Northern Ireland Legal Quarterly

Volume 65 Number 2

EDITOR
PROFESSOR SALLY WHEELER

Queen's University Belfast
School of Law

Contents

Special issue: Mental Disorder and Criminal Justice

Guest Editors: Ben Livings and Nicola Wake

NILQ 65(2): 137–40

Editorial

Ben Livings

Senior Lecturer, University of New England, Australia

Nicola Wake

Senior Lecturer, University of Northumbria

The genesis and impetus for the works presented in this special edition of the *Northern Ireland Legal Quarterly* lie in the Mental Disorder and Criminal Justice conference convened by the guest editors and jointly supported by Northumbria University's Centre for Evidence and Criminal Justice Studies and the University of Sunderland. Whilst the works themselves stand as the contributions of the individual scholars, they have been influenced by and represent some of the views put forward at this conference, which was held at Northumbria University in Newcastle in October 2013.

The criminal law has long struggled to find a way to accommodate the mentally disordered defendant, both according to its substantive doctrinal strictures and in a broader procedural sense. It is telling that the M'Naghten Rules still stand as the leading authority when it comes to the defence of insanity in England and Wales, in spite of widespread criticism from practitioners and scholars. The publication of this special edition comes at a time of potential reform, however; in recent years, the Law Commission has undertaken extensive reviews of the current law on 'unfitness to plead' and 'insanity and automatism'. Previously, the Law Commission provided comprehensive reports on the partial defences to murder which fed into amendments made to that area under ss 52–56 of the Coroners and Justice Act (CJA) 2009.

The case law generated post the 2009 Act reforms has highlighted that it is essential this area of law reform is informed by an understanding drawn from a wide knowledge base and the collection of essays presented herein offers a range of perspectives on the place of mental disorder within the criminal justice system. This edition brings together scholarship from a number of jurisdictions, all of which, in their different ways, have sought to develop criminal justice responses to this intractable area of the criminal law.

The collection commences with a contribution from Thom Brooks. Whilst Brooks does not specifically engage with mental disorder, his work offers a way of looking at criminal responsibility that may provide a platform on which to structure and critique the response of the criminal law to mental disorder. 'What is wrong about the "criminal mind"?' asserts that those who ascribe to the retributivist model and justification for the criminal law argue for a strong link between a criminal's mindset at the time of an offence and the community's response through punishment. This is a claim about desert, where the retributivist claims that the deserving should be punished to the degree deserved. For Brooks, any judgment about desert requires consideration of factors concerning 'the

criminal mind', where mitigating and aggravating factors may change the available penal options because of a criminal's mindset. Brooks argues that this retributivist link has been widely influential and sentencing policy in England and Wales reflects it, but that it is based on a mistake. The retributivist link presupposes that a criminal possesses some degree of moral responsibility, explaining why mitigating factors may decrease possible penalties as these factors may be evidence for diminished moral responsibility. This is a problem because the absence of moral responsibility is no defence to most criminal offences, such as strict liability offences. Moreover, offences requiring intent are often set at such a low threshold as to elude the higher standard of moral responsibility that is appealed to by retributivists. Thus, it is not a crime's threat or harm to morals that is most salient, but instead its threat or harm to rights, grounded in autonomy. Brooks contends that a more fruitful approach to desert and sentencing practices would be rights-based; wrongs are in fact not found in the criminal mindset, but rather the potential and actual infringement of rights.

In 'The insanity defence in operation', Ronnie Mackay reflects upon his engagement with, and involvement in, law reform in successive projects, including the most recent attempts of the Law Commission. As a leading contributor to the reform agenda, Mackay brings a historical perspective, assessing the impact of successive legislation upon the use and success of the insanity plea in the courts of England and Wales. Mackay points out that little is known about how the M'Naghten Rules operate in practice; his paper is designed to redress this gap in knowledge by exploring available data on successful pleas of insanity over several decades. In so doing, he assesses the value of empirical studies and what they may bring to the reform agenda.

The joint contribution of John Child and Alan Reed, 'Automatism is never a defence', challenges the accepted view of the law of automatism as a defence. Child and Reed argue that automatism be viewed as relating to the requirements of a prima facie offence. They explain that, where the defendant is not at fault for her lack of voluntariness, the term 'automatism' is simply a shorthand explanation that the defendant does not satisfy an essential element of every offence; namely voluntary conduct. Where the defendant is at fault for his or her lack of voluntariness, the automatism rules (within the current law) become an inculpatory tool through which to substitute the missing offence elements and construct liability. Having recognised that automatism plays an inculpatory role within the law, Child and Reed analyse this role and conclude that it is defective, as prior fault automatism lacks the equivalent blameworthiness necessary to fairly substitute for even missing basic intent offence elements. From here, Child and Reed discuss the possibility of a new automatism offence, to recognise the criminal blameworthiness of the defendant's conduct in certain cases, but to do so in a coherent manner that appropriately criminalises and labels the defendant.

