clause needed to be added to the M'Naghten rule in order for it to excuse these volitional impairments, and many states expanded the M'Naghten rule by incorporating some reference to volition.

The most common way in which states expanded the M'Naghten rule was by adding an "irresistible impulse" test. According to one formulation of the irresistible impulse test, a person is not criminally responsible if he is

> impelled to do the act by an irresistible impulse, which means before it will justify a verdict of acquittal that his reasoning powers were so dethroned by his diseased mental condition as to deprive him of the will power to resist the insane impulse to perpetrate the deed, though knowing it to be wrong.[34]

This test seems to view mental disease as a kind of force that overcomes the agent's ability to resist because either the force of the impulse is too strong or the agent's willpower is too weak(ened) or both.[35]

It is not clear what it means for an impulse to be irresistible. One possibility is that an impulse is irresistible when the particular agent is unable to resist it in the actual circumstances of the offense. Another possibility, which counts fewer impulses as irresistible, is that the particular agent must be unable to resist it in any circumstances. And there might be an objective standard built into the defense as well: an impulse does not count as irresistible unless it is too strong to be resisted by any normal or reasonable person.

The notion of an impulse is also problematic because it suggests suddenness. In contrast, many volitional diseases do not operate quickly. Just as people can be addicted to cigarettes for a long period of time without feeling any sudden impulses to smoke immediately (especially if they smoke before any impulses arise), so some volitional diseases display themselves in long-term patterns of behavior rather than

in quick impulses. While a kleptomaniac might not feel any sudden rush to steal, she still might be unable to stop herself from stealing for any extended period of time.

One reason why she might be unable to stop herself is that the tension mounts until it is too great to resist. A second reason for her inability to stop herself might be that the tension does not increase but her willpower weakens, so her ability to resist the continuous tension diminishes. On a recent view, willpower is like a muscle that can get tired when it is used.[36] Yet a third possible reason why she might be unable to avoid misconduct is that the pressure is unrelenting for long periods and fighting it requires more attention and hope than she can keep up for long enough. After all, I can raise ten pounds easily, but I cannot keep it raised for an hour. The weight does not get heavier, and I might not get so tired that I literally cannot hold it up any more, but I always eventually let it down because my attention lapses or I lose hope and become resigned to the inevitable. Likewise, some forms of mental illness might create persistent urges that can be resisted for a while but not forever. Such mental diseases might make people unable to avoid certain acts without either weakening the will or causing any irresistible urge.

These kinds of cases are excused from criminal responsibility under some "loss of control" tests. One leading case, *Parsons v. State*, held that a person who knows what he is doing and that it is wrong

> may nevertheless not be legally responsible, if the following two conditions concur: (1) if, by reason of the duress or such mental disease, he had so far lost the power to choose between the right and wrong, and to avoid doing the act in question, as that his free agency was at the time destroyed; (2) and if, at the same time, the alleged crime was so connected with such mental disease, in the relation of cause and effect, as to have been the product of it solely.[37]

This rule (the "Parsons rule") is broader than an irresistible impulse rule, which applies only when there is sudden impulse or weakened will power, because the Parsons rule does not mention either impulse or resistance, so it covers any loss of "the power to choose" regardless of what causes that loss.

The Parsons rule also requires causation. This requirement was probably implicit when the M'Naghten rule referred to "a defect of reason, *from* disease of the mind,"[38] but the Parsons rule makes this requirement explicit in its clause (2). The point is to ensure that the mental disease is causally relevant to the crime. For example, even if someone suffers from kleptomania at the time of his crime, this mental disease hardly helps to remove his legal responsibility if his crime was rape. In order to remove responsibility, the mental disease must not merely be present at the time of the crime but must also actually cause the crime in an appropriate way.

The term "solely" at the end of the Parsons rule might seem too strong because acts almost never have one lone cause, especially if causes are understood as necessary ("but for") conditions. More likely, the Parsons rule uses the term "solely" to refer not to necessary conditions but to the "triggering" or "proximate" cause of the alleged crime. On this view, the term "solely" in the Parsons rule really means *primarily*. It is meant to exclude certain kinds of causal overdetermination: namely,

situations in which the defendant's mental disease is not the primary cause of, or motivation behind, her crime. Suppose, for example, that a patient with kleptomania murders her psychiatrist in order to avoid having to pay her psychiatry bills. To be sure, she would not have committed the murder but for her mental illness; for, without the illness, she never would have met her psychiatrist in the first place. But she is still responsible for this murder because her mental defect, kleptomania, is not the *primary* cause of her action. Instead, her act is primarily caused by a factor that is *not* either identical with or the product of her kleptomania: greed.

The Product Test

Although loss-of-control tests are broader than cognitive tests, they still do not cover all kinds of mental illness. For example, some mental illnesses affect emotion more than they affect will or cognition. One example might be depression, including the extreme postpartum depression that is reported to lead some mothers to kill their children shortly after birth.

Moreover, many psychiatrists object to the artificiality of dividing our mental worlds into cognition, volition, and emotion. According to these psychiatrists, these categories are vague and intertwined. For these reasons, they suggest that an insanity defense should refer to mental disease generally, without any commitment to subdivisions.

Some legal authorities who accepted this view adopted the "Durham rule," which is also known as the Product Test:

> [A]n accused is not criminally responsible if his unlawful conduct was the product of mental disease or mental defect.[39]

This simple test requires only two elements: (1) mental disease or defect, and (2) a causal connection between the mental disease or defect and the act.

The main problem with the Durham rule stems from its very point. While it is simpler than the M'Naghten and Parsons rules, and while it also gives psychiatrists the freedom to testify as they want and juries the freedom to consider what they want, the Durham rule fails to give psychiatrists or juries much, if any, guidance on how to testify or decide cases.

When tough decisions have had to be made, courts have felt forced to come up with their own ways of answering questions about the Durham rule. The first question is whether a mental disease causes an act when it is only a necessary and not a sufficient condition of the act. It might seem that a mental disease would not excuse in such cases because to say that the mental disease is not sufficient for the act is just to say that it did not necessitate or compel the act, in which case the act could have been avoided. Still, courts tended to follow general theories of legal causation, according to which the main decision on the causal prong of the Durham test was

whether or not the mental disease was a necessary condition of the act. If "the accused would not have committed the act he did commit if he had not been diseased as he was,"[40] then the act was seen as a product of the disease.

The next question is what counts as a mental disease. While it might seem tempting just to say that "mental disease" in the Durham rule has the same meaning as the medical term "psychosis," this interpretation gives too much legal power to psychiatrists, since psychiatrists decide which conditions to label "psychosis." Critics accordingly charged the Durham rule with imposing "trial by label."[41] Moreover, when psychiatrists classify mental illnesses, they typically have in mind the treatments that they administer. The classifications that work for treatment might not work for criminal justice or other legal purposes.[42] Partly for this reason, a leading decision in the Durham tradition defined "mental disease" in legal terms:

> [A] mental disease or defect includes any abnormal condition of the mind which substantially affects mental or emotional processes and substantially impairs behavior controls.[43]

The goal of this definition was to provide the guidance to witnesses and juries that the original Durham rule by itself had failed to provide.

There are three things to notice about this definition of mental disease. First, unlike the M'Naghten rule (as it is commonly interpreted), this definition does not give priority to cognition but rather refers to "emotional" and broadly "mental" processes. Second, the conjunction "and" implies that a condition does not count as a mental disease unless behavioral controls are impaired. Third, the term "substantially" has been added, presumably because mental diseases are rarely, if ever, completely overwhelming.

This recognition of partial mental illness is crucial. A kleptomaniac might be able to keep himself from stealing when he knows that a policeman is standing at his elbow but still not able to stop himself when he knows only that a policeman is likely to be watching from the other side of the room or that nobody is watching. Or he might be able to go for a week without stealing but not for a whole month. It is not clear whether a kleptomaniac who steals once a week even though he is compelled to steal only at least once a month is morally or legally responsible. Such incomplete incapacities do not guarantee that the person will steal under all circumstances, but they still affect the person's behavior in many circumstances of everyday life, where cues and temptations abound. Whether or not an individual defendant retains enough capacity for responsibility is, in effect, left up to the jury to determine, possibly in light of community standards.[44]

THE MODEL PENAL CODE

Many of these modifications were brought together by the American Law Institute in its popular Model Penal Code (MPC) formulation of the insanity defense, which said:

> A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality (wrongfulness) of his conduct or to conform his conduct to the requirements of the law.[45]

The American Legal Institute included the parenthetical "wrongfulness" after "criminality" because it wanted to leave states free to choose either term. The MPC also added a restriction:

> The terms "mental disease or defect" do not include an abnormality manifested only by repeated criminal or otherwise anti-social conduct.[46]

This restriction was intended to keep contract killers, serial killers, sociopaths, and psychopaths from being found not guilty by reason of insanity just because they committed so many crimes. It is not clear, however, that it still applies to psychopaths today, since contemporary science has found many manifestations of psychopathy beyond criminal or otherwise antisocial conduct.[47] In any case, despite this clarification, the MPC never provides any positive definition of mental disease.

The MPC rule is still clearly broad. It refers to appreciation rather than knowledge, does not require complete loss of capacity, and covers volitional as well as cognitive defects. In this way, the MPC rule seems to excuse acts resulting from all of the kinds of mental disease discussed above, with the possible exception of purely emotional impairments.

This MPC rule was a great hit. By 1980, the MPC rule or some close relative was adopted in twenty-eight states as well as by every federal court of appeals that had addressed the issue. In the majority of states, the rule was adopted legislatively.[48] (The M'Naghten rule was used in every other state except New Hampshire and the District of Columbia, which stuck with their versions of the Durham rule.) The popularity of the MPC rule at that time might have been due in part to its flexibility. States could interpret this rule in various ways to fit their preferences.

