

The effect of mental illness under US criminal law

PAUL H ROBINSON

Colin S Diver Professor of Law, University of Pennsylvania

The effects of mental illness can sometimes make it impossible for the state to prove the culpability requirements for an offence. For example, an actor who hallucinates that a knife is a clothes brush may not have the required culpability for homicide if he kills someone thinking that he is brushing lint from the victim's chest. Similarly, mental illness can mitigate murder to a lesser form or to manslaughter if the actor killed under the influence of an 'extreme *mental* or emotional disturbance'. Finally, mental illness can form the basis for a general excuse, for example, the insanity defence. Unlike the other two doctrines, the insanity defence operates without regard to – that is, despite the defendant's satisfaction of – the elements of the offence definition (indeed, the excuse is only necessary if the defendant otherwise satisfies the offence requirements). In order to successfully raise the insanity defence, the actor need only satisfy the conditions set out in the defence provision.

1 General excuse of insanity

The insanity defence reflects the standard structure and requirements common in disability excuses, namely, that a *disability* – in this instance, mental disease or defect – causes an *excusing condition*, that is, a particular kind of dysfunction in relation to the offence conduct.

A. THE REQUIRED DISABILITY: MENTAL DISEASE OR DEFECT

In this context, 'mental disease or defect' is a legal concept, not a medical one, and is thus for the jury rather than medical experts to resolve – though the jury will no doubt be influenced by the expert witnesses that they hear. Many experts testifying as to whether the defendant suffers from a mental disease or defect will rely on the classification system contained in the *Diagnostic and Statistical Manual* of the American Psychiatric Association (APA), now in its fifth edition (DSM-5).¹ The APA gives the following definition of 'mental disorder':

A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor

1 APA, *Diagnostic and Statistical Manual of Mental Disorders* xi–xl 5th edn (American Psychological Publishing 2013) (hereafter DSM-5).

or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above.²

Intoxication can cause mental dysfunction, but it is commonly excluded as a basis for the insanity defence, because it is not a form of mental disease or defect.³ It is dealt with instead under the law's special intoxication defence. The habitual and excessive use of intoxicants, however, may cause a mental disease with resulting dysfunction apart from the intoxication, and this mental disease can be the basis for an insanity defence.⁴ Addiction, for example, has been recognised as a mental disease.⁵

The insanity defence also typically excludes any 'abnormality manifested only by repeated criminal or otherwise anti-social conduct'.⁶ In other words, being a habitual criminal is not in itself a sufficient indication of a cognizable disability. Such an abnormality may be a mental disease for clinical treatment purposes, but to recognise it as a mental disease for the purposes of the insanity defence would generate results inconsistent with the theory of excuses as serving to exculpate blameless offenders. Such habitual criminality by itself may be fully volitional conduct, and thus fully blameworthy.

B. THE REQUIRED EXCUSING CONDITIONS: ALTERNATIVE FORMULATIONS

It is not enough for the defence that an actor suffers from a mental disease or defect, even one that causes some dysfunction. To be held blameless, the actor's mental illness must cause effects so strong that it would not be reasonable to expect the actor to have avoided the criminal law violation. This *excusing condition*, or required effect of the mental illness, has been formulated in several different ways for the insanity defence. The most significant tests include: the McNaghten test; the McNaghten test plus the 'irresistible impulse' test; the Durham 'product' test; the Model Penal Code formulation; and the more recent federal insanity formulation.

In *McNaghten's Case*, the House of Lords held that an actor has a defence of insanity if '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, [he] did not know he was doing what was wrong*'.⁷ This 1843 formulation has been adopted and maintained by many US jurisdictions today, as detailed below.

As the quote indicates, this test can be satisfied in two ways: the mental disability may prevent the defendant from understanding (1) the nature or (2) the wrongfulness of his or her conduct. For both instances, the focus is on the defendant's impaired perception or cognition. The McNaghten test was an advance over prior case law, which set the standard for the defence as having no more understanding than 'an infant, a brute, or a wild beast'.⁸ McNaghten gave the jury specific criteria to focus on rather than vague analogies.

2 DSM-5 20.

3 Model Penal Code § 2.08(3).

4 See e.g. *United States v Lyons* 731 F 2d 243 (5th Cir 1984) (en banc) (evidence on lack of substantial capacity to appreciate criminality of conduct due to physiological impairment due to drug addiction can be submitted to jury); *People v Griggs* 17 Cal 2d 621 (1941) (insanity from long continued intoxication must be treated similarly by the court as insanity produced by another cause).

5 See DSM-5 485–87. Some cases, however, expressly reject the notion that addiction can qualify an actor for an insanity defence. See e.g. *United States v Moore* 486 F 2d 1139 (DC Cir 1973).

6 Model Penal Code § 4.01(2).

7 *McNaghten's Case* [1843] 8 Eng Rep 718, 722 [author's emphasis added].

8 *Arnold's Case* [1724] 16 How State Tr 695, 765.

As early as 1887, the McNaghten test was criticised as failing to reflect advances in the behavioural sciences. Mental illness, it was observed, can remove the *power to choose* as well as the *knowledge* of one's situation or of right and wrong.⁹ To permit a defence in cases involving loss of the power to choose, some jurisdictions supplemented the McNaghten test with what is sometimes described as the 'irresistible impulse' test.¹⁰ Under this formulation – under which a 'control prong' is said to be added to McNaghten's 'cognitive prong' of the insanity defence – an actor obtains the defence if he or she satisfies the McNaghten test or:

(1) if, by reason of the duress of 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*.¹¹

The McNaghten-plus-irresistible-impulse test was criticised in turn as not fully reflecting still more recent advances in the behavioural sciences. For example, the court in *Durham v United States* observed that mental dysfunctions, of both the cognitive and control types, are always a matter of degree and are not, as was previously thought, absolute in their effect.¹² Further, the court found that the McNaghten and irresistible-impulse tests improperly focus on particular symptoms rather than on the key question of whether the mental illness, whatever its nature, *had the effect of causing the offence*. *Durham* then articulated a 'product' test for insanity, under which an accused 'is not criminally responsible if his unlawful act was the product of mental disease or mental defect'.¹³

Durham, however, was widely criticised as overstating the grounds of exculpation. The critics argued that it should not suffice that the mental illness is a 'but for' cause of the offence; the mental illness must also cause a degree of impairment sufficiently severe that it renders the defendant blameless for the offence – a reasonable person suffering this kind and degree of dysfunction could not reasonably have been expected to have avoided the violation. The product test was adopted in only a few jurisdictions, and arguably remains in use in only one.¹⁴

In *United States v Brawner*,¹⁵ the Court of Appeal for the District of Columbia Circuit overruled its earlier decision in *Durham*, rejected the *Durham* 'product test' and adopted instead the test contained in Model Penal Code § 4.01(1) (also known as the American Law Institute or ALI test):

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]*¹⁶ *of his conduct or to conform his conduct to the requirements of law.*

9 *Parson v State* 2 So 854 (AL 1887).