In 'A new partial defence for the mercy killer: revisiting loss of control', Ben Livings assesses the extent to which recent revisions to the partial defences of diminished responsibility and provocation (now loss of control) may provide a potential defence to the mercy killer. Despite its *de jure* classification as murder, mercy killing is rarely prosecuted as such, and it has become linked to mental disorder by virtue of a 'benign conspiracy', which allowed for its accommodation under the auspices of diminished responsibility. This approach may obviate the injustice that would undoubtedly be caused in the event of a murder conviction, but lacks transparency and may hinder the development of a better response through open debate. The inception of the CJA 2009 brings the possibility of change, as it seeks to narrow the applicability of the partial defences and bring clarity to their operation. The prevailing view appears to be that the

CJA 2009 has indeed narrowed the applicability of the partial defences, and that this may restrict the availability to mercy killers of an increasingly medicalised diminished responsibility plea. In the face of such a possibility, Livings contends that the move from provocation to loss of control, also ushered in by the CJA 2009, may have resulted in a plea that is broader in application, and which may avail the mercy killer. He argues that a view of the narrowing of the plea is based in assumptions of legislative intent and the baggage of history, neither of which have necessarily made it into the law in its new form. If loss of control is to apply, Livings suggests that it may prove a more appropriate avenue than that offered by the benign conspiracy.

The work James Chalmers presents in 'Insanity and automatism: notes from over the border and across the boundary' is comparative in nature, and looks to the Scottish experience of reform to the law of insanity and automatism, undertaken a few years before the Law Commission's recent discussion paper on the same subject. Chalmers examines statutory reform of the Scottish law of insanity, which replaced the common law plea with a new defence. Against this background, his work provides an account of the Scottish law of insanity and automatism, before discussing the recent statutory reforms, and offering contrasts with the proposals now suggested for English law. It concludes by discussing the area where the two reform projects diverge most significantly: the boundary between automatism and the reformed defences of (in Scotland) mental disorder excluding criminal responsibility and (in the Law Commission's proposals) 'not criminally responsible by reason of recognised medical condition'.

Offering another international and comparative perspective is Warren Brookbanks, whose contribution looks to a novel procedural aspect of mental disorder within the criminal justice system of New Zealand. In '"Special hearings" under New Zealand's Criminal Procedure (Mentally Impaired Persons) Act 2003', Brookbanks examines a particular aspect of the decade-long operation of legislation that made broad changes to the procedural treatment of the mentally disordered defendant. His particular aim within this legislation is to look at s9, an ostensibly straightforward provision that tests 'evidential sufficiency' prior to a determination of unfitness to stand trial. The 'evidential sufficiency hearing' was designed to determine whether the offender had 'caused' the acts or omissions constituting the *actus reu*s elements of the offence he or she was charged with, before they were at risk of a finding of unfitness to stand trial. The apparent statutory purpose of the s9 hearing was to divert mentally impaired offenders from the criminal justice system where the evidence was insufficient to sustain a criminal charge. Brookbanks contends that, in practice, its drafting has led to its becoming a highly contentious and intensely litigated provision. For Brookbanks, this points to a failure on the part of the legislation, which has not fulfilled expectations.

Paul Robinson provides a US perspective and examines the various ways in which an offender's mental illness can have an effect on liability and offence grading under American criminal law. This is not a straightforward task, as the 52 American jurisdictions have adopted a variety of different formulations of the insanity defence. A similar diversity of views is seen in the way in which different states deal with mental illness that negates an offence culpability requirement, a bare majority of which limit a defendant's ability to introduce mental illness for this purpose. Beyond this, Robinson also looks to the modern successor of the common law provocation defence, which now allows certain forms of mental illness to mitigate murder to a lesser form of murder or to manslaughter.

Stephen Morse also writes from a US perspective, and his concern is with the effect of neuroscience on normative questions that arise within the criminal law. Morse points to the basis of the criminal law's ideas of culpability in 'folk-psychological' concepts such as

desire, belief and intention, which means that 'the law treats persons generally as intentional creatures and not simply as mechanistic forces of nature'. Given that the criminal law is unlikely to accept a radical overhaul of its base concepts, Morse asserts that the adoption of science must be undertaken with great care, whether this comprises an externalising (the use of scientific or clinical experts to come to decisions) or internalising (deferring to scientific criteria as comprising legal criteria) approach. In light of this, he is sceptical of the utility and influence of neuroscience, and particularly where grand claims are made as to the possibilities it offers for providing answers to the normative questions that lie at the heart of the criminal law. Although he argues that 'neurolaw' presents limited possibilities and does not pose a genuinely radical challenge to the law's concepts of the person and responsibility, Morse goes on to make a case for 'cautious optimism about the contribution that neuroscience may make to law in the near and intermediate term', but emphasises that this must involve translation of the science into the folk–psychological framework and criteria of the criminal law.