To understand the MPC rule, we need to ask, first, what the phrase "he lacks substantial capacity" means. The need to refer to *capacities* should be clear. Although the M'Naghten rule hinged on actual knowledge rather than ability to know, the criminal law clearly should not excuse agents who do not happen to know that their acts are wrong if those agents could be reasonably expected to know that their acts are wrong. Such agents can be held responsible for their failure to exercise their ability to know the wrongfulness of their conduct. In contrast, agents who lack the capacity to know that their acts are wrong obviously cannot be held responsible for failing to exercise this capacity, a capacity that they don't have. This point helps to explain why the MPC is formulated in terms of a capacity to appreciate rather than in terms of actual appreciation.

Capacities clearly come in degrees. The MPC rule highlights this point by referring to *substantial* capacity. Still, this MPC clause is ambiguous between a relative measure and an absolute measure. It might seem to say either that agents are not responsible if their capacity falls substantially below what is normal or that agents

should be excused only when the capacity that they retain, if any, is not substantial. To see the difference, imagine that degrees of capacity can be located on a scale from one to ten, any decrease of three or more points counts as substantial, and most people are at capacity level nine. Then someone with capacity level five lacks substantial capacity in the former, relative sense (he is more than three below the normal nine) but retains a substantial capacity in the latter, absolute sense (he is more than three above the bottom). What is important here is that even if an agent retains some capacity to appreciate wrongfulness, she still might be eligible for the insanity defense under the MPC clause if her retained capacity is either substantially below what is normal or too minimal to count as substantial.[49]

Next we need to ask which capacity matters. The MPC rule refers to two specific capacities: the capacity to conform one's conduct to the law and the capacity to appreciate the wrongfulness of one's act. The conformity clause, also known as the *volitional prong*, seems to be intended to capture the point of irresistible impulse and loss of control tests like those discussed above. If agents lack substantial capacity to control their conduct, and if they (like most of us) run into situations where they need to exercise control in order to conform to the law, then they lack capacity to conform to the law. Certain compulsions, possibly including kleptomania, are believed to fall under this clause.[50]

The other clause refers instead to appreciation of wrongfulness. This part of the MPC rule, often called the *cognitive prong* , is supposed to correspond to both parts of the M'Naghten rule, since an agent cannot appreciate the wrongfulness or criminality of an act if she cannot appreciate the nature and quality of the act, especially if the relevant parts of the nature and quality of an act are those that matter to criminality and wrongfulness. In the examples discussed above, if a husband cannot know that he is squeezing a lemon instead of his wife's neck, or if Joy Baker cannot know that she is shooting her aunt rather than the devil, then they both lack substantial capacity to appreciate the wrongfulness of the act.

In contrast, when an agent does know the relevant nature and quality of the act, then it is less clear what it means to say that she lacks substantial capacity to appreciate its wrongfulness. The term "wrongfulness" in the MPC rule is subject to the same interpretations as the term "wrong" in the M'Naghten rule discussed above. Four kinds of wrongfulness were distinguished: legal, social, personal, and moral. Many jurisdictions do not clearly distinguish these interpretations, but courts could interpret the MPC rule so that it requires substantial appreciation of any of these kinds of wrongfulness (or any combination of them).

One alternative for jurisdictions that use the MPC rule is to hold that it is enough for criminal responsibility that an agent appreciates that her act is illegal or criminal—that is, *legally* wrong. The MPC explicitly allowed this option by including the word "criminality" in addition to "wrongfulness" in its proposed insanity defense. Some jurisdictions adopt this interpretation explicitly.[51] The rationale behind this view is presumably either retributive (because agents who can know the law are rational enough to be morally responsible for not obeying it) or deterrent (because threats of criminal punishment can deter criminal behavior if the agent

can know which acts risk punishment). On this view, the defendant need not also understand the moral basis for the law.

In contrast, the MPC rule also allows states to refer to wrongfulness instead of criminality, and some states explicitly adopt this option.[52] Within this group, several jurisdictions specify that by "wrongfulness" they mean *social* wrongfulness, so that defendants are eligible for the insanity defense if and only if they lack substantial capacity to appreciate that their acts violate the moral code in the relevant society—that is, that people in that society do or would see the act as immoral.[53] The rationale is presumably that it would not be fair to hold people morally responsible for violating the prevailing social morality if they cannot appreciate what the prevailing social morality forbids. A defendant might, for example, personally believe that wife-beating is not immoral. Nonetheless, on this approach, it would still be legitimate to punish him for wife-beating as long as he can appreciate the fact that most of his society believes that wife-beating is immoral.

The third possible interpretation of "wrongfulness" in the MPC is that an agent is not criminally responsible for an act unless she has substantial capacity to appreciate that the act is *personally* wrong in the sense that it violates her personal moral beliefs. As with the corresponding interpretation of the M'Naghten rule, no jurisdiction adopts this interpretation of the MPC rule.[54]

Finally, the MPC could be interpreted as suggesting that an agent is not criminally responsible for an act unless she has substantial capacity to appreciate that the act is just plain *morally* wrong. On this interpretation, it is not enough that the agent can appreciate that the act is contrary to law or contrary to social or personal moral beliefs. To be criminally responsible, the agent must be able to appreciate the facts about the act that make it morally wrong and also that these facts do make it morally wrong. Some jurisdictions seem to adopt or allow this moral interpretation at least implicitly insofar as they specify that the relevant kind of wrongfulness is moral rather than legal wrongfulness or criminality and do not add that the relevant kind of wrongfulness is only social wrongfulness.[55]

A rationale for this moral wrongfulness interpretation might come from an expressive theory of punishment, according to which a major, if not *the* major, purpose of punishment is to express moral condemnation of criminal acts.[56] A penalty such as a parking fine merely inflicts a loss as a way of controlling behavior. We usually do not morally condemn people who get parking tickets, as long as they pay their fines. Criminal punishments also inflict losses, but that is not all they do. They also typically express moral condemnation of acts and agents. We condemn rapists as bad people even if they suffer losses for what they do. Fines would not be enough for rape even if they worked to prevent it. We also wish to proclaim to society that rape is morally wrong and, ideally, to make rapists understand its wrongness and feel remorse for their crimes.

Given such an expressivist purpose for punishment, it seems inappropriate to punish agents who lack the capacity to understand why their acts are immoral because they cannot understand what is being expressed. We can fine people for parking illegally even if they cannot understand the reasons why this parking space

is reserved, but moral condemnation seems inappropriate if a person lacks the capacity to understand any moral reason why the violated law is a law in the first place. Insofar as criminal punishment goes beyond behavior control by penalties and is supposed to express moral condemnation, it seems appropriate to excuse those who cannot understand any moral reasons behind the criminal laws that they break because they cannot understand what would be expressed by the punishment.

After specifying the kind of wrongfulness that needs to be appreciated, a separate question of interpretation still needs to be answered. Jurisdictions need to specify what it means to *appreciate* that kind of wrongfulness. The term "appreciate" is the MPC's replacement for the term "know" in the M'Naghten rule, so it helps to begin by asking what counts as knowing wrongfulness.

One possibility is that the relevant kind of knowledge or appreciation is no more than the ability to answer questions correctly. We do, after all, say that someone knew the answer when he got it right on a game show or a standardized test. This abstract and isolated kind of knowledge, however, hardly seems sufficient for moral or criminal responsibility. Giving the correct answer is compatible with failing to *understand* the answer or its meaning. Such a thin mental state might be enough for awarding a prize on a quiz show, but it seems inadequate to warrant criminal punishment.

A few courts have explicitly adopted a similar view, usually in dicta, that even if a person possesses a bare "theoretical," abstract, or cognitive understanding of wrongfulness, she may still be found legally insane if she lacks a deeper emotional or affective understanding of wrongfulness.[37] The change from "know" in the M'Naghten rule to "appreciate" in the MPC rule is arguably an attempt to move beyond a purely abstract account of knowledge. Appreciation requires the person not only to know the right answers to questions but also to understand those answers. What, then, is this understanding?

Understanding might be shown, among other ways, by drawing appropriate inferences from answers. Thus, if a person says not only that his act is wrong but also that its wrongfulness implies that he ought to feel bad about it, that he deserves punishment for it, and so on, then he knows and appreciates its wrongfulness at a deeper level than someone who calls it wrong but cannot say what this answer implies. This deeper level of knowledge or appreciation still remains purely cognitive insofar as drawing out implications is a cognitive skill.

One problem with this "inferential" view of understanding is that a person might be able to draw out implications without understanding what the implications mean any more than she understood what implied them. Such abstract inferences without understanding are exemplified by psychopaths, who often make normal moral judgments and draw normal inferences from them but still do not emotionally appreciate what they say. For example, when asked if he experienced remorse for a murder that he had committed, one psychopath said, "Yeah, sure, I feel remorse." Pressed further, he said that he didn't "feel bad inside about it." Similarly, when asked whether he had ever committed a violent offense, a man

serving time for theft answered, "No, but I once had to kill someone."[58] Hence, real appreciation requires more than merely drawing out abstract inferences.

Another interpretation of understanding solves this problem by requiring not only cognition and inference but also "affective knowledge" or "emotional appreciation."[59] On this view, a defendant does not appreciate the wrongfulness of an act if, for example, she does not "internalize the enormity" of the criminal act.[60] Someone who announces that an act is wrong and mimics common inferences about the need for remorse and punishment still might not have appropriate feelings or motivations.