10 The phrase 'irresistible impulse' is something of a misnomer. As the quotation in the text illustrates, nothing in the defence requires that the 'duress of mental illness' be impulsive. It may be, and frequently is, a long process that both creates the mental illness and has the mental illness cause the criminal conduct.

11 *Parson* (n 9) 866–67 [author's emphasis added]

12 *Durham v United States* 214 F 2d 862 (DC Cir 1954).

13 *Ibid* 874.

14 *State v Shackford* 506 A 2d 315 (NH 1986) (holding that the insanity test is a matter 'to be weighed by the jury upon the question whether the act was the offspring of insanity', quoting *State v Jones* 50 NH 369, 398–99 (1871)); *State v Pike* 49 NH 399 (1870).

15 471 F 2d 969 (DC Cir 1972).

16 The alternative language provided by the Code – 'criminality [wrongfulness]' – arises from disagreement over whether the test should look to the actor's awareness that the conduct is legally wrong or that it is morally wrong. For a discussion of the issue, see e.g. *State v Crenshaw* 659 P 2d 488 (WA 1983).

This formulation concedes that there are *degrees* of impairment, as *Durham* had emphasised, but also requires a minimum degree of impairment: namely, the actor must ‘lack substantial capacity’ to behave properly. The Model Penal Code test reverts to the structure of the McNaghten-plus-irresistible-impulse test in specifically noting that the dysfunction may affect either cognitive or control capacities. It differs from McNaghten-plus-irresistible-impulse, however, in that the earlier formulation would appear to require absolute dysfunction, i.e. the total absence of knowledge of criminality or the total loss of power to choose.¹⁷ The Model Penal Code test, in contrast, requires only that the actor lack ‘substantial capacity’ to control his conduct or ‘appreciate’ its criminality. As detailed further below, the test has gained wide acceptance, rivalling the popularity of the McNaghten and McNaghten-plus-irresistible-impulse formulations.

Another formulation of the insanity defence, which was considered but rejected by the ALI, calls for the jury’s general assessment of an actor’s responsibility and blameworthiness for the offence. It would provide the defence if the actor, because of mental disease or defect, lacked sufficient capacity to be ‘justly held responsible’ for his or her conduct.¹⁸ The approach is similar to the Model Penal Code’s approach in other contexts, where the Code’s rules explicitly call on the decision-maker for a normative judgment. Its causation test, for example, asks the jury to decide whether a result’s occurrence is too remote or accidental ‘to have a [just] bearing on the actor’s liability or on the gravity of his offence’.¹⁹ Such an open formulation was rejected in the insanity context, however, because it was thought that the jury could and should be given greater guidance. The version ultimately included in the Code focused the jury’s attention on the nature and effect of the dysfunction – with specific reference to cognitive or control dysfunction – and avoided having a jury incorporate considerations of general sympathy (or bias) that might slip in under the broader ‘justice’ standard.

Some jurisdictions that previously adopted the Model Penal Code test have cut back on it. For example, the federal insanity statute, which was enacted by Congress to replace the holding in *Branner*, noted above, which adopted the ALI test, follows the ALI’s ‘appreciates’ language, rather than McNaghten’s ‘know’ language, thereby seeming to allow exculpation for degrees of cognitive dysfunction short of complete loss.²⁰ On the other hand, the federal statute drops the ‘lacks substantial capacity’ language, which makes it closer to the apparently absolute requirement of McNaghten. Most importantly, the federal formulation drops the control prong of the defence altogether, reverting to the single cognitive prong. This reflects scepticism as to whether behavioural scientists can measure an actor’s degree of control impairment and as to whether jurors can understand testimony about, or effectively judge, a defendant’s degree of impairment.²¹

A few jurisdictions have abolished the insanity defence (but continue to allow mental disease or defect to provide a defence if it negates a required offence culpability element,

17 Some writers have observed that the irresistible impulse test may not be as absolute in application as it appears. By requiring that the actor has no power to choose, it certainly urges the jury to be demanding in the level of dysfunction that they will require, but it seems unusual, if not impossible, that an actor would lose *all* power to choose. Typically, control impairments make an actor’s decisions to remain law-abiding more difficult, but rarely take away all decision. See Model Penal Code § 4.01, comment 3 (1985) (explaining that a workable test calling for complete loss of ability to know or control is not possible and that such a test would impose unrealistic restriction on scope of proper inquiry).

18 This proposal appears as alternative (a) to para (1) of Model Penal Code § 4.01 (Tentative Draft No 4, 1955).

19 Model Penal Code § 2.03(2)(b) and (3)(b). Similar instances of broad language calling for a normative judgment can be found in Model Penal Code §§ 2.02(2)(c) and (d) (defining recklessness and negligence), 2.12(2) (defining *de minimis* infraction), 3.02(1)(a) (defining lesser evils defence) and 2.09(1) (defining duress defence).

20 18 USC § 17.