The diversity of the articles presented here demonstrates the multiplicity of challenges that mental disorder, in its many varied forms, presents for the criminal justice system.

NILQ 65(2): 141–51

What is wrong about the 'criminal mind'?

THOM BROOKS[1]

Durham University

1 Introduction

Should we punish offenders because they have a 'criminal mind'? Retributivists argue for a strong link between a criminal's mindset at the time of an offence and our community's response through punishment. This view claims that punishment can be justified depending on the possession of a criminal mind which can be affected by factors that may affect culpability, such as mitigating factors. Retributivism is a powerful influence on our sentencing practices reflected in policy, such as the sentencing purposes listed in the Criminal Justice Act 2003.[2] These include purposes like 'the punishment' of offenders that are linked to the sentencing purposes originally found in the Model Penal Code, such as the purpose of differentiating offenders through 'a just individualization in their treatment'.[3]

It is argued that this view is based on a mistake about what makes the criminal mind relevant for punishment. It will be argued that a currently popular view of retribution endorsed by Feinberg and Duff – 'retributivist expressivism' – incorrectly links punishment to a criminal's possession of moral responsibility. This is a problem because its absence is no defence to strict liability offences, the largest subset of crimes. It is not a crime's threat or harm to morals that is most salient, but instead its threat or harm to our rights. We can make sense – and make better sense – of desert and sentencing practices through a rights-based approach. This is because the wrongs we are primarily concerned with in fact are not found in the criminal mindset, but rather the potential and actual infringement of rights.

This article focuses on the distinctively retributivist claim that morality forms the connection between crimes and harms: crimes are not any kind of harm, but specifically a kind of *immorality* demanding punishment. The criminal mind is, thus, an immoral mental state that deserves punishment as its justified response – one especially influential variety of retribution we might call 'retributivist expressivist' and developed by Joel Feinberg, Antony Duff and others. Retributivist expressivists argue that punishment has an *expressivist* function of communicating societal disapproval to criminal offenders for their moral

1 An earlier version of this essay was presented to the Oxford Jurisprudence Group. My thanks to the audience and, most especially, to John Gardner, Les Green and Fred Schaeur for their comments. This piece develops my critique of retributivist expressivism in new directions, expanding and clarifying my views in Thom Brooks, 'Criminal Harms' in Thom Brooks (ed), *Law and Legal Theory* (Brill 2013).

2 See Criminal Justice Act 2003, s 142.

3 American Law Institute, *Model Penal Code* (ALI 1962), s 1.02.

wrongdoings where punishment is proportionate to the immorality of their crimes. The following sections argue that it is not compelling both as a view about the criminal law, but also as a theory about punishment. The article concludes with a brief examination about how criminal harms might be better understood. It is argued that crimes might be a kind of harm, but not the kinds of harm endorsed by retributivists and, more specifically, expressivists. Sentencing should not aim at punishing the criminal mind, but the threat to our rights. Mindsets can matter, but not how retributivists think. The primary claim is that retributivist expressivism should be rejected and plausible alternatives may be available.

2 Retributivist expressivism

The Victorian judge James Fitzjames Stephen presents an early exposition of *retributivist expressivism*:

> The sentence of the law is to the moral sentiments of the public in relation to any offence what a seal is to hot wax. It converts into a permanent final judgment what might otherwise be a transient sentiment . . . *the infliction of punishment by law gives definite expression and solemn justification to the hatred which is excited by the commission of the offence.*[4]

Retributivist expressivism claims that punishment is the expression of public hatred for a criminal offence that is communicated to an offender. The public's voicing of its 'moral sentiments' confirm its anger: criminalisation and punishment is justified, in part, by the public's moral aversion to a crime. Crimes are not merely harms, but harmful to morals and so deserve punishment where an essential role is played by the expression of public anger to offenders through their punishment.