Consider, by analogy, a person on a roof who says that it would be stupid to jump off because it would cause him great pain and disability. But then he jumps off anyway, and he gives no reason other than that he thought it would be fun. This act seems to be strong evidence that the jumper did not really know or appreciate just how stupid jumping was, even if he called it "stupid" and drew appropriate inferences. What he lacked was appropriate emotions or feelings, especially fear, and motivations, such as desire to avoid to pain and disability. His cognitive abilities did not connect to his decision-making, and this disconnection is grounds for saying that he did not fully know or appreciate how stupid his act was.

This point still holds if the jumper does not regret his jump after it breaks his legs. That would be even more evidence that he does not fully appreciate what is wrong with his act, since what is wrong is that he lost much more than he gained by jumping.

Now extend these points to appreciation of moral wrongfulness as opposed to imprudence. Imagine that instead of jumping, someone pushes a victim off the roof. Just as the jumper could not be thought to fully appreciate the stupidity of jumping off the roof when he lacked the appropriate emotions and motivations (fear and self-protection), the pusher does not seem to fully appreciate the immorality of pushing another person off the roof if he equally lacks the appropriate emotional concern for this other person's well-being. This concern seems necessary to fully appreciate why it is morally wrong to push the victim off the roof.

Law has often underestimated the importance of emotion, but contemporary psychology calls for correcting this oversight.[61] There are several reasons to think that emotion is necessary for the kind of capacity to appreciate that is required for moral responsibility and therefore should be required for criminal responsibility. First, emotions and cognition are hard to separate because emotions are necessary for normal cognition even apart from morality. Antonio Damasio, among others, has argued that people whose emotions are reduced or destroyed by frontal lobe damage, for example, cannot properly assess risks and benefits.[62] Second, emotions are necessary for triggering inhibitions and therefore for consistent law-abiding action and for deterrability in the long run.[63] Third, emotions help us pick out what is salient, and a sense of salience is necessary for consistent decision-making in the face of overwhelming amounts of information.

More generally, the same basic argument for excusing insane defendants applies as well to defendants whose mental illnesses affect emotions: to blame people for

acts that result from mental illness, even when that mental illness manifests itself only or primarily in an emotional deficit, is like blaming a sick person for blowing his nose in a library. Of course, not all emotional deficits excuse all crimes, but some emotional deficits still might excuse some crimes at least partially. That is why most jurisdictions provide for the defense of extreme emotional disturbance, a condition that reduces a higher degree crime to a lesser degree crime of the same kind (e.g., first-degree homicide to second-degree homicide).

The point can also be brought out by analogy to children and animals. Imagine that your five-year-old son yells, "I hate you!" He is angry but not so angry as to be out of control. He knows that these words will hurt you deeply, and that is exactly why he utters them. Still, we do not hold him fully morally responsible. The reason for excusing him seems to be that he lacks certain relevant emotions. He does not fully appreciate what he is doing or why it is wrong. If that is why we (partially) excuse this child, then similar lacks in adults that result from mental disease or defect should excuse them as well. If lack of substantial capacity to emotionally appreciate wrongfulness explains why children are not (fully) responsible for their actions, then it should also remove or at least reduce responsibility in adults.

A Federal Rule

Despite its flexibility and popularity, the MPC rule ran into problems. First, it turned out to be hard to apply to actual cases. Many commentators blamed the MPC rule for the *Hinckley* decision, which they saw as misguided. As a result, in reaction to this decision, several states abandoned or weakened the MPC rule by avoiding or dropping the volitional prong and returning to a formulation closer to the old M'Naghten rule.[64] (See below on these reforms.) Second, other critics complained that the notion of "substantial" capacity was too vague and favored, in the interest of clarity and easy application, requiring total incapacity for an insanity defense.

For these reasons, the U.S. Congress in 1984 adopted the following statutory definition of insanity:

> It is an affirmative defense to a prosecution under any Federal statute that, at the time of the commission of the acts constituting the offense, the defendant as a result of a severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts. Mental disease or defect does not otherwise constitute a defense.[65]

The restriction to "severe" mental illness, the deletion of the volitional prong, and the last sentence were all intended to reduce the number of defendants who would be eligible for the insanity defense and thereby to avoid future decisions like that in *Hinckley*.

Insanity as Irrationality

In addition to these developments in law, some legal theorists have proposed a novel way to understand insanity. They emphasize that mentally ill people typically act irrationally, and they claim that this irrationality is the basis on which the insane lack responsibility and should be excused.[66] Different proponents of this general approach disagree about the nature of irrationality, but they agree that insanity should be understood in terms of irrationality of some kind.

Still, we need to look carefully at the kind of irrationality that is supposed to remove responsibility. Consider someone who kicks a computer out of anger. This act of kicking is irrational insofar as it hurts his foot and the computer for no good reason. Nonetheless, the kicker seems responsible for any damage to the computer. Indeed, he also seems just as responsible if what he kicks is not his computer but rather a child. As long as he has (1) the ability to think things through and realize that he has no reason to kick the child and many reasons not to, and (2) the ability to act on his reasons, we can reasonably expect him to exercise these abilities and hold him fully responsible if he fails to exercise them, regardless of the irrationality of his act. Since irrationality does not remove responsibility in cases like this, irrationality by itself cannot be what removes responsibility in cases of insanity.[67]

What removes responsibility, if anything, is a lack of the *capacity* to be rational. People lack this capacity if they cannot form rational beliefs or rationally consider the criminality or wrongfulness of their acts (a defect in theoretical rationality) or if they cannot act according to the reasons that they have (a defect in practical rationality). These are exactly the lacks that remove responsibility according to the MPC rule. So it is not clear that these theoretical revisionists are so far from the MPC rule after all.

The notion of irrationality might give us additional insight into the MPC approach by revealing an underlying unity among defenses involving mental incapacities such as insanity, mental retardation, intoxication, somnambulism, and hypnosis.[68] Still, no jurisdiction has officially adopted this suggestion or defined insanity directly in terms of irrationality. This omission might be because the term "rational" is vague and controversial. Still, it might be possible for the rationality approach to be developed in fruitful ways.

Arguments for an Insanity Defense

Partly because of the difficulties in defining "insanity," there are serious doubts about whether or not our laws should recognize insanity as a separate defense at all. People whose criminal acts are due to mental disease can sometimes be found not guilty on the basis of other, less controversial conditions of responsibility, such

as that the agent lacked the specific mens rea for the crime. That finding makes a big difference with regard to how the legal system treats defendants. If they are found to lack mens rea, then they are acquitted and released. In contrast, if they are found not guilty by reason of insanity, then they are committed to a secure mental facility, possibly for longer than they would have spent in prison.[69] This difference makes it important to ask whether, and why, we should have a separate defense for insane people who would not be excused on any other ground, including lack of mens rea.

The main argument for an insanity defense is simple:

> If the insanity defense were abolished, the law would not take adequate account of the incapacitating effects of severe mental illness. Some mentally ill defendants who were psychotic and grossly out of touch with reality may be said to have "intended" to do what they did but nonetheless may have been so severely disturbed that they were unable to understand or appreciate the significance of their actions. These cases do not arise frequently, but when they do a criminal conviction, which signifies the societal judgment that the defendant deserves to be punished, would offend the basic moral intuitions of the community. Judges and juries would be forced either to return a verdict of conviction, which they would regard as morally obtuse, or to acquit the defendant in defiance of the law. They should be spared that moral embarrassment.[70]

Insanity removes moral responsibility, and criminal punishment without moral responsibility is both morally and practically problematic. If the (main) purpose of punishment is either to repay moral wrongs or to express moral condemnation, then punishment seems inappropriate when the punished person was not morally responsible. Still, this conclusion is controversial because it is not clear that the (main) purpose of punishment is retribution or expression of moral condemnation rather than more consequentialist goals like rehabilitation, incapacitation, specific deterrence, general deterrence, victim healing, or protection of society.

But why does insanity remove moral responsibility? One answer is that moral responsibility requires the ability to do otherwise (or control) and insanity negates the ability to do otherwise.[71]

This requirement of the power to do otherwise for moral responsibility seems to stand behind the voluntary act requirement. For example, an epileptic is not responsible for hitting someone during an unforeseeable seizure. There is no voluntary act in this case, but why not? Our criminal law does not count bodily movements as voluntary acts in certain cases, such as seizures, reflexes, hypnosis, and sleepwalking, because they do not result from conscious will. But why should those conditions remove responsibility? The reason seems to be that we cannot avoid doing what does not result from conscious will, which is what enables us to control what we do. This rationale for the voluntary act requirement applies also to acts that result from mental illnesses if their agents could not avoid having the illness and then acting as the illness dictates.

Similarly, consider cognitive excuses. Reasonable mistakes excuse in a way that unreasonable mistakes do not. Why? A natural answer is that we can avoid

unreasonable mistakes by being careful, but we cannot avoid making mistakes that are reasonable unless we exercise more caution than it is reasonable to expect of us. Likewise, if mental diseases make people unable to avoid breaking the law, we cannot reasonably expect them to have done otherwise and therefore should not hold such people responsible or punish them for their illegal actions.

This point seems to motivate formulations of the insanity defense that hinge on capacity, control, or what is irresistible, but it also extends to cognitive insanity defenses or cognitive prongs of insanity defenses. If a person is so deluded that she cannot know whether what she is cutting is a steak or a person, then she cannot avoid cutting people who sit next to her while she eats steak. And if she cannot know that it is morally wrong to cut people, then we cannot reasonably expect her to choose her actions according to appropriate moral reasons.