21 See e.g. *Lyons* (n 4).

as discussed in Part 2 below).²² Constitutional challenges to such abolition have been successful in some cases,²³ but not in others.²⁴ Whether or not the federal or state constitutions bar it, abolition is a questionable policy. To the extent that the criminal law claims to express conclusions about an actor's blameworthiness – the characteristic that traditionally has distinguished criminal law from civil law – it cannot impose criminal liability and punishment on clearly insane offenders without destroying its moral credibility. If society has a need to protect itself against dangerous persons who are predicted to commit future crimes, it can and should do so. Typically, dangerous persons are blameworthy offenders. In the unusual case where an offender is dangerous yet blameless, as is true for some insane offenders, civil commitment is an alternative means of protecting the community while retaining the criminal law's moral credibility.²⁵

No single insanity formulation is dominant. Twenty-two of the 52 jurisdictions apply the traditional McNaghten test,²⁶ with three adding the irresistible-impulse element.²⁷ Of the 17 jurisdictions adopting a control prong, 14 have done so by adopting the somewhat broader Model Penal Code (ALI) formulation. Thus, just under a third of the states with insanity defences have adopted the ALI formulation in its entirety.²⁸ Four jurisdictions have no insanity defence, though they continue to allow mental illness to negate an offence culpability element.²⁹ One state adopts what appears to be something close to the *Durham* product test.³⁰

The most prominent alternative formulations of the insanity defence are summarised in Table 1 overleaf.

C. ADDING A VERDICT OF 'GUILTY BUT MENTALLY ILL'

In the 1970s and 1980s, a number of jurisdictions adopted a verdict of 'Guilty But Mentally Ill' (GBMI).³¹ The verdict replaces the insanity defence in only a few states. More frequently, it provides the trier of fact with an additional verdict in cases where mental

22 See Idaho Code § 18-207; Kan Stat Ann § 22-3220; Mont Code Ann § 46-14-102; Utah Code Ann § 76-2-305(1); See, generally, Paul H Robinson, *Criminal Law Defences* vol 2 (West Group 1984) § 173(a) n 5.

23 See e.g. *State v Strasburg* 110 P 1020 (WA 1910).

24 See e.g. *State v Korell* 690 P 2d 992 (MT 1984).

25 See Paul H Robinson, 'Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice' (2001) 114 *Harvard Law Review* 1429; Paul H Robinson, 'The Criminal-Civil Distinction and Dangerous Blameless Offenders' (1993) 83 *Journal of Criminal Law and Criminology* 693.

26 Ariz Rev Stat Ann § 13-502; Cal Penal Code § 25; Colo Rev Stat Ann § 16-8-101; Fla Stat Ann § 775.027; Ga Code Ann § 16-3-2; Iowa Code Ann § 701.4; La Rev Stat Ann § 14; Minn Stat Ann § 611.026; *Roundtree v State* 568 So 2d 1173 (MS 1990); Mo Ann Stat § 552.030 (modifying the standard language slightly to 'incapable of knowing and appreciating'); *State v Harms* 650 N W 2d 481 (NE 2002); N J Stat Ann § 2C:4-1; Ohio Rev Code Ann § 2901.01; Okla Stat Ann tit 21 § 152; 18 Pa Cons Stat Ann § 314; SC Code Ann § 17-24-10; SD Codified Laws § 22-1-2; Tex Penal Code Ann § 8.01; *Price v Commonwealth* 323 SE 2d 106 (VA 1984); Wash Rev Code Ann § 9A.12.010.

27 *State v White* 270 P 2d 727 (NM 1954).

28 Ark Code Ann § 5-2-312; Conn Gen Stat Ann § 53a-13; *Patton v US* 782 A 2d 305 (DC 2001); Haw Rev Stat Ann § 704-400; Ky Rev Stat Ann § 504.020; Md Code Ann, Crim Proc § 3-109; *Commonwealth v Brown* 434 NE 2d 973 (MA 1982); Mich Comp Laws Ann § 768.21a; Or Rev Stat § 161.295; *State v Johnson* 399 A 2d 469 (RI 1979); Vt Stat Ann Tit 13 § 4801; *State v Samples* 328 SE 2d 191 (WV 1985); Wis Stat Ann § 971.15; Wyo Stat Ann § 7-11-304.

29 Idaho Code § 18-207; Kan Stat Ann § 22-3220; Mont Code Ann § 45-2-101; Utah Code Ann § 76-2-305.

30 *State v Shackford* 506 A 2d 315 (NH 1986).

31 See Lisa A Callahan et al, 'Measuring the Effects of the Guilty But Mentally Ill (GMI) Verdict: Georgia's 1982 Reform' (1992) 16 *Law and Human Behavior* 447; Ronnie Mackay and Jerry Kopelman, 'The Operation of the "Guilty But Mentally Ill" Verdict in Pennsylvania' (1988) 16 *Journal of Psychiatry and Law* 247, 248.

| | Absolute | Substantial impairment |
|------------------------------|---|--|
| Cognitive dysfunction | McNaghten 'does not know nature of conduct or that it is wrong' | ALI part 1 'lacks substantial capacity to <i>appreciate</i> criminality of his conduct' |
| | Federal: 'unable to appreciate' criminality . . . | |
| Control dysfunction | + irresistible impulse 'lost the power to choose right from wrong' | + ALI part 2 'lacks substantial capacity to conform his conduct to the requirements of law' |

Table 1: Alternative formulations of the insanity defence

illness is an issue. The special verdict may be returned where a defendant is mentally ill, but where his or her mental illness is insufficient to provide either an insanity defence or a defence of mental illness negating an offence element (discussed in part 2). Following a GBMI verdict, the court typically has the same sentencing options that would follow from a typical 'guilty' verdict. GBMI convicts must be examined by psychiatrists before beginning to serve the sentence and, if found to be in need of treatment, are then imprisoned in a criminal mental health facility. In most jurisdictions, however, such evaluation and treatment occurs for *all* convicted offenders, not just those receiving a GBMI verdict,³² in which case a GBMI verdict is indistinguishable from a guilty verdict in its practical effect. (As noted above, similar civil commitment required-examination procedures also often exist for defendants *acquitted* under a Not Guilty by Reason of Insanity (NGRI) verdict.) In fact, although the GBMI verdict may seem designed to help mentally ill convicts, one of the few practical distinctions between GBMI convicts and typical sane convicts is that the GBMI convicts tend to receive *longer* sentences.³³

The GBMI verdict raises some significant concerns. First, one must question why the fact-finder in a criminal trial is an appropriate body to determine whether an offender is in need of a psychiatric examination. The expertise of the jury is in finding the facts of past events and in applying that community's notion of blameworthiness. The need for psychiatric treatment is a clinical issue, appropriate for prison psychiatrists, for example. It

32 See, for example, Cal Penal Code § 2684 (prescribing terms for transfer to state hospital of mentally ill prisoners); DC Code Ann § 24-302; Mich Comp Laws Ann §§ 330.2001-330.2006; 50 Pa Cons Stat Ann § 4408; see also National Advisory Commission on Criminal Justice Standards and Goals, Corrections 184 (1973).