This approach brings together two different, but interrelated ideas. The first is retributivist desert, a contested concept.[5] The *standard view* of retributivist desert is presented above: punishment is justified where offenders deserve it because of their moral responsibility for some evil act or omission. Therefore, punishment should be set proportionate to their moral responsibility and the resulting evil. What justifies criminalisation of some act or omission is its immorality; likewise, what justifies punishment of some act or omission is its degree of immorality, too. Murderers and thieves should be punished because each performs evil actions, but the former should be punished more severely because the evil of murder is greater than the evil of theft. So the murder possesses *more* retributivist desert because of this combination and this can be an object of communication between the public and offenders.

The second idea is communicated expression between the public and the offender. Punishment should be communicated to offenders as an *expression* of public disapprobation for criminal wrongdoings. This idea has gained contemporary prominence through the work of Joel Feinberg, who argues that punishment has a 'symbolic significance' that the idea of punishment as an expression of public anger captures well.[6] Punishment expresses the community's moral condemnation of crime where crime is a harm *to morals*. These two ideas can be taken together: punishment is justified as the communication of retributivist expressivism. Punishment is an activity communicated by the state to offenders pertaining to their retributivist desert for moral wrongdoing that is best communicated through the expression of legal punishment.

4 James Fitzjames Stephen, *A History of the Criminal Law of England* (London 1883) 81–82 (emphasis added).

5 Thom Brooks, *Punishment* (Routledge 2012) 15–34.

6 Joel Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton University Press 1970) 98.

Some caveats are necessary. First, this justification of punishment understands 'punishment' in a specific way. Feinberg famously distinguishes 'punishment' from 'penalties'.[7] Punishment refers only to the use of prison: to speak of the justification of punishment is to address possible defences of hard treatment. Penalties encompass everything else, such as community sentences, monetary fines and verbal warnings. All retributivist expressivists are committed to the view that there is a difference in kind between *punishment as imprisonment* and *punishment as a response to crime*. The first understands punishment as only the use of prisons; the second views punishment as a set of responses to crime inclusive of penalties and hard treatment. Retributivist expressivists claim only the former (and not the latter) count as 'punishment'.

This perspective is problematic for two reasons. First, it is incorrect to say that *only* imprisonment can 'express' public condemnation to offenders. The discovery of a fine for illegal parking may be a penalty in Feinberg's terminology, but it is hard to see how receiving a fine cannot be understood by the person fined as some token of condemnation. The parking ticket is a symbol of public disapproval by imposing a financial burden that clearly does not communicate acceptance or public reward. So, perhaps a fine cannot express condemnation as loudly as a sentence, it may be more convincing to view penalties and imprisonment as falling along some scale of punishment.

Secondly, Feinberg's distinction falsely presents us with an either/or distinction. A penal outcome is *either* a penalty or punishment and not both. The problem is that this departs from the reality in many jurisdictions where convicted offenders who are imprisoned often receive some form of penalty as well. The point is that penalties and hard treatment are not either/or options for magistrates or judges, but part of a catalogue of potential penal outcomes that are available and regularly distributed together rather than one or the other.

A second caveat is that there is a curious relationship between the theories of retribution, expressivism and communicative theories of punishment exposing overlaps and contrasts. This claim is more than merely terminological. Each is understood in somewhat different ways although each also presents itself as a distinct alternative to rival theories of punishment despite possessing a shared view of desert. Retributivists do not always accept the view that punishment must require some further justification beyond its being deserved, such as an expression of public disapproval. Expressivists and communicative theorists accept this perspective, but it remains unclear how their views are distinct from retributivists. Expressivists claim punishment is an expression of public disapproval, but only for some disapprovals and not others – and then only in proportion to a specified range of behaviour where public disapproval is relevant.[8] In fact, expressivism accepts that only the deserving should be punished and to the degree deserved. Therefore, it is difficult to see what work 'expressivism' does for expressivist theories as such because offenders receive what is deserved and no less. To say that what is deserved is presented to offenders as an expression of public condemnation does not offer us anything new by emphasising any expressivist character. Retributivists and so-called 'expressivists' defend the same penal outcomes grounded in similar views about desert.[9]

The communicative theory of punishment, foremost championed by Antony Duff amongst others, claims that punishment must be 'communicated'.[10] Duff argues

7 Feinberg (n 6).

8 Brooks (n 5) 115–17.

9 Willie Charlton makes a compelling case that expressivists dress up retribution with 'undue glamour' in this way. See W E W St G Charlton, 'Rules and Punishments' [1966] Think 3.

10 See R A Duff, *Punishment, Communication and Community* (OUP 2001).

punishment should not only 'express' public disapproval from the public to offenders, but we should consider an offender's time served in prison as a 'communication' of penance expressed from offenders to the public. Nevertheless, communicative theories also limit the justification of punishment to the deserving and in proportion to what is deserved. While crucial to highlight the stated differences between retribution, expressivism and communicative theories, it is important to note their potential overlap. Indeed, there may be some redundancy in the idea of 'retributivist expressivism'.