Moreover, even if she knows both that she is cutting a person and that this is wrong, she still might lack the capacity to refrain from cutting others (perhaps because of alien-hand syndrome or some other psychological compulsion). The clauses in insanity defenses that refer to capacity, control, or resistibility seem tailored to capture such conditions, conditions under which a person cannot avoid performing illegal acts. So if both (1) the ability to do otherwise is necessary for moral responsibility and (2) moral responsibility is necessary for justified criminal punishment, then we should not criminally punish people who lack the requisite abilities because of insanity.

Of course, abilities come in degrees. It is harder for some people than others to obey the law. Circumstances also matter. Some insane people can avoid breaking the law by seeking treatment during their lucid periods or by avoiding situations in which they will end up doing harm. Still, if insane people lack the degree of ability or control that is needed in order for them to be morally responsible on some occasions, then they should be excused from criminal punishment for acts that they perform on these occasions (assuming that these conditions can be proved by the applicable standards of evidence).

It might be argued that even insane people who *can* conform their conduct to the law still might be excused for some of their bad acts. Sometimes a person is capable of refraining from acting, but it is still unreasonable to expect that person to refrain from acting. Consider coercion. If a robber says, "Your money or your life," then you are still capable of refusing to hand over the money. You might even have a chance of escaping with the money. Still, you cannot reasonably be expected to take such a chance, which is why we do not blame you for handing over the money. Likewise, this reasonable-expectations excuse constitutes another rationale for an insanity defense. When people are insane in such a way that they are unable to respond to reasons (as discussed above). it seems unreasonable to expect them to act on different reasons and thereby to avoid the acts that their mental illnesses cause them to perform.

A separate set of arguments for the insanity defense comes from consequentialists. If someone really is insane in a way that removes the ability to avoid doing illegal acts, then threats of punishment will not deter this person from breaking the

law. He will continue to commit the illegal acts regardless, in which case punishment would be pointless.[72] Similarly, if the purpose of punishment is moral education of criminals or rehabilitation more generally, then these purposes cannot be served by punishing insane people who simply cannot learn or come to know right from wrong.[73] Indeed, punishment and confinement might even exacerbate some mental illnesses, perhaps by removing support networks.

Arguments against an Insanity Defense

Other consequentialists argue on the opposite side of the issue that even if punishment cannot specifically deter, morally educate, or rehabilitate certain insane people, punishing the insane can still promote other benefits. These other benefits are sometimes claimed to justify abolishing or severely restricting the insanity defense.

One benefit is supposed to be general deterrence—that is, deterring people in society other than the specific individual who is punished.[74] Punishing the insane can promote general deterrence in at least two ways. First, threats of punishment can deter *some* insane criminals. Even if fully insane people cannot be deterred or taught moral lessons, those who are only partially incapacitated can be affected by threats of punishment. Such threats can provide incentives that are strong enough to counteract their insane inclinations. The above consequentialist arguments *for* the insanity defense arguably seem to forget that people have various degrees of control (as discussed above), and threats of punishment can affect probabilities and rates of misbehavior by those who still have some vestige of control.

The second way in which punishing the insane might promote general deterrence is by preventing sane criminals from hoping (usually mistakenly) that they might be able to avoid punishment by faking insanity. A few criminals have reported that they considered in advance the possibility of faking insanity if they were caught.[75] There is no reliable evidence of how often this happens, but the insanity defense arguably interferes with deterrence to some extent if there are any criminals who commit crimes that they would not otherwise have committed if they had not thought that they could avoid punishment by faking insanity. This is just one instance of the general point that whenever the law allows yet another way to avoid punishment, this extra avenue of impunity increases the perceived chances of escaping punishment and thereby decreases the power of general deterrence.

A second consequentialist argument against the insanity defense is that a verdict of not guilty by reason of insanity arguably creates the impression that the criminal "got away with it." After millions of television viewers watched Hinckley shoot President Reagan in 1981 and then be found not guilty by reason of insanity, many people thought that he got off at least partly because he was rich and had tricky lawyers and psychiatrists on his side. Whether or not this widespread belief

was true, the impression that rich, devious people get away with crimes can decrease people's respect for the law and thereby their motivation to obey the law.

Third, many critics of the insanity defense fear that dangerous people are incapacitated for less time in mental institutions than in jails and, therefore, that the insanity defense might end up putting dangerous people back out on the street. This belief was not true in the nineteenth century, when people committed to mental institutions often stayed there for the rest of their lives. Today, in contrast, new drugs and new requirements for involuntary commitment combine to make release seem easier and quicker, and people who are safe while on their medications can become very dangerous when they go off them.

Fourth, there is the fundamental problem of determining whether or not a particular defendant really is insane. In many cases, psychiatrists testify on both sides, and juries don't know which experts to believe. In addition, psychiatrists themselves often admit that even if they can determine what a person actually believed and intended to do, they cannot determine whether or not a particular defendant had the ability to do what she did not actually do or to believe what she did not actually believe. It is not even clear what kind of evidence could support claims about such abilities and counterfactuals. These evidentiary difficulties and difficult judgment calls make some results seem arbitrary. Indeed, some convicted criminals in prisons have more serious psychiatric problems than others who were found not guilty by reason of insanity.

All of these practical problems are often presented as additional reasons to restrict the insanity defense as well as to change the procedures surrounding it. Can scientific progress, like that in neuroscience for example, help solve these problems? Some hope so, but it is not at all clear yet. While we wait for better diagnoses and treatments, we face a difficult choice between narrowing the insanity defense so as to increase crime prevention and widening the insanity defense so as to ensure that we do not punish anyone who is not really morally responsible.

Proposed Reforms

When such powerful arguments conflict, it is reasonable to try to give each side its due. The main goals are to minimize the risk of punishing people who are not morally responsible and also to maximize the prevention of crime through deterrence, moral education, and the incapacitation of dangerous people. These goals are supposed to be furthered in varying degrees by the various reforms that have been proposed.

The most radical proposal is to abolish the insanity defense entirely. Many prominent scholars have advocated abolition.[76] Even if the insanity defense is abolished, however, defendants can still be found not guilty if they lack a mental condition (such as intent or knowledge) that is required for a crime. Insanity also might

reduce a crime, say, from first-degree or second-degree murder to manslaughter under doctrines of diminished responsibility. Still, the end result of abolition is that people who commit crimes because of insanity will more often be punished.[77]

There are at least three less radical approaches to reforming the insanity defense. First is to reduce its scope by reducing the number of mental conditions that preclude guilt. A proposal of this kind was advocated by the American Bar Association and the American Psychiatric Association.[78] The proposal is to drop the volitional prong of the MPC rule, which (as we have seen) excuses defendants who lack substantial capacity to conform their conduct to the law.[79] Without this clause, the only people who would be found not guilty by reason of insanity would be those who lack substantial capacity to appreciate either what they are doing or that it is wrong, which is closer to the original M'Naghten rule. This proposal is motivated not only by a desire to reduce the number of insanity verdicts but also by supposedly special difficulties in determining whether an individual who *did not* conform to the law still *could have* conformed to the law.

The main problem with this proposal is that even people who know right from wrong do not seem to be morally responsible or, therefore, justly punishable for their actions if they truly could not have avoided performing them. If the reason why cognitive incapacities remove responsibility is that they render one unable to avoid breaking the law, then it is hard to see why volitional or emotional incapacities would not also remove responsibility for the same reason.

A second proposal for reforming the insanity defense is to model it after another feature of the federal rule discussed above. Model Penal Code jurisdictions could drop the word "substantial" so that only a complete lack of capacity to appreciate or conform would remove responsibility. They could also change the word "appreciate" back to "know" in order to ensure that lack of emotional appreciation of the wrongfulness would not remove responsibility. Unfortunately, it is not clear how much difference such wording changes would make to decisions by juries or judges.

A third proposal for reforming the insanity defense is to adopt a new verdict, usually called "guilty but mentally ill." The exact difference between the verdicts of guilty but mentally ill and not guilty by reason of insanity varies, and the former might either replace or be allowed in addition to the latter so that juries could decide between these two verdicts.[80] Separate rules would then have to be formulated for the different verdicts. There might even be separate trials with different procedures for determining guilt and insanity. But however this new verdict is formulated and handled, the point is to allow juries to label defendants as "guilty." The continued use of this key term is supposed to reduce the impression that a criminal "got away with it" and thereby promote respect for the criminal justice system. In addition, people who would be found guilty but mentally ill would sometimes be required to serve out their sentence in prison if they were released from a mental institution before a certain period. This verdict would reduce the chances that dangerous people would return to the streets too soon and would take away much of the incentive to try to fake insanity.

Still, as always, opponents of such proposals argue that if someone really was insane when she committed the crime, then she was not morally responsible. In that case, sending her to prison after she is cured amounts to punishing an innocent person.[81]

Finally, those who fear that the insanity defense lets too many dangerous people back on the streets often propose stricter requirements for release from mental institutions. While the government is now usually required to show that involuntary mental patients are dangerous to themselves or to others, some reformers want people who are found not guilty by reason of insanity to be confined until the patients themselves prove that they are no longer dangerous. Jurisdictions could also require approval by a judge or a panel before release, and some even suggest that the victim or her relatives be reserved a place on the panel. All of these steps would create longer stays for those found not guilty by reason of insanity.

This proposal has its own problems. First, it might lead to inequities insofar as defendants who harmed wealthy victims would probably be confined longer than those who harmed poor victims simply because wealthy victims tend to have more friends who are better able to attend meetings of release panels and to speak at them more articulately and forcefully against release. Second, opponents object, as before, that these people are not guilty in the first place. And if the government cannot show that they are dangerous, as in other cases of involuntary confinement, then it should not be allowed to restrict their freedom. Third, it will be extremely hard for anyone in prison for a crime to prove that she is no longer dangerous at all. What kind of evidence could one cite for such a negative claim?