33 See Henry J Steadman et al, *Before and after Hinckley: Evaluating Insanity Defence Reform* (Guildford Press 1993) 8. This is likely due to a suspicion that mentally ill individuals are unusually dangerous and need to be incapacitated to prevent them from committing more crimes.

is not within the lay judgment of the jury,³⁴ and asking the jury to undertake such an inquiry can distract and confuse it in its task of assessing blameworthiness.

A second concern arises from an analysis of the legislative history for the GBMI verdict. The history suggests that the verdict was not even designed to perform such a psychiatric screening function, but rather arose as a device to reduce NGRI acquittals after constitutional mandates limited the use of civil commitment to preventively detain disturbed offenders.³⁵ The limitations on civil commitment were thought to create a risk of dangerously insane acquittees being released back into the community. Adding the GBMI verdict combats this perceived danger indirectly, not by loosening civil commitment standards, but by diverting mentally ill offenders from civil commitment to the criminal justice system. A jury choosing between an NGRI verdict and a GBMI verdict may select the latter, not because the jury finds the defendant blameworthy, but because the latter verdict seems to guarantee what may be the obvious need for treatment and confinement.

The difficulty with the GBMI verdict is that it invites jurors to consider matters unrelated to blameworthiness at a time when blameworthiness should be the sole issue before them. Moreover, the verdict plays on jurors' ignorance of the consequences of an NGRI verdict (or a standard 'guilty' verdict), encouraging the misperception that a GBMI verdict is the only way to incapacitate dangerously mentally ill persons while also providing necessary psychological treatment. (Adding to the potential confusion is the likelihood that the jury will inadvertently confuse the statutory definition of 'mental illness', relevant to the GBMI verdict, and the definition of 'mental disease or defect', relevant to the insanity defence.) The use of such an improper compromise verdict may do as much to undermine the insanity defence as total abolition would. If effective abolition is the objective, abolishing the insanity test openly would better further the interests of informed debate and reform.

The underlying purpose of adding the GBMI verdict, reducing insanity acquittals, is driven by fears that the insanity defence is granted too often and possibly subject to abuse. Yet the empirical evidence suggests that such fears are ill-founded. For example, people generally believe, inaccurately, that the insanity defence is a commonly offered defence in criminal trials: one study found that people estimate that 38 per cent of all defendants charged with crime plead NGRI.³⁶ In reality, an insanity plea is exceedingly rare, raised in a

34 Referrals for such professional evaluation of offenders who may need treatment can be done more effectively and efficiently by the court after receiving the pre-sentence report. Indeed, several jurisdictions have established specific post-trial procedures to provide treatment for mentally ill offenders who are sentenced to prison.

35 See e.g. Donald H J Hermann and Yvonne S Sor, 'Convicting or Confining? Alternative Directions in Insanity Law Reform: Guilty But Mentally Ill Versus New Rules for Release of Insanity Acquittes' (1983) *BYU L Rev* 499, 582 ('The rationale for the GBMI verdict stems from a legislative concern that the insanity defense is too easily proved, while the abolition of automatic commitment of insanity acquittees in some states has made civil commitment of persons found NGRI more difficult.'). See also *People v Ramsey* 375 NW 2d 297 (MI 1985) (a major purpose of GBMI statute is to lessen the number of persons relieved of all criminal responsibility by way of NGRI verdict); *State v Neely* 819 P 2d 249, 252 (NM 1991) (suggesting that legislature's purpose in enacting GBMI statute was 'to reduce the number of improper or inaccurate insanity acquittals and to give jurors an alternative to acquittal when mental illness is believed to play a part in an offence'); *State v Hornsby* 484 SE 2d 869, 872 (SC 1997) (purposes of GBMI statute were to reduce the number of insanity acquittals and provide mental health care for GBMI inmates); *Robinson v Solem* 432 NW 2d 246, 248 (SD 1988) ('[O]ur legislature intended to provide an alternative verdict available to a jury to reduce the number of offenders who were erroneously found not guilty by reason of insanity?'); *People v Smith* 465 NE 2d 101, 106 (Ill App 1984) ('In the instant case, the legislature intended to provide a statute that reduced the number of persons who were erroneously found not guilty by reason of insanity and to characterize such defendants as in need of treatment.').

36 See Valerie P Hans, 'An Analysis of Public Attitudes toward the Insanity Defence' (1986) 24 *Criminology* 393, 406; see also Eric Silver et al, 'Demythologizing Inaccurate Perceptions of the Insanity Defence' (1994) 18 *Law and Hum Behavior* 63, 67–68.

fraction of a per cent of even felony cases.³⁷ Also contrary to popular belief, in the few cases where an insanity plea is introduced, more than half involve non-violent offences.³⁸ In addition, even in the rare cases in which the insanity defence is sought, it is usually not granted,³⁹ yet the public perception is that it is commonly successful.⁴⁰ Claims that the defence is abused and employed to manipulate juries are also belied by the fact that most NGRI pleas are not contested,⁴¹ and the vast majority of NGRI verdicts – 93 per cent, in one study – are reached through negotiated pleas or rendered by judges in bench trials, rather than by juries.⁴² The evidence directly refutes fears of rampant abuse and courtroom manipulation; in fact, most NGRI acquittees have significant prior histories of treatment for mental illness.⁴³

Protecting the public from potentially violent offenders, sane or insane, is an important, indeed irreproachable, goal. The GBMI verdict may protect the public from some dangerously insane offenders, but it does so not through rational reform of civil commitment, but rather by subverting the insanity defence and thereby perverting the criminal justice system to condemn, through criminal conviction, violators who may be blameless. Such condemnation of blameless offenders may have the long-term effect of weakening the criminal law's moral credibility, undermining its general condemnatory force and ability to harness the powerful forces of social influence and internalized norms.