This idea has gained in popularity for several reasons. The first is that criminal law has been perhaps the most natural law-friendly area of law for many theorists. Most traditional crimes are widely consistent with what any reasonable moral theory would disapprove, if not denounce. Crimes such as murder, theft or rape are perhaps the first to come to mind for most people. A natural overlap between the most serious moral wrongs and crimes might suggest some deeper connection between the criminal law and morality and so provide a more welcoming environment for natural law jurisprudence. Plus, this link between the morality and lawfulness is brought together in the idea of a criminal mind: the mental state of an offender is the key to unlocking our knowledge of his or her desert for a criminal wrong.

A second reason is the concern for providing a more robust justification for punishment within the context of a modern, liberal democracy. Retribution is about the enforcement of some moral perspective on all: we punish crimes as moral wrongs as determined by some viewpoint. But which one? The reassessment of retribution as retributivist expressivism opens up the possibility of a more compelling answer: wrongs are determined by 'us' and expressed through 'our' shared communication of public disapproval. In this way, retributivist expressivism can speak to some common agreement about morals found in our shared expression of public denunciation towards criminal wrongs that might avoid problems associated with the fact of reasonable pluralism and diverse moral views.[11]

The following two sections consider the idea of retributivist expressivism from a more critical perspective. First, it is argued that this idea provides a problematic and unpersuasive view about the criminal law. Secondly, it is argued that retributivist expressivism does not offer a compelling theory about punishment more generally and so we require a new model.

3 Retributivist expressivism and the criminal law

Does retributivist expressivism offer a view consistent with the criminal law? It may appear to be consistent at first glance, but this is a mistake. For example, most crimes we might think about may include murder, theft, rape or criminal damage. Any reasonable view of morality will at least disapprove, if not strongly denounce, them all. Different moral theories may disagree on the precise reasons while reaching the same conclusions. The deontologist might denounce murder as a failure to treat persons with the dignity they possess and respect owed. Utilitarians might denounce murder as a means to maximise happiness and minimise pain. Reasonable religious believers might claim murder is wrong because of widely shared beliefs about the value of life. Or not. These are only three general illustrations to show how different persons may agree murder is morally wrong and for different reasons. Many other examples could be given.

There is clearly something to be said for the claim that many criminal offences correspond toward much of what most of us would find morally problematic, especially in regard to more serious criminal offences. The question is then not whether there is an

11 See John Rawls, *Political Liberalism* paperback edn (Columbia UP 1996) 153.

overlap between crimes and immorality,[12] but rather a much deeper query about the causal link: are crimes harms to morals, such as the examples considered thus far?

This view has defenders among retributivist expressivists. For instance, Antony Duff argues that 'the criminal law aims to "enforce morality" in the sense that . . . it is inconsistent with the central moral values of the political community'.[13] Duff claims that the criminal law enforces public morality through punishment: 'The law "prohibits" murder, rape, and the like because such conduct is wrongful in a way that properly concerns the law – wrongful in terms of the shared values of the political community.'[14] We criminalise acts or omissions because they fail a moral standard that is found in the shared values we have and that find expression in our public disapproval.

But such a perspective has powerful critics. H L A Hart was surely correct to say: 'Has the development of the law been influenced by morals? The answer to this question plainly is "Yes".'[15] But is illegality linked with immorality? The answer to this more fundamental question is plainly 'no'. Much of the criminal law governs actions and omissions that need not be considered immoral by any reasonable moral view. When we consider crimes in general, our first thoughts may likely focus on so-called 'other-regarding' harms. These are crimes involving the infliction of some harm to someone, such as murder, theft and rape. Not all crimes have this character. Some are self-regarding, such as drug offences. Other crimes might lack victims or persons wronged (self or other), such as traffic offences.

Consider illegal parking. This is a traffic offence and part of the criminal law. It might not be the first type of offence to immediately spring to mind, but it is an offence that more of us have direct knowledge about: we can normally expect far more instances of illegal parking than murders for most, if not all, political communities. What does it mean to say this offence is *immoral* or even a *moral wrong*? Perhaps illegal parking through double parking prevents someone who has lawfully parked his or her vehicle from free movement. Or illegal parking on a narrow street might prevent normal access for traffic to travel. Both might be instances where the offender demonstrates a clear disregard for the respect for others to some degree.