Restrictions on courtroom evidence have also been proposed. One possibility is to limit defense testimony about the background and emotional history of the defendant. Some critics claim that sad stories of a defendant's childhood can stir up emotions that distort verdicts. If all that is relevant is the defendant's mental condition at the time of the crime, then it might seem irrelevant to talk about his life at other times, much less how his parents treated him when he was young.

Psychiatrists, however, often respond that a person's mental condition cannot be understood properly in isolation from circumstances that gave rise to it. If so, juries need to know the defendant's background in order to reach a justified verdict. Informing the jury of the defendant's past crazy behavior can also help to dispel fears that the defendant might now be faking insanity.

Other restrictions on evidence are also possible. If problems arise when psychiatrists are forced to use legal language, then the law could prohibit psychiatric experts from saying anything directly about the legal issues. Or if insanity trials are ruined by battles between defense and prosecution experts, then the court could assign its own impartial panel of experts. This panel might testify in addition to, or instead of, the experts of both sides.

Such restrictions on evidence, however, fly in the face of a strong tradition of allowing great leeway to defendants in presenting their cases. Without such leeway, defendants will be at the mercy of judges who are not always inclined to select experts sympathetic to defendants and thereby give defendants the fairest possible trials.

Possibly the most popular reform is to shift the burden of proof. Before *Hinckley*, most state and federal courts required the prosecutor to prove the defendant's sanity beyond a reasonable doubt.[82] This burden was often hard to carry because insanity is obscure and experts conflict. It also seems natural to presume that people are sane in the absence of any evidence to the contrary.[83] For these reasons, revisionists proposed either to lighten the burden of proof on the prosecution or to shift the burden to defendants to prove their *in*sanity, as in the original M'Naghten rule.

Since *Hinckley*, most states and the federal system have required the defense to prove that the defendant is insane, usually either by a preponderance of evidence or by clear and convincing evidence, as in the federal rule discussed above. To critics, such shifts seem to conflict with the traditional view that every element necessary for someone to be guilty must be proven by the prosecution beyond a reasonable doubt. This stringent requirement is usually justified by the dangers of convicting innocent people who cannot prove their innocence. Reducing the burden on the prosecution or shifting the burden to the defense might increase the chance of punishing people who are not guilty, if insane people really are not guilty. Of course, defenders of this proposal respond that the insanity defense should be seen as an affirmative defense, comparable to entrapment, and the defense always bears the burden of proof for affirmative defenses.[84]

All of these reforms are contentious, partly because each brings dangers as well as promises of improvement. The proper course, then, seems to depend on how many mistakes of various kinds are likely to be made and also on the relative importance of the different kinds of mistakes. Any balancing act will be difficult and debatable.

NOTES

We happily thank Larry Crocker, John Deigh, and David Dolinko for very helpful comments on previous drafts.

1. *Regina v. M'Naghten*, 10 Cl. & Fin. 200, 9 Eng. Rep. 718 (1843).
2. *U.S. v. Hinckley*, 1:81cr00306 (D.D.C., Aug. 24, 1981).
3. *Jones v. U.S.*, 463 U.S. 354 (1983). Given the risks, it might seem surprising that anyone would plead not guilty by reason of insanity to a minor criminal charge. Experts estimate, however, that as many as 86 percent of insanity pleas are for nonviolent felonies and misdemeanors (National Mental Health Association, *Myths and Realities: A Report of the National Commission on the Insanity Defense* [1963], pp. 20–21). Even for serious crimes, the insanity defense is rarely invoked, though its rate might be underestimated because prosecutors frequently dismiss charges against defendants who plead insanity if they agree to civil confinement in a mental institution. *See* Joshua Dressler, *Understanding Criminal Law, 4th ed.* (LexisNexis, 2006), p. 367.
4. *See U.S. v. Freeman*, 357 F.2d 606, 619–620 (2d Cir. 1966) ("At bottom, the determination whether a man is or is not held responsible for his conduct is not a medical

but a legal, social or moral judgment. Ideally, psychiatrists—much like experts in other fields—should . . . furnish the raw data upon which the legal judgment is based. It is the psychiatrist who informs as to the mental state of the accused—his characteristics, his potentialities, his capabilities. But once this information is disclosed, it is society as a whole, represented by judge or jury, which decides whether a man with the characteristics described should or should not be held accountable for his acts."). For a more compromising view, see *State v. Worlock*, 117 N.J. 596, 606 (1990) ("The insanity defense is a legal standard incorporating moral considerations often established by medical testimony. Notwithstanding advances in psychiatry, mental states remain inscrutable, especially when an abnormal state is in issue. Generally, the determination of a defendant's ability to distinguish between right and wrong depends on psychiatric testimony. Such testimony provides insight into a defendant's mental condition and enables the fact-finder to differentiate defendants who can choose between right and wrong from those who cannot. Ultimately, the insanity defense requires satisfaction of a legal, not a medical, standard. As applied, the 'right and wrong' test blends legal, medical, and ethical concepts. Perhaps for this reason, courts have generally admitted any credible medical testimony on the insanity defense."). Some courts do refer to "medical insanity," but then they distinguish it from legal insanity (though this is misleading for reasons given in the text). *See, e.g., Epps v. State*, 984 So. 2d 1042, 1048 (Miss. Ct. App. 2008) ("It is well established in Mississippi law that just because a criminal defendant is considered medically insane does not mean that he is M'Naghten insane."); *People v. Irwin*, 4 N.Y.S.2d 548, 550 (N.Y. 1938) ("In New York State insanity, to constitute a defense to crime, must be legal insanity as distinguished from what might be loosely or colloquially termed medical insanity."); *Fuller v. State*, 423 S.W.2d 924, 925 (Tex.Cr.App. 1968) ("[M]ere mental deficiency or derangement, though it may constitute a form of insanity known to and recognized by medical science, does not excuse one for crime."); *State v. Crenshaw*, 98 Wash.2d 789, 793 (1983) ("The insanity defense is not available to all who are mentally deficient or deranged; legal insanity has a different meaning and a different purpose than the concept of medical insanity."); *Simecek v. State*, 243 Wis. 439, 447 (1943) ("[U]nder the settled law of this state insanity from a medical standpoint does not necessarily constitute legal insanity in the sense that it constitutes a defense in prosecutions for crime. One may be medically insane and yet be criminally responsible for his acts.").

5. This analogy comes from Herbert Fingarette, *The Meaning of Criminal Insanity* (Berkeley: University of California Press, 1972), pp. 38–39.

6. *See* Dressler, *supra* n. 3, at 363.

7. William Lambard, *Eirenarcha, or the Offices of the Justices of Peace* (London: Newbery and Binneham, 1581).

8. *Regina v. M'Naghten*, *supra* n. 1, at 722.

9. On Gettier conditions, see Robert K. Shope, *The Analysis of Knowing: A Decade of Research* (Princeton: Princeton University Press, 1983).

10. This case is described by Richard Bonnie, "The Moral Basis of the Insanity Defense," 69 *ABA Journal* 194 (Feb. 1983).

11. Baker might have thought either that her aunt was the devil or that her aunt was a human person "possessed" by the devil. In the former but not the latter case, Baker might not know that she was shooting a human person.

12. Her act did not meet the legal requirements for self-defense because, although her motive was self-preservation, it was unreasonable for her to think that her life was in immediate danger. Still, her act might have been justified self-defense if her beliefs had been reasonable and true.

13. *Regina v. M'Naghten*, *supra* n. 1. *See also Webb v. State*, 270 Ga. 556, 557 (1999) ("[I]t has long been recognized that 'if the delusion is as to a fact which would not excuse the act with which the prisoner is charged, the delusion does not authorize an acquittal of the defendant.'") (citations omitted); *Finger v. State*, 117 Nev. 548, 576 (2001) ("[I]f a jury believes [that the defendant] was suffering from a delusional state, and if the facts as he believed them to be in his delusional state would justify his actions, he is insane and entitled to acquittal. If, however, the delusional facts would not amount to a legal defense, then he is not insane. Persons suffering from a delusion that someone is shooting at them, so they shoot back in self-defense are insane under M'Naghten. Persons who are paranoid and believe that the victim is going to get them some time in the future, so they hunt down the victim first, are not."); *Overton v. State*, 56 S.W.2d 740, 741 (Tenn. 1933) ("Overton, being possessed by an insane delusion that Scott was seeking his life, should be judged as if Scott was actually seeking his life; that, believing Scott was attempting his life, Overton was justified, according to this belief, in cutting up Scott by way of defense.... 'Under [the M'Naghten case] and those following it a homicide committed under an insane delusion is excusable, if the notion embodied in the delusion and believed to be a fact, if a fact indeed, would have excused the defendant.'") (citation omitted);

14. Consider also a contemporary murderer who has an insane delusion such as that he is Jack the Ripper. *See* Fingarette, *supra* n. 5.

15. Notice that this point does not rule out moral testimony. Just as we can know the temperature on the basis of testimony, even though to know the temperature is not to know what anyone believes about the temperature, so too we can know that an act is wrong on the basis of testimony, when we are justified in trusting the person who gives the testimony, even though knowing that the act is wrong is distinct from knowing that anyone believes that it is wrong.

16. *See* Shope, *supra* n. 9.