The proper solution to the problem of dangerous but insane offenders lies not in the distortion of criminal justice, but in the adoption of civil commitment standards and procedures that will adequately protect offenders and the public. While some serious constitutional limitations on civil commitment do exist, the Supreme Court has held that the same limitations do not apply to commitment after an acquittal based on an insanity defence. Civil commitment after an NGRI verdict is made easier in part because the insanity acquittee's past offence provides evidence of dangerousness that may not exist in the normal civil commitment case.⁴⁴ These relaxed requirements are enough to protect the community through civil commitment without the need to distort the criminal justice process.

37 See Lisa A Callahan et al. 'The Volume and Characteristics of Insanity Defence Pleas: An Eight-State Study' (1991) 19 Bull Am Acad Psychiatry and L 331, 334. Note that this is less than 1 per cent of all *felony* cases, while the lay subjects estimated insanity pleas for 38 per cent of all persons charged with *any* crime. See also Richard A Pasewark and Hugh McGinley, 'Insanity Plea: National Survey of Frequency and Success' (1985) 13 J Psychiatry and L 101 (reporting median rate of 1 plea per 873 reported crimes); Stephen G Valdes, 'Frequency and Success: An Empirical Study of Criminal Law Defences, Federal Constitutional Evidentiary Claims, and Plea Negotiations' (2005) 153 U Pa L Rev 1709 (note) (study of 400 judges, prosecutors, and defence counsel reports insanity claims offered in less than 1 per cent of cases).

38 See Steadman et al (n 33) 111; see also Callahan et al (n 37) 336.

39 One study reports that the average acquittal rate for an insanity plea is 26 per cent. See Callahan et al. (n 37) 334. Pasewark and McGinley report a success rate of 15 per cent of pleas. See Pasewark and McGinley (n 37) 106; Valdes (n 37) (reporting success rate of 24 per cent).

40 See e.g. Hans (n 36) 406 (reporting study indicating that public believes over 36 per cent of all NGRI claims, constituting perceived 14 per cent of all criminal cases, result in NGRI verdict); Mary Frain, 'Professor Says Insanity Defence Seldom Works' *Telegram and Gazette* (Worcester, MA, 19 January 1996) B1 (quoting chair of psychiatry at the University of Massachusetts Medical Center as saying that general public believes the insanity defence is used in 20 to 50 per cent of all criminal cases).

41 See Michael J Perlin, 'A Law of Healing' (2000) 68 University of Cincinnati Law Review 407, 425 ('Nearly 90% of all insanity defence cases are "walkthroughs" – stipulated on the papers.').

42 See Callahan et al (n 37) 334.

43 See e.g. Michael R Hawkins and Richard A Pasewark, 'Characteristics of Persons Utilizing the Insanity Plea' (1983) 53 Psychology Reports 191, 194; Steadman et al. (n 33) 56.

44 See, generally, Robinson, 'Punishing Dangerousness' (n 25).

2 Mental illness negating an offence element

Just as an actor's mistake can cause her to lack the culpability required for an offence, so too can an actor's mental illness cause the absence of an offence culpability requirement.⁴⁵ If an actor is hallucinating and believes she is hitting moles, when she is in fact lethally beating her daughter, she does not have the culpable state of mind – knowingly causing death of another person – required for the offence of murder. The hallucination induced by her mental illness 'negates' (shows that she did not have) the culpable state of mind required for the offence.

Model Penal Code § 4.02 expressly authorises this use of mental illness evidence:

Evidence that the defendant suffered from a mental disease or defect is admissible whenever it is relevant to prove that the defendant did or did not have a state of mind which is an element of the offense.

As in the case of mistake, one might argue that there is no need for this type of section. Model Penal Code § 1.12(1) already provides that:

No person may be convicted of an offense unless each element of such offense is proved beyond a reasonable doubt. In the absence of such proof, the innocence of the defendant is assumed.

Thus, even absent a special provision, an actor cannot be convicted without proof of all elements, and the absence of any element – such as a culpability element, because of mental illness – will provide a defence. As will become apparent in the following discussion, however, the Model Code's provision affirmatively permitting mental illness to negate a required culpability element was a wise addition because it would alter the previously existing law in many jurisdictions.

This absent element defence – mental illness negating a required culpability element – is given a special name in many jurisdictions, such as 'diminished capacity', 'diminished responsibility', 'partial responsibility', or 'partial insanity', yet, all of these labels can be misleading. The terms *diminished* and *partial* suggest that the defence is designed to provide a mitigation or partial defence, perhaps for mental illness short of insanity. But there is nothing partial about the defence.⁴⁶ The actor's mental illness either negates a required element of the offence charged or it does not. The doctrine accordingly either provides a complete defence to the offence charged or it does not. An actor may end up with liability for a lesser offence, of course, if his or her mental illness is such that it negates the culpability of the offence charged but not the culpability required for a lesser offence. On the other hand, the mental illness might be such that it negates the culpability required for all lesser offences or it might be such that it does not negate the culpability of any offence, including the most serious offence charged.

The latter case illustrates the important limitations of the defence: an actor may be seriously mentally ill yet have no diminished capacity defence if, despite mental illness, the actor satisfies the required culpability elements of the offence charged. Mental illness that impairs an actor's ability to control his or her conduct, for example, is unlikely to negate an offence culpability requirement, because such requirements typically concern specific cognitive functioning (for example, being aware of facts or consequences) rather than matters of control. Typically, only cognitive dysfunction will cause an actor not to know the nature, circumstances, or potential consequences of his or her conduct, and therefore not satisfy a culpability element.

45 A hypothetical commonly given in the literature to illustrate this kind of situation is the man who, because of mental illness, believes he is squeezing a lemon when in fact he is squeezing his wife's neck.