But this need not be true in every case. Illegal parking is not defined by our distinguishing the sinful from the virtuous, but often in utilitarian terms. We ask: how can traffic move most freely through specific spaces? The fact that a one-way street restricts movement to a single direction need not be because this is morally good or desirable, but rather because the road might be narrow and restricting traffic to a single direction maximises our ability to travel around the vicinity most easily all things considered. So the law might enforce criminal law without any obvious connection with morality. Note that many country roads, such as in my adopted Britain, may be as narrow as any city street, but only the latter might not permit parking on either side and be restricted to one-way travel. That we park here or there and drive one direction or another might often be settled almost by chance and luck than morality and virtue. Note further that illegal parking might be morally justified or even morally required depending upon context, such as enabling a life-saving rescue. Illegal parking is not best explained with reference to its *immorality*, but rather its *practicality* as our rights may have more relevance than any moral wrongs. We will return to this point in the next section.

12 See Thom Brooks, 'Legal Positivism and Faith in Law' [2014] 77 MLR 139.

13 Duff (n 10) 67.

14 Ibid 58.

15 H L A Hart, *Law, Liberty, and Morality* (OUP 1963) 1.

Consider a very different crime like treason.[16] Every state punishes it with the most severe amount available to that state. Is treason immoral or morally wrong? The answer is clearly 'no'. Again, there may be cases where treason is morally justified or morally required, such as acts of treason against Nazi Germany or some similar evil state. The criminality of treason is not essentially about moral wrongdoing, but more about practical concerns.

The examples of illegal parking and treason are chosen to identify a spectrum of crimes that any state would include in its criminal law punished in a relatively small way in the case of illegal parking and commanding the most severe sanction in cases of treason. They are also crimes whose 'wrongness' is relatively independent to immorality. If retributivist expressivism cannot account for crimes like these, then it might have a significant problem as a theory of punishment. This is because theories of punishment are theories about practices. Perhaps it is to be expected that there will be some gap between our ideal view of punishment and practices found in any particular state.

The fact there is a gap is not the problem; the problem is the size of this gap. Retributivist expressivism runs into trouble not only with crimes on both ends of the scale, but many in between. This is perhaps especially true with so-called crimes involving self-regarding harms. These may include drug offences and (more controversially) prostitution. Few proponents of drug criminalisation argue for full legalisation of all currently banned drugs. Likewise, few proponents of legalising prostitution call for full deregulation. So how does the community express its disapproval for acts a person has performed that might cause no harm or inconvenience to the community? How much public anger will stir our collective moral sentiments? The answer is unclear at best.

A serious problem for retributivist expressivism is that it is a theory of punishment that rests on a particular foundation consistent with natural law – and at odds with the existing criminal law. It is a theory of punishment that can address only some, but not all, crimes we would want to include in the criminal law. None of this is to suggest that crimes are not harmful in any way. But the view of crimes as harms to morality stretches too far beyond the criminal law as we find it and even perhaps as we would want it to be. This point is important because it highlights the problem at the heart of retributivist expressivism: its claim that we should punish the possession of a criminal mind to the degree an offender has desert because an expression of public anger rests on a view that what is punishable is a wrongful criminal mind. But this defends an implausible position that crimes are best understood as kinds of harms to morals because this does not hold, or at least not among all crimes we would want a view about punishment to account for.

4 Against retributivist expressivism

Retributivist expressivism offers a poor match with the criminal law. But is it a compelling theory about punishment? The idea of expressive communication entails the public speaking with one voice. Offenders receive a message expressed by the public about how much it disapproves of their crimes. However, it is a mistake to argue that the public speaks with a single, unified voice in the way suggested. This point is defended well by Hart:

> It is sociologically very naive to think that there is even in England a single homogeneous social morality whose mouthpiece the judge can be in fixing sentence . . . Our society, whether we like it or not, is morally a plural society; and the judgements of the relative seriousness of different crimes vary within it far more than this simple theory recognises.[17]

16 See Treason Act 1351.

17 H L A Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Clarendon 1968) 171.

Modern society is characterised by the fact of reasonable pluralism.[18] No political community possesses one 'social morality' and not others, even if it may privilege one or some. Every community is pluralist and contains reasonable disagreement about moral and political values. One consequence is that there is no single voice from which the community might speak. This is because the political community contains more than one moral view. Perhaps there may be agreement – even an 'overlapping consensus' – that an offender should be punished to some degree, but the reasons for this decision may likely be several and perhaps conflicting.

A more nuanced problem with retributivist expressivism is the claim that the public can express its disapproval speaking as one voice and that this expression will communicate a particular message. Indeed, we may be unable to guard against communicating *unintended* meanings.[19] It is a mistake to claim that punishment expresses a single message to any messenger, but it may instead express multiple messages arising from the reasonable pluralism that exists in any modern political community. Moreover, we should not insist that punishment expresses only the message (or messages) we intend to express because there may be unintended messages communicated as well.