17. In applying their own moral standards, judges and juries need not claim that their moral beliefs are objective, much less universal or infallible. The issue here is only about *whose* moral standards—society's, the defendant's, or the legal decision-maker's—are used to determine whether the defendant's moral beliefs count as moral knowledge. To hold that juries and judges may or should use their own moral standards in this way is not to assume that their moral standards are true or objective, although this approach is sometimes (mis)understood in this way. *See, e.g.*, *State v. Bott*, 310 Minn. 331, 336 (1976) ("The instruction used, although it employed the phrase 'contrary to law,' was clearly directed to the issue whether defendant appreciated his act was wrong in a moral sense. The trial court's use of this phrase was clarified by its explanation that '(k)nowledge that the act was wrong refers to the moral side.' The court apparently accepted the philosophic concept of a natural law under which certain rules of conduct—moral rules—inhere throughout the very fabric of the universe and which legislatures and courts have undertaken to reflect. An act which contravenes a properly enacted criminal statute, in the view of the naturalist, contravenes the moral law as well. While we are persuaded that the jury understood the instruction in this way, we think the better practice is to instruct the jury in the terms of the statute but with the explanation that the word 'wrong' is used in the moral sense and does not refer simply to a violation of statute.").

18. For the two courts that have explicitly refrained, see *Ivery v. State*, 686 So.2d 495, 501 (Ala. Crim. App. 1996); and *State v. Abercrombie*, 375 So.2d 1170, 1178–79 (La. 1979).

19. *See, e.g.*, *State v. Hamann*, 285 N.W.2d 180, 183–184 (Iowa 1979); *Finger v. State*, *supra* n. 13, at 576; *State v. Hodgen*, 47 N.C. App. 329, 334 (1980); *Ruffin v. State*, 270 S.W.3d 586, 592 (Tex. Cr. App. 2008); *State v. Crenshaw*, *supra* n. 4, at 793–796.

20. *See, e.g., U.S. v. Ewing*, 494 F.3d 607, 620 (7th Cir. 2007) ("'[C]riminality' or 'contrary to law' is too narrow a definition of wrongfulness, and 'subjective personal morality' is too broad. The second of the alternative definitions of wrongfulness—contrary to objective societal or public morality—best comports with the rules established in M'Naghten's case."); *State v. Tamplin*, 986 P.2d 914 (Ariz. Ct. App. 1999); *People v. Stress*, 205 Cal.App.3d 1259, 1274 (Cal. Ct. App. 1989) ("Although seldom addressed by the courts, which have generally left the word 'wrong' undefined in jury instructions, the question is whether moral wrong is to be judged by society's generally accepted standards of moral obligation or whether the subjective moral precepts of the accused are to be employed. While the inherent 'slipperiness' of the terminology in this area may leave some doubt, it appears most courts mean that the defendant is sane if he knows his act violates generally accepted standards of moral obligation whatever his own moral evaluation may be."); *People v. Serravo*, 823 P.2d 128, 134–37 (Colo. 1992); *State v. Worlock, supra* n. 4, at 609–11; *People v. Wood*, 12 N.Y.2d 69, 76 (1962); *People v. Adams*, 26 N.Y.2d 129, 135–136 (1970).

21. Some retributivists might argue that the purpose of punishment is to restore justice by counterbalancing what Kant calls "inner viciousness" in *The Metaphysical Elements of Justice*, trans. John Ladd (Indianapolis: Bobbs-Merrill, 1965), p. 103. On this view, people are not morally responsible or, therefore, criminally punishable for doing what they did not or could not personally believe to be morally wrong. But this position would produce practical problems. *See, e.g., People v. Coddington*, 23 Cal.4th 529, 608–609 (Cal. 2000), *rev'd on other grounds* by *Price v. Superior Court*, 25 Cal.4th 1046 (2001) ("'[M]oral obligation in the context of the insanity defense means generally accepted moral standards and not those standards peculiar to the accused.'... 'The fact that a defendant claims and believes that his acts are justifiable according to his own distorted standards does not compel a finding of legal insanity.' As we explained in [a previous case], this aspect of the M'Naghten test... is necessary 'if organized society is to formulate standards of conduct and responsibility deemed essential to its preservation or welfare, and to require compliance, within tolerances, with those standards.'") (citations omitted); *People v. Wood*, 12 N.Y.2d 69, 77 (1962) ("[T]he jury... could reasonably have found that defendant was operating under a standard of morality he had set up for himself and which applied only to him. The law does not excuse for such moral depravity.")

22. *See, e.g., State v. Ulm*, 326 N. W. 2d 159, 161 (Minn. 1982) ("To convict a defendant who raises mental illness defense, defendant must know that his act was wrong in a moral sense and not merely know that he has violated a statute."); S.C. Code Ann. § 17-24-10(A) (2009) ("It is an affirmative defense to a prosecution for a crime that, at the time of the commission of the act constituting the offense, the defendant, as a result of mental disease or defect, lacked the capacity to distinguish moral or legal right from moral or legal wrong or to recognize the particular act charged as morally or legally wrong.").

23. *Rex v. Hadfield*, 27 How. St. Tr. 1281 (K.B. 1800).

24. *See supra* n. 13 and surrounding text.

25. *See, e.g., McElroy v. State*, 242 S.W. 883, 884 (Tenn. 1922) ("It is no new thing for criminals to attempt to justify their conduct upon the excuse that they acted on the command of God. They have frequently claimed to have seen visions and to have gotten instructions from the Almighty to do wicked things.... Such excuses were not taken into account in olden times, nor do we think they can now be allowed to avail. Men cannot put themselves beyond the reach of the law by the indulgence of such vain imaginations. To permit this would impair the practical administration of justice.").

26. *See, e.g., People v. Skinner*, 39 Cal. 3d 765, 778–81 (1985) ("Where defendant, who killed his wife, clearly could not distinguish right and wrong with regard to his act, believing that marriage vow 'till death do us part' bestowed on marital partner God-given right to kill other partner if he or she was inclined to violate marital vows, defendant was entitled to judgment of not guilty by reason of insanity, without any further hearing to determine whether he was capable of knowing or understanding nature and quality of his act."); *People v. Schmidt*, 216 N.Y. 324, 340 (N.Y. 1915) ("Obedience to the law is itself a moral duty. If, however, there is an insane delusion that God has appeared to the defendant and ordained the commission of a crime, we think it cannot be said of the offender that he knows the act to be wrong."); *State v. Potter*, 68 Wash. App. 134, 149 (1992) ("[T]he trial court properly instructed the jury that if the deific decree overcame the defendant's cognitive ability to know that his act was 'wrong,' the jury could find him not guilty by reason of insanity."); *State v. Cameron*, 100 Wash.2d 520, 526–27 (1983) ("[O]ne who believes that he is acting under the direct command of God is no less insane because he nevertheless knows murder is prohibited by the laws of man. Indeed, it may actually emphasize his insanity. In the instant case, there is considerable evidence (although not unanimous) from which the jury could have concluded that petitioner suffered from a mental disease; that he believed his stepmother was satan's angel or a sorceress; that he believed God directed him to kill his stepmother; that because of the mental disease it was impossible for him to understand that what he was doing was wrong; and... that his free will had 'been subsumed by [his] belief in the deific decree.'") (citations omitted)

27. *See, e.g., State v. Wilson*, 242 Conn. 605, 616 (1997) ("[A] defendant may establish that he lacked substantial capacity to appreciate the 'wrongfulness' of his conduct if he can prove that, at the time of his criminal act, as a result of mental disease or defect, he substantially misperceived reality and harbored a delusional belief that society, *under the circumstances as the defendant honestly but mistakenly understood them*, would not have morally condemned his actions.") (emphasis in text); *People v. Serravo, supra* n. 20, at 139 ("In our view, the 'deific-decree' delusion is not so much an exception to the right-wrong test measured by the existing societal standards of morality as it is an integral factor in assessing a person's cognitive ability to distinguish right from wrong with respect to the act charged as a crime.").

28. *See, e.g., State v. Worlock, supra* n. 4, at 611 ("In the exceptional case, such as the deific exception in which the defendant claims that he or she acted under a command from God, the court should instruct the jury that 'wrong' encompasses both legal and moral wrong."); *Cunningham v. State*, 1879 WL 3956, at *6 (Miss. 1879) ("If a crazed enthusiast violates the law, impelled by a madness which makes him deem it the inspired act of God, he has only done that which his diseased and deluded imagination taught him was right; and if the act would be proper in one so divinely inspired, and was the direct and necessary consequence of the delusion, there can be no punishment, because, however rational on other subjects, he was on that subject incapable of having a criminal intent."). *See also supra* n. 13.

29. On the relationship between reasons-responsiveness and ability, capacity, and control, see John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998); and Philip Pettit, *A Theory of Freedom: From the Psychology to the Politics of Agency* (Cambridge: Polity Press, 2001). Of course, reasons-responsiveness comes in degrees, and someone might respond to certain reasons but not others. We will ignore such complications here.

30. *See U.S. v. Freeman, supra* n. 4, at 624 ("[T]he gravamen of psychiatric objections to the M'Naghten Rules... was not that they looked to the cognitive feature of the

personality, undeniably a significant aspect of the total man, but that they looked to this element exclusively.") (citation omitted); *State v. Gardner*, 616 A.2d 1124, 1126 (R.I. 1992) ("M'Naghten was predicated upon an outdated psychological concept. As we stated, '[i]nsanity affects the whole personality of the defendant, including will and emotions,' not just cognitive capacity.") (citation omitted); *State v. Colby*, 6 Ohio Misc. 19, 22 (1966) ("The Court has reached the conclusion that the present 'right and wrong' test, the M'Naghten rules, for deciding criminal responsibility is 'based on an entirely obsolete and misleading conception of the nature of insanity, since insanity does not only, or primarily, affect the cognitive or intellectual faculties, but affects the whole personality of the patient, including both the will and the emotions. An insane person may therefore know the nature and quality of his act and that it is wrong and forbidden by law, and yet commit it as a result of the mental disease.'") (citation omitted).