46 Paul H Robinson, *Criminal Law Defences* vol 1 (West Group 1984) § 22.

The history of the doctrine of diminished capacity is for the most part the history of attempts to *limit* the use of evidence of mental illness to negate an offence element. American jurisdictions take a variety of positions on the admissibility of evidence of mental illness to negate an offence element. About 40 per cent of the jurisdictions, typically those with modern criminal codes, admit any evidence of mental disease or defect that is relevant to negate any culpability element of an offence,⁴⁷ as the Model Penal Code recommends. Another 30 per cent allow such evidence to be admitted, but purport to limit such admission to negating only ‘specific intent’⁴⁸ – a concept that has little meaning in modern codes – or, even more restrictively, to negate only the malice or premeditation element of murder (in jurisdictions whose definition of murder requires such elements).⁴⁹ The final 30 per cent purport to exclude the admission of mental-illness evidence to negate any offence element.⁵⁰ While some of these efforts have been held unconstitutional,⁵¹ the US Supreme Court in *Clark v Arizona* held that the federal constitution did not require states to allow admission of evidence of mental illness.⁵²

The common law treated diminished capacity as analogous to voluntary intoxication, where culpability was imputed to the actor. However, the analogy is flawed. The imputation of culpability may well be justified in the context of voluntary intoxication; the actor has culpably caused his own intoxication and that culpability ought to be taken into account. Current law commonly takes this as a rationale for imputing to him recklessness as to the

47 See Alaska Stat § 12.47.020; Ark Code Ann § 5-2-303; *State v Burge* 487 A 2d 532 (CT 1985); Colo Rev Stat Ann § 18-1-803; Haw Rev Stat Ann § 704-401; Idaho Code § 18-207; *Robinson v Commonwealth* 569 SW 2d 183 (Ky App 1978); Me Rev Stat Ann tit 17-A § 38; *Hoey v State* 536 A 2d 622 (MD 1988); Mo Ann Stat § 552.030; Mont Code Ann § 46-14-102; *Finger v State* 27 P 3d 66 (NV 2001) (finding abolition of the insanity defence unconstitutional and holding that evidence not meeting the legal insanity standard may be admitted at trial to negate an offence element); *Novosel v Helgemoe* 384 A 2d 124 (NH 1978) (applying only in bifurcated trials); NJ Stat Ann § 2C:4-2; Or Rev Stat § 161-300; *State v Perry* 13 SW 3d 724 (TN Crim App 1999); Utah Code Ann § 76-2-305; *State v Smith* 396 A 2d 126 (VT 1978); *United States v Pohlott* 827 F 2d 889 (3d Cir 1987) (holding that in codifying an insanity excuse, 8 USCA § 17, Congress abolished defences of ‘diminished capacity’ and ‘partial responsibility’ but did not intend to preclude admission of psychiatric evidence relevant to negate an element of the offence).

48 Cal Penal Code § 28; *Veverka v Cash* 318 NW2d 447 (IA 1982); *State v Dargatz*; 614 P 2d 430 (KS 1980); *People v Atkins* 325 NW 2d 38 (MI App 1982); *People v Segal* 429 NE 2d 107 (NY 1981); *Commonwealth v Walzack* 360 A 2d 914 (PA 1976); *State v Corraera* 430 A 2d 1251 (RI 1981); *State v Huber* 356 NW 2d 468 (SD 1984); *State v Bottrell* 14 P 3d 164 (WA App 2000).

49 *People v Leppert* 434 NE 2d 21 (Ill App 1982) (considering defendant’s claim that, due to mental defect, he lacked the requisite intent to attempt murder); *Commonwealth v Baldwin* 686 NE 2d 1001 (MA 1997); *Washington v State* 85 NW 2d 509 (NE 1957); *State v Beach* 699 P 2d 115 (NM 1985); *State v Shank* 367 SE 2d 639 (NC 1988); *LeVasseur v Commonwealth* 304 SE 2d 202 (VA 1979).

50 *Barnett v State* 540 So 2d 810 (AL Crim App 1988); *State v Schantz* 403 P 2d 521 (AZ 1965); *Bates v State* 386 A 2d 1139 (DE 1978); *Bethea v United States* 365 A 2d 64 (DC 1976); *Zamorra v State* 361 So 2d 776 (FL App 1978); *Hudson v State* 319 SE 2d 28 (GA App 1984); *Brown v State* 448 NE2d 10 (IN 1983); *State v Murray* 375 So 2d 80 (LA 1979); *State v Bouwman* 328 NW 2d 703 (MN 1982); *Garcia v State* 828 So 2d 1279 (Miss App 2002); *State v Wilcox* 438 NE2d 523 (Ohio 1982); *Gresham v State* 489 P 2d 1355 (Okla Crim App 1971); *Gill v State* 552 SE 2d 26 (SC 2001); *Warner v State* 944 SW 2d 812 (TX App 1997); *State v Flint* 96 SE 2d 677 (WV 1957) (providing statement against diminished capacity defence, which has since been questioned but not overruled, in *State v Simmon* 309 SE 2d 89 (WV 1983)); *Muench v Israel* 715 F 2d 1124 (7th Cir 1983) (finding that Wisconsin may constitutionally reject the diminished capacity defence and refuse to admit evidence proving defendant’s inability to form requisite intent); *Price v State* 807 P 2d 909 (WY 1991). To date, North Dakota courts have not explicitly spoken to this issue – their position remains unclear.

51 See e.g. *Hendersbott v People* 653 P 2d 385 (CO 1982) (finding unconstitutional Colorado statute barring evidence of mental illness to negate *mens rea* requirement for nonspecific intent crimes); *Finger* (n 47) (holding Nevada statute unconstitutional). But see *Muench* (n 50) (rejecting a constitutional challenge to Wisconsin’s practice of excluding evidence of mental illness relevant to a *mens rea* requirement).

52 *Clark v Arizona* 548 US 735 (2006).

objective elements of an offence, for example. While one may question some aspects of this rationale for imputing culpability, it provides at least the semblance of a rational reason.

No similar claim can be made in the context of mental illness negating an offence element. An actor is rarely to blame for causing his own mental disease or defect. What, then, is the rationale for treating the mentally ill actor as if he has a required culpability when in fact he does not? As with all doctrines of imputation, the process of imputation in itself is not objectionable (this is the basis for complicity liability, for example), but it may become so if not adequately justified.