One possible response is offered by Duff. He argues that 'we should not hope to find any criterion, or neat set of criteria, of criminalization' that addresses this concern about what kinds of wrong are to serve as public wrongs deserving of punishment.[20] So perhaps there are many different messages communicated. Criminalization rests on difficult to unpick foundations and his communicative version of retributivist expressivism still remains the most compelling theory.

This is unsatisfactory. Supposing there might not be a single standard of 'immorality' derivable from the community's shared values, there remains (*a*) no argument or evidence of what values are shared by our community, (*b*) no satisfactory recognition that the values held by community members may be in conflict nor how such conflicts might be resolved and (*c*) no clear view about the problem of securing the communication of intended meanings while guarding against the expression of unintended meanings.

Retributivist expressivism has a much deeper problem: it is either redundant or incoherent. Retributivist expressivism claims punishment is to be proportionate to the amount of public denunciation appropriate. For example, Feinberg says: 'What justice demands is that the condemnatory aspect of the punishment suit the crime, that the crime be of a kind that is truly worthy of reprobation.'[21] Punishment is an expression of public disapproval.

However, not all public disapproval should be expressed as punishment. Only that which is 'truly worthy' – or, in other words, that which is *deserved* – can be the subject of punishment. Punishment is then proportionate to an offender's desert. The problem here is that the degree of appropriate public condemnation need not always be equal to the moral wrongfulness of a criminal offence. After all, we can 'exact retribution . . . without denunciation' and 'denounce a crime without exacting retribution'.[22]

Retributivist expressivists do not argue that, if public condemnation exceeded moral wrongfulness, punishment should be set more severely than deserved. Nor do they argue that, if moral wrongfulness exceeded public condemnation, punishment should be set much

18 Rawls (n 11) 153.

19 See Charles W Collier, 'Speech and Communication in Law and Philosophy' [2006] 12 Legal Theory 1, 8–9.

20 R A Duff, 'Answering for Crime' [2006] CVI Proceedings of the Aristotelian Society 85, 89.

21 Feinberg (n 6) 118.

22 John Cottingham, 'Varieties of Retribution' [1979] 29 Philosophical Quarterly 238, 238.

lower to bring it closer in line to public disapproval. Note that public disapproval does little, if any, work: punishment is justified when it is retributively deserved and punishment should be in proportion to what is deserved. Expressivism collapses into retributivism.

Now consider Duff's variant on retributivist expressivism. He argues that his communicative view helps to develop retributivism although it is also distinct from it.[23] Retributivists look backward only to the past crime whereas Duff's theory also looks forward to the future in order 'to persuade offenders that they should repent'.[24] Duff says:

> If he is convicted, his conviction communicates to him (and to others) the censure that he has been proved to deserve for his crime. He is expected (but not compelled) to understand and accept the censure as justified: to understand and accept that he committed a wrong for which the community now properly censures him. His trial and conviction thus address him and seek a response from him as a member of the political community who is both bound and protected by its laws.[25]

For Duff, imprisonment serves 'the communicative aims of punishment *more adequately* than . . . mere convictions or symbolic punishments; a communicative conception of punishment thus provides for its *complete* justification'.[26] Imprisonment serves communicative aims by providing 'an opportunity' for criminals 'to examine their souls', but not 'invade' them.[27] We are told 'punishment must go deep with the wrongdoer and must therefore occupy his attention, his thoughts, his emotions, for some considerable time'.[28] For Duff, the ability to 'go deep' is available only through imprisonment: demonstrating repentance is not enough.

The first problem with his particular theory of punishment is empirical. Imprisonment is said to 'more adequately' satisfy the communicative aims of bringing about a change of heart in offenders than the use of other sanctions. This is an empirical claim for which the only available evidence suggests that, in fact, imprisonment does not perform this task better than alternatives.[29] It is highly surprising to find Duff offering such a claim in light of the weight of evidence against this position.

Nevertheless, Duff is often at some pains to argue that the empirical foundations of his theory of punishment rest on how prison should be rather than how it is. He says:

> Such an objection would have force if my claim were that the familiar kinds of hard treatment punishment which are salient in our existing penal systems actually serve to induce repentance and self-reform, but that is not my claim . . . My claim is rather that suitably designed and administered kinds of hard treatment should, and in principle could, serve those aims . . . it does not depend on proof that our existing penal systems serve those aims.[30]

The problem with this position is not simply that imprisonment as currently practised fails to satisfy the communicative aims he sets for punishment (and it clearly fails these aims).