31. David Margolick, "Lorena Bobbitt Acquitted in Mutilation of Husband," *New York Times* (January 22, 1994), pp. 1, 7. Cognitive disease might also have been present because "[i]n testimony, Mrs. Bobbitt said she had not realized what she had done until later when she fled their home and was in her car." There was no evidence, however, that any cognitive deficit contributed to her act.

32. *See* Joel Feinberg, "What Is So Special about Mental Illness?" in *Doing and Deserving* (Princeton: Princeton, 1970), pp. 272–292.

33. *See* John Deigh, "Moral Agency and Criminal Insanity," in his *Emotions, Values, and the Law* (New York: Oxford University Press, 2008), pp. 196–219.

34. *Smith v. U.S.*, 36 F.2d 548, 549 (D.C. 1929).

35. Some tests in this tradition refer only to weakening of resistance rather than to increase of impulse. *See, e.g., Davis v. U.S.*, 165 U.S. 373, 378 (1897) (holding that a defendant is not responsible if his will "has been otherwise than voluntarily so completely destroyed that his actions are not subject to it, but are beyond his control.").

36. *See* M. T. Gailliot, N. L. Mead, and R. F. Baumeister, "Self-Regulation," in *Handbook of Personality: Theory and Research, 3d ed.*, O. P. John, R. W. Robbins, and L. A. Pervin eds. (New York: Guilford Press, 2008), pp. 472–491.

37. *Parsons v. State*, 81 Ala. 577, 597 (1887).

38. Our emphasis.

39. *Durham v. U.S.*, 214 F.2d 862, 874–875 (1954). There was precedent for this proposition in *State v. Pike*, 49 N.H. 399, 402 (1969); *see also Royal Commission on Capital Punishment 1949–53, Report 131* (Cmd. 8932) (London: HMSO, 1953).

40. *Carter v. U.S.*, 252 F.2d 608, 617 (D.C. 1957). This decision raises the problem of deviant causal chains, discussed above.

41. *See, e.g., U.S. v. Brawner*, 471 F.2d 969, 977–78 (D.C. 1972).

42. *See supra* n. 4 on the issue of whether insanity is a legal or medical classification.

43. *McDonald v. U.S.*, 312 F.2d 847, 851 (D.C. 1962).

44. *See, e.g., State v. Johnson*, 121 R.I. 254, 268 (1979).

45. American Law Institute, *Model Penal Code* (Philadelphia; The American Law Institute, Final Draft, 1962), § 4.01(1).

46. American Law Institute, *Model Penal Code* § 4.01(2). As noted in *State v. Johnson*, *supra* n. 44, at 269 n. 10, seventeen states that adopted the Model Penal Code rule added this clause by 1979: Alaska, Arkansas, Connecticut, Hawaii, Illinois, Indiana, Maine, Maryland, Missouri, Montana, Oregon, Tennessee, Texas, Utah, Vermont, Wisconsin, and Wyoming. Still, this clause remains controversial.

47. On nonbehavioral manifestations of psychopathy, see *The Psychopath: Emotion and the Brain*, by James Blair, Derek Mitchell, and Karina Blair (Oxford: Wiley-Blackwell,

2005). Whether or not psychopaths should be found not guilty by reason of insanity is discussed and debated in a special issue of *Neuroethics* 1, 3 (2008), pp. 149–212.

48. *See State v. Johnson, supra* n. 44, at 262 n. 4.

49. What counts as substantial is not clear, but it is illuminating that a minority of the American Law Institute proposed alternative wording: "A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect, his capacity either to appreciate the wrongfulness of his conduct or to conform his conduct to the requirements of the law is so substantially impaired that he cannot justly be held responsible." This alternative wording might have helped to clear up the ambiguity between absolute and relative measures (discussed in the text).

50. For more detail on volitional incapacity, see Michael Corrado, "Responsibility and Control," 34 *Hofstra L. Rev.* 59 (2005); and "The Case for a Purely Volitional Insanity Defense," 42 *Tex. Tech L. Rev.* 481 (2009).

51. Arkansas, Iowa, Maryland, Nevada, Oregon, and Virginia. *See* Ark. Code Ann. § 5-2-312(a)(1)(B) (2009) ((a)(1) "It is an affirmative defense to a prosecution that at the time the defendant engaged in the conduct charged he or she lacked capacity as a result of mental disease or defect to: (A) Conform his or her conduct to the requirements of law; or (B) Appreciate the criminality of his or her conduct."); *State v. Dye*, No. 08–0887, 2009 WL 3337617, at *4 (Iowa Ct. App., Oct. 7, 2009) ("The words 'right' and 'wrong' in [Iowa Code Ann. §] 701.4 refer to legal right and wrong, not moral right and wrong."); Md. Code Ann., [Criminal Procedure] § 3-109(a)(1) (2009) ((a) "A defendant is not criminally responsible for criminal conduct if, at the time of that conduct, the defendant, because of a mental disorder or mental retardation, lacks substantial capacity to: (1) appreciate the criminality of that conduct"); Nev. Rev. Stat. Ann. § 174.035(5)(b)(2) (2009) ((5) "Under [a plea or defense of not guilty by reason of insanity], the burden of proof is upon the defendant to establish by a preponderance of the evidence that: (a) Due to a disease or defect of the mind, he was in a delusional state at the time of the alleged offense; and (b) Due to the delusional state, he either did not: (1) Know or understand the nature and capacity of his act; or (2) Appreciate that his conduct was wrong, meaning not authorized by law."); Or. Rev. Stat. Ann. § 161.295(1) (2009) ("A person is guilty except for insanity if, as a result of mental disease or defect at the time of engaging in criminal conduct, the person lacks substantial capacity either to appreciate the criminality of the conduct or to conform the conduct to the requirements of law."); Va. Code Ann. § 4801(a)(1) (2009) ("(a) The test when used as a defense in criminal cases shall be as follows: (1) A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks adequate capacity either to appreciate the criminality of his conduct or to conform his conduct to the requirements of law.").

52. As noted in *State v. Johnson, supra* n. 44, at 269 n. 9, seventeen states that adopted the Model Penal Code rule chose the term "wrongfulness": Alaska, Connecticut, Delaware, Hawaii, Idaho, Indiana, Maine, Michigan, Missouri, New York, Ohio, Tennessee, Texas, Utah, Wyoming, West Virginia, and Wisconsin.

53. States that interpret "wrongfulness" in the Model Penal Code rule in terms of the community's or society's sense of morality include Arizona, California, Colorado, Connecticut, Indiana, New York (which provides for an insanity *defense* that is a hybrid of M'Naghten and Model Penal Code), Rhode Island, South Carolina, and Washington. *See, e.g., State v. Wilson*, 242 Conn. 605, 616 (1997) ("[W]e conclude that a defendant may establish that he lacked substantial capacity to appreciate the 'wrongfulness' of his conduct if he can prove that, at the time of his criminal act, as a result of mental disease or defect, he substantially misperceived reality and harbored a delusional belief that society, under

the circumstances as the defendant honestly but mistakenly understood them, would not have morally condemned his actions.").

54. *See supra* n. 21 for possible reasons.

55. *See, e.g., U.S. v. Freeman, supra* n. 4, at 622 n. 52 ("We have adopted the word 'wrongfulness' in Section 4.01 as the American Law Institute's suggested alternative to 'criminality' because we wish to include the case where the perpetrator appreciates that his conduct is criminal, but, because of a delusion, believes it to be morally justified."). Failure to appreciate moral wrongness would also be enough in jurisdictions that excuse those who lack capacity to appreciate either moral or legal wrongness. *See, e.g.,* S.C. Code Ann. § 17-24-10(A) (2009) ("It is an affirmative defense to a prosecution for a crime that, at the time of the commission of the act constituting the offense, the defendant, as a result of mental disease or defect, lacked the capacity to distinguish moral or legal right from moral or legal wrong or to recognize the particular act charged as morally or legally wrong.").

56. *See* Antony Duff, *Punishment, Communication, and Community* (New York: Oxford University Press, 2000). Of course, neither moral condemnation nor retributive motives are always perfectly sensitive to the mental problems that afflict perpetrators. We often are so angry or devastated that we want to hold killers responsible even when they are not actually morally responsible and even when we know that they are not morally responsible. This natural psychological tendency, however, conflicts with our moral belief that just punishment requires moral responsibility.

57. *See, e.g., State v. Wilson, supra* n. 27, at 614 ("As Herbert Wechsler, chief reporter for the Model Penal Code, stated in his model jury charge: 'To appreciate the wrongfulness of conduct is, in short, to realize that it is wrong; to understand the idea as a matter of importance and reality; to grasp it in a way that makes it meaningful in the life of the individual, not as a bare abstraction put in words.'") (citation omitted); *People v. Adams,* 26 N.Y.2d 129, 135 (1970) ("'A new dimension is accorded the word 'know' by following it with 'or appreciate.' This is designed to permit the defendant possessed of mere surface knowledge or cognition to be excused, and to require that he have some understanding of the legal and moral import of the conduct involved if he is to be held criminally responsible.'") (citation omitted); *State v. Dyer,* 16 Or. App. 247, 258 (1974) ("The effect of the word 'understand' in the [State v.] Gilmore[, 242 Or. 463 (1996)] formulation is to allow a full range of testimony as to emotional as well as intellectual cognition of the act, a practice which the drafters wished to continue by using 'appreciate.'"); *State v. Johnson, supra* n. 44, at 268 (1979) ("Our [new] test [of the insanity defense] consciously employs the more expansive term 'appreciate' rather than 'know.' Implicit in this choice is the recognition that mere theoretical awareness that a certain course of conduct is wrong, when divorced from appreciation or understanding of the moral or legal impact of behavior, is of little import.").