One form of attack on the use of mental-illness evidence to negate an offence element is to claim incompatibility between behavioural science and criminal law. It is argued, for example, that behavioural science admits gradations of responsibility while the criminal law does not; it must decide to impose liability or not.⁵³ But this argument rests upon a mistaken view of the diminished capacity defence. As noted above, the actor's mental illness either negates an offence element or it does not; it does not ask the criminal law to admit of gradations.

Another argument is that the behavioural sciences are not yet sophisticated enough for the criminal law to rely upon them. A similar argument stresses the tendency of the diminished capacity doctrine to take 'full decision-making authority' from the jury and shift it to the expert witness.⁵⁴ But what is asked is not to have criminal law rely on behavioural science; what is asked is that juries be given access to such evidence along with all other relevant evidence, so the *jury* can decide whether the required offence mental element is present. The jury has full authority to reject any psychiatric evidence (and is particularly likely to do so where psychiatrists disagree, as frequently occurs).⁵⁵

Another challenge to allowing mental illness to negate culpability focuses on the nature of culpability: it is a legal (and moral) concept, not a scientific one. Accordingly, it is argued, culpability must necessarily be decided on an objective standard, and application of an objective standard means that personal abnormalities cannot be taken into account.⁵⁶ While this view of culpability might have been true at early common law, it does not accurately describe the nature of current doctrine. It is true that an actor's state of mind cannot be known directly, and it is true that culpability must be proven by objective *evidence*. However, this does not demand an objective standard for culpability. Current law rejects such common law rules as the presumption that 'an actor intends the natural and probable consequences of his conduct'.⁵⁷ It requires instead a finding by the jury that, based on all the evidence, the jury believes that the actor actually had the culpable state of mind required by the offence definition. The members of the jury may call upon their own life experiences in reaching their factual conclusions, including judgments about how people normally function. But the issue they are asked to decide is not whether the ordinary person would have had the required culpable state of mind, but rather whether this defendant actually did have it.

53 See *Bethea* (n 50). Following an argument with his estranged wife, defendant shot her five times at close range; at trial defendant claimed at the time of the shooting his mental condition was such as to preclude a finding of 'sound memory and discretion' and 'deliberate and premeditated malice' as required for the offence.

54 *Ibid* 89.

55 For a discussion of the role of experts and the jury in applying the insanity defence, see *Brunner* (n 15).

56 See *ibid* 1002.

57 See, for example, *Regina v Walllett* [1968] All ER 296 (the Criminal Justice Act of 1967 excluded common law presumption; jury instruction on ordinary person standard for determining intent in murder case was in violation of Act; murder conviction reduced to manslaughter); *Sandstrom v Montana* 442 US 510 99 S Ct 2450, 61 L Ed 2d 39 (1979) (jury instruction in accordance with common law presumption had effect of either conclusive presumption of intent or shift of burden of persuasion, and therefore unconstitutional because violates 14th Amendment requirement that state prove every offence element beyond reasonable doubt).

One final form of challenge, to some the most persuasive arguments against permitting mental illness to negate an element, focuses on the need to protect society from the mentally ill person who commits a crime. We must bar the defence of diminished capacity because such a dangerous person must be convicted of something in order to provide authority for incarceration or treatment or whatever else is necessary to prevent him from causing further harm. Further, the danger to the public in permitting such a defence is particularly high because everyone who commits a brutal offence suffers some degree of mental abnormality, so allowing such a defence would frustrate the needed criminal law jurisdiction over those from whom we most need to protect ourselves.

Even if it were true that everyone who commits a brutal offence is mentally abnormal, it does not follow that all such persons will get a defence of this sort. Only those who are mentally ill and, as a result, do not have a *required offence element*, are entitled to a defence for mental illness negating an element. Further, 'mental illness' typically is defined expressly to exclude abnormality manifested only by anti-social conduct. Most important, the proper way to protect ourselves from dangerous mentally ill people is not through distortion of the criminal law system by convicting blameless people, but rather through providing an effective system for civil commitment of the dangerously mentally ill.⁵⁸ (Note that this protection-of-the-public line of argument would call for abolition of not only the use of mental illness negating an element, but also of the insanity defence.)

Some courts have concluded that barring the use of mental illness evidence to show the absence of a required offence element is unconstitutional.⁵⁹ This is said to follow from the cases holding that the state is constitutionally required to prove all elements of an offence beyond a reasonable doubt. Such a constitutional rule may be broader than is appropriate; it would seem to bar all forms of imputation of a required offence element. Many doctrines of imputation – such as the doctrine of complicity, which imputes the conduct of another – are well justified and universally accepted. If there is to be a constitutional rule, it ought to focus instead upon the adequacy of the justification for the imputation. Given the difficulty in showing a basis of blameworthiness of an actor whose mental illness negates an offence element, imputation of the negated element seems unwise. Whether the rejection of such bad policy ought to be enshrined as a constitutional rule is another matter.

3 Extreme mental or emotional disturbance manslaughter

Criminal homicide occurs when one culpably 'causes the death of another human being'.⁶⁰ For *murder* liability to arise, causing death must be the actor's 'conscious object' (purpose), or he must be 'practically certain' (knowing) that his conduct will cause death.⁶¹ Some codes limit murder to the intentional (purposeful) form.⁶²

58 See Paul H Robinson, *Distributive Principles of Criminal Law* (OUP 2008) ch 6; Paul H Robinson, 'Punishing Dangerousness' (n 25); *Branner* (n 15) 1429.

59 See, for example, *Hendershott* (n 51) (denial of defendant's request to present mental impairment evidence to negate requisite culpability held violation of due process; exclusion of mental impairment evidence rendered prosecution's *mens rea* evidence uncontestable as matter of law and lessened prosecution's burden to something less than mandated by due process). The US Supreme Court has not yet taken a position on the issue. Some federal circuit courts have held, however, that it is a violation of due process if jury instructions put the burden to prove mental illness on the defendant where the jury might conclude that this relieves the state of proving all required culpability elements. *Humanik v Beyer* 871 F 2d 432 (3rd Cir) cert denied, 493 US 812 (1989).

60 Model Penal Code § 210.1(1).

61 See Model Penal Code § 2.02(2)(a)(i) and (b)(ii) (defining purposely and knowingly as to a result).

62 See e.g. NY Penal Code § 125.25(1) ('intent'); Ohio Rev Code Ann § 2903.02(A) ('purposely'). Some of these states do not refer to 'knowing' killings, but define murder also to include some non-intentional killings, such as the depraved indifference category. See e.g. NY Penal Code § 125.25(2).

A common exception to the paradigm of an intentional killing as murder is found in the common law doctrine of provocation, under which intentional killings would be mitigated from murder to the lesser crime of *manslaughter* – specifically, what is commonly known as *voluntary manslaughter* – if the killer was ‘provoked’ into committing the crime. The mitigation reflected the position that passion frequently obscures reason and, in some limited way, renders the provoked intentional killer less blameworthy than the unprovoked intentional killer.

Modern codes, following the Model Penal Code, give a broader mitigation than the common law provocation doctrine. The Model Penal Code’s manslaughter mitigation applies where:

murder is committed *under the influence of extreme mental or emotional disturbance for which there is reasonable explanation or excuse*. The reasonableness of such explanation or excuse shall be determined from the viewpoint of a person in the actor’s situation under the circumstances as he believes them to be.⁶³

This modern formulation certainly covers all cases in which the common law would have given a mitigation, but goes much further, to provide a much broader basis for the mitigation.

The doctrine has two components. First, the killing must have been committed ‘under the influence of extreme mental or emotional disturbance’. A defendant will not be eligible for the mitigation if he did not personally suffer such a disturbance or if it did not drive or dictate his act, even if the circumstances would have created such a disturbance in most other people and would have driven them to violence. Second, there must be a ‘reasonable explanation or excuse’ for the disturbance. No mitigation is available if the disturbance has no reasonable basis or is peculiar to the actor.⁶⁴

The Model Penal Code broadens the common law mitigation in several important respects. Unlike the common law rules, it does not explicitly require, or exclude, particular situations; there are no conditions that are inadequate as a matter of law to provide a mitigation. It also drops the common law rule barring the mitigation if the killing occurs some period of time after the provoking event. In other words, the Code postulates that an actor’s mental or emotional disturbance does not necessarily decrease with time; indeed, it might increase.⁶⁵

Further, nothing in the Code’s mitigation limits it to cases where the actor kills the source of the provocation, as the common law does. The Code’s position is that if the actor’s killing is less blameworthy by virtue of the influencing conditions, then such reduced blameworthiness exists no matter who is killed. Indeed, the Code does not even require a provoking act as such; the relevant ‘disturbance’ may arise from any source so long as it satisfies the rule’s requirements. (The underlying theory of this version of the mitigation does not appear to be one related to a possible partial justification. Defensive force defences, for example, may justify force against an aggressor but not against anyone else.⁶⁶ The mitigation’s basis, rather, is more akin to excuse defences, which look to the actor rather than the objective circumstances and apply regardless of the identity of the offence’s victim.)⁶⁷

63 Model Penal Code § 210.3(1)(b).

64 See e.g. *People v Casassa* 404 NE 2d 1310 (NY 1980) (trial court found defendant acted under required disturbance but no reasonable explanation or excuse for it, thus denied mitigation; affirmed on appeal).

65 Model Penal Code § 210.3 comment at 48 (Tentative Draft No 9, 1959).

66 In some provocation situations, of course, an actor may also have a self-defence claim.

67 For an illustrative application of the Model Penal Code’s extreme emotional disturbance test, see e.g. *State v Ott* 686 P 2d 1001 (OR 1984) (defendant’s conviction for murder of wife reversed and new trial ordered because trial judge failed to permit jury to consider the ‘personal’ characteristics of defendant; extreme emotional disturbance must be judged from perspective of actor’s situation).

The Model Penal Code mitigation uses a ‘reasonableness’ standard, as the common law doctrine does, but instead of adopting a purely objective understanding of reasonableness, modern rules partially individualise the standard through the requirement that the reasonableness of the explanation or excuse is to be determined ‘under the circumstances as [the actor] believes them to be’ and ‘from the viewpoint of a person in the actor’s *situation*’. These two phrases provide significant opportunities for a court, or jury, to take account of the particular characteristics of the defendant and the specific conditions in which the defendant acted. The Code’s drafters intended the second phrase – in particular ‘in the actor’s situation’ – to permit a trial judge great leeway in partially individualising the reasonable-person standard.⁶⁸

Most of the applications of the mitigation will not involve mental illness; the majority will concern the classic cases of emotional rage or distress. But by broadening the mitigation to ‘extreme *mental or* emotional disturbance’, the modern formulation allows mental illness to be taken into account. This will include cases that previously would have gone under the label of ‘diminished responsibility’ or something of the sort.⁶⁹ It was common for states to allow this mental-illness mitigation to reduce first-degree murder (which often required premeditation) to second-degree murder, but only a minority took the Model Penal Code’s approach of allowing it to mitigate murder to manslaughter.⁷⁰

Conclusion

Here, then, are three kinds of doctrines that can allow mental illness to provide a mitigation or excuse under US criminal law, either by satisfying the special requirements of the extreme-mental-or-emotional-disturbance mitigation of murder to manslaughter, by negating a required offence culpability element, or by satisfying the conditions of a general insanity defence. In each instance, there is great variation in how the states formulate the doctrine. The most that can be said is that nearly all states have a general insanity defence, the majority of states allow mental illness to negate at least some offence elements, and most states allow some forms of mental illness to provide some kind of mitigation to murder.

68 As the Model Penal Code commentary explains: “There is an inevitable ambiguity in “situation.” If the actor were blind or if he had just suffered a blow or experienced a heart attack, these would certainly be facts to be considered in a judgment involving criminal liability, as they would be under traditional law. But the heredity, intelligence or temperament of the actor would not be held material in judging negligence, and could not be without depriving the criterion of all its objectivity. The code is not intended to displace discriminations of this kind, but rather to leave the issue to the courts.” Model Penal Code § 2.02 comment 4 (1985) 242. See also *State v Ott* (n 67) (noting that a similar problem exists with recklessness, and that discriminations similar to those required by the negligence standard must be made).

69 See Model Penal Code (n 68)

70 *Ibid.*