58. Robert Hare, *Without Conscience* (New York: Atria, 1993), p. 125.

59. American Law Institute, *Model Penal Code and Commentaries* (1985), § 4.01(1), p. 166; *see also State v. Wilson, supra* n. 27, at 614 ("The drafters of the Model Penal Code purposefully adopted the term 'appreciate' in order to account for the defendant whose 'detached or abstract awareness' of the wrongfulness of his conduct 'does not penetrate to the affective level.'").

60. *See* Christopher Slobogin, "The Integrationist Alternative to the Insanity Defense," 30 *Am. J. Crim. L.* 315 (2003).

61. *See supra* n. 30 on emotional mental illness, and Terry Maroney, "Emotional Competence, 'Rational Understanding,' and the Criminal Defendant," 43 *Am. Crim. L. Rev.* 1375 (2006).

62. *See, e.g.*, Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Avon Books, 1994), pp. 212–217.

63. *See* Christopher Slobogin's arguments that deterrability is essential for responsibility in "A Jurisprudence of Dangerousness," 98 *Nw. U. L. Rev.* 1 (2003).

64. States whose insanity defenses contain only a cognitive prong, usually M'Naghten, and lack a volitional prong include California, Louisiana, Minnesota, Nebraska, New York, North Carolina, Ohio, Oklahoma, South Dakota, Texas, and Washington. *See* Ohio Rev. Code Ann. § 2945.391 (2009) ("Proof that a person's reason, at the time of the commission of an offense, was so impaired that the person did not have the ability to refrain from doing the person's act or acts, does not constitute a defense."); Okla. Stat. Ann. tit. 21, § 154 (2009) ("A morbid propensity to commit prohibited acts existing in the mind of a person who is not shown to have been incapable of knowing the wrongfulness of such acts forms no defense to a prosecution therefor."); S.D. Codified Laws § 22-5-7 (2009) ("A morbid propensity to commit prohibited acts existing in the mind of a person who is not shown to have been incapable of knowing the wrongfulness of such acts forms no defense to a prosecution therefor."); *People v. Severance*, 138 Cal.App.4th 305, 324 (2006) ("The irresistible impulse test... has long been discredited in California as a test for legal insanity."); *State v. George*, 768 So.2d 748, 757 (La. Ct. App. 2000) ("[U]nder the law and jurisprudence of this state, the defendant cannot claim the defense of irresistible impulse."); *State v. LaTourelle*, 343 N.W.2d 277, 282 (Minn. 1984) ("This court has rejected the 'irresistible impulse' definition."); *State v. Canbaz*, 270 Neb. 559, 574 (2005) ("the theory of 'irresistible impulse'... is a defense not recognized in Nebraska."); *People v. Hakner*, 34 N.Y.2d 822, 823 (1974) (noting that New York Legislature rejected irresistible-impulse defense); *State v. Helms* 284 N.C. 508, 514–515 (1974) ("Under the law of this State there is no halfway house on the road to insanity which affords sanctuary to those who know the right and still the wrong pursue.... '[T]he law is far from excusing criminal acts committed under the impulse of [irresistible] passions.'") (citation omitted); *Saenz v. State*, 879 S.W.2d 301, 308 (Tex. Crim. App. 1994) (irresistible impulse is "doctrine that is no longer recognized in Texas."); *State v. Edmon*, 621 P.2d 1310, 1314 (Wash. Ct. App. 1981) ("This 'irresistible impulse' defense is not accepted in Washington.").

65. 18 U.S.C. § 20. This move was anticipated in *U.S. v. Lyons*, 731 F.2d 243, 15 Fed. R. Evid. Serv. 859 (5th Cir. 1984) (en banc).

66. *See* Feinberg, *supra* n. 32; Fingarette, *supra* n. 5; Michael Moore, *Law and Psychiatry* (Cambridge: Cambridge University Press, 1974); Charles Culver and Bernard Gert, *Philosophy and Medicine* (New York: Oxford University Press, 1982); Jennifer Radden, *Madness and Reason* (London: Allen and Unwin, 1985); and Stephen Morse, "Diminished Rationality, Diminished Responsibility," 1 *Ohio St. J. Crim. L.* 289 (2003).

67. For more arguments against assimilating insanity and irrationality, see Walter Sinnott-Armstrong, "Insanity vs. Irrationality," 1 *Public Affairs Quarterly* 1 (1987); and Slobogin, n. 60.

68. *See* Herbert Fingarette and Ann Fingarette Hasse, *Mental Disabilities and Criminal Responsibility* (Berkeley: University of California Press, 1979).

69. *See Jones v. U.S.*, *supra* n. 3. Prosecutors could be authorized or even required to bring civil commitment proceedings in parallel with criminal proceedings whenever the defendant's mental condition poses a danger to others. This proposal, however, would hardly be free of practical problems.

70. Bonnie, *supra* n. 10.

71. *See* Fischer and Ravizza and Pettit, *supra* n. 29. Harry Frankfurt ("Alternate Possibilities and Moral Responsibility," 66 *Journal of Philosophy* 829 [1969]) and his followers have proposed various counterexamples to the proposition that moral responsibility entails the power to do otherwise, but they need not detain us here because they are not directly relevant to mental illness or conditions of legal responsibility.

72. Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (Garden City, N.Y.: Doubleday, 1961, originally published 1789) and Slobogin, "A Jurisprudence of Dangerousness," n. 63.

73. *See* Jean Hampton, "The Moral Education Theory of Punishment," 13 *Philosophy and Public Affairs* 208 (1984).

74. *See* H. L. A. Hart, "Principles of Punishment," in *Punishment and Responsibility* (Oxford: Oxford University Press, 1968), pp. 18–20.

75. Fakers or malingerers were cited often after Hinckley as reasons to abolish the insanity defense. *See, e.g.,* Orrin Hatch, "The Insanity Defense Is Insane," *Reader's Digest* (Oct. 1982).

76. Pre-Hinckley abolitionists include Thomas Szasz, *Law, Liberty, and Psychiatry* (New York: Collier-Macmillan, 1963); Goldstein and Katz, "Abolish the Insanity Defense, Why Not?" 72 *Yale L.J.* 853 (1963); Alan Dershowitz, "Abolish the Insanity Defense," 9 *Crim. L. Bull.* 434 (1973); and Norval Morris, *Madness and the Criminal Law* (Chicago: University of Chicago Press, 1982). Morris and Bonnie criticize each other's views in "Debate: Should the Insanity Defense Be Abolished?" 1 *Journal of Law and Medicine* 117 (1986–87).

77. Three states have abolished the insanity defense: Idaho, Kansas, and Montana. Defendants in these states may still use mental illness to show "diminished capacity" to form the specific intent that is required for the offense being charged.

78. *See American Bar Association Policy on the Insanity Defense,* Approved by the ABA House of Delegates, February 9, 1983; American Psychiatric Association, *Statement on the Insanity Defense* (Washington, D.C.; American Psychiatric Association, 1982).

79. To date, the only state that seems to have adopted the Model Penal Code's version of the insanity defense but without the volitional prong is New York. *See also U.S. v. Lyons,* n. 65.

80. States that allow a verdict of guilty but mentally ill include Alaska, Delaware, Georgia, Illinois, Indiana, Kentucky, Maryland, Michigan, Mississippi, Nevada, New Mexico, Oregon, Pennsylvania, South Carolina, South Dakota, Utah, and Washington. Most of these states allow not guilty by reason of insanity as well. *See* "'Guilty but Mentally Ill' Statutes: Validity and Construction," 71 A.L.R. 4th 702; "Mental Illness Not Amounting to Insanity; 'Guilty but Mentally Ill' Verdict," 21 Am. Jur. 2d, Criminal Law § 38; "Guilty but Mentally Ill or Mentally Retarded," 22 C.J.S., Criminal Law § 130.

81. *See, e.g., Kansas v. Crane,* 534 U. S. 407 (2002); and *Kansas v. Hendricks,* 521 U. S. 346 (1997).

82. Dressler, *supra* n. 3, at 369. Recall that the burden was on the defense in the original M'Naghten rule.

83. *But see* Eule, "The Presumption of Sanity: Bursting the Bubble," 25 *UCLA L. Rev.* 637 (1978).

84. In *Leland v. Oregon,* 343 U.S. 790 (1952), the Supreme Court held that Oregon may require a defendant raising an insanity defense to prove insanity beyond a reasonable doubt. *See also* James M. Varga, "Due Process and the Insanity Defense; Examining Shifts in the Burden of Persuasion," 53 *Notre Dame Lawyer* 123 (1977). Of the forty-seven states that provide for an insanity defense, most of them explicitly designate it as an affirmative defense, a defense that the defendant has the affirmative burden of proving, whether

beyond a reasonable doubt, by clear and convincing evidence, or by a preponderance of the evidence. Only two states, Nevada and Utah, tend to treat the insanity defense not so much as an affirmative defense but more, like the three abolitionist states (Idaho, Kansas, and Montana), as a diminished capacity defense. The difference is that the diminished capacity defense does not impose an affirmative burden on the defendant but rather constitutes a means by which the defendant can try to establish that the prosecutor has not met *her* affirmative burden of proving that defendant had, at the time of the alleged offense, the specific intent required to be guilty of the offense.