23 Duff (n 10) 25, 30.

24 Ibid 30.

25 Ibid 80. See also Thom Brooks, 'The Right to Trial by Jury' [2004] 21 Journal of Applied Philosophy 197 and Thom Brooks (ed), *The Right to a Fair Trial* (Ashgate 2009).

26 Duff (n 10) 82 (emphasis added).

27 Ibid 87 and see 122–23, 133.

28 Ibid 108.

29 Brooks (n 5) 123–48.

30 R A Duff, 'On Defence of One Type of Retribution: A Reply to Bagaric and Amarasekara' [2000] 24 Melbourne University Law Review 411, 420.

Instead, there seems no compelling reason to accept – given what we know – that imprisonment will 'always "more adequately"' serve communicative aims better than any alternative approach.

There is a further concern with Duff's theory. He says:

> But how can his punishment reconcile him to his victim or the wider community if it is obvious that he is unrepentant and unapologetic? . . . The offender has been subjected to what would constitute an appropriately reparative apology if he undertook it for himself. His fellow citizens should therefore now treat him as if he had apologized . . . He might not have paid the apologetic debt that he owed . . . But something like that debt has been exacted from him, and those who exacted it should now treat him *as if the debt has been paid*.[31]

What is striking about this passage is that the importance of repentance (and apology) drops out of the picture and appears to do little, if any, work. One concern is that simply serving a length of time in prison is transformed into some 'apologetic debt'. The problem is not simply that there is no evidence that placing offenders in prison is always more likely to inculcate a greater respect for the law than alternatives, but that ultimately it is of no concern whether offenders *do* repent and receive whatever moral communication we had sent, if any. Becoming repentant disappears: 'doing time' is 'repentance' by definition.

For Duff, communicative theories are different from others: expressivism is only about the expression of the public to offenders, but communication includes an expression from the offender to us. But, in fact, no such communication from offender to us may be expressed or perhaps even be possible. The 'communicative' theory of punishment may include no actual communication at all and yet its presence is at the heart of its claims about the justification of punishment.

5 Crimes as harms to rights

The idea of crimes as harms to morals does not cohere well with current criminal law nor how we might want it changed. Nor do the theories of punishment that endorse this idea offer us a compelling view about crime, morality or punishment. Crime is often considered to be a harm. This may have much to do with the attractiveness of the harm principle whereby an individual is free unless he or she might harm another. Many criminal offences are harms, such as murder or actual bodily harm. However, this perspective captures too much, for not all harms are or should be criminal. One example is prize-fighters who harm each other when boxing. Few (besides me) believe this should be criminalised. Perhaps all crimes are harms, but not all harms are crimes.

We have considered at some length one attempt to clarify more sharply the idea of crimes as harms in the formulation of crimes as harms *to morality*. We found that this is perhaps too narrow because it omits much of what we would want the criminal law to include. Many crimes are immoral (on various and competing views), but not all are so. Moreover, not all immorality is or should be criminalised: no one (including me) believes telling a white lie to keep a surprise birthday party a secret is wrong in every instance.

Another attempt is the idea of crimes as harms *to rights*. This is a position I defend at length elsewhere and summarise here.[32] This view understands that the criminal law is one part of a wider effort to protect and maintain our rights. Crimes are harms to our rights

31 Duff (n 10) 123–24.

32 Brooks (n 5) 123–48.

considered as substantial freedoms warranting protection and preservation.[33] Punishment is a response to crime that, as *a response*, may take multiple forms. But the justification of punishment is as a response to crimes in order to enable the protection and preservation of our rights. And, thus, crimes are a kind of rights violation that may require intervention.

All crimes are rights violations and some rights are more central than others. For example, some rights, such as a right against being murdered, are necessary to make possible other rights. Our more fundamental rights may warrant greater protection and, thus, more greater responses via criminal justice. This perspective is captured by the nineteenth-century British Idealist T H Green, who says: '[Punishment] is a disapproval founded on a sense of what is necessary for the protection of rights.'[34] The relation of rights and punishments is clarified by another nineteenth-century British Idealist named James Seth:

> This view of the object of punishment gives the true measure of its amount. This is found not in the amount of moral depravity which the crime reveals, but in the importance of the right violated, relatively to the system of rights of which it forms a part . . . The measure of the punishment is, in short, the measure of social necessity; and this necessity is a changing one.[35]

For Seth, we reflect on the central importance of the rights we have and protect through the criminal law. Some rights require greater protection than others as their violation may endanger our fundamental freedoms more than other crimes. We then punish murder more than theft because the former represents a greater threat to our rights than the latter.

Note that this new model can overcome problems associated with retributivist expressivism. Punishment might metaphorically 'express' public disapproval, but this has a definite shape – and it is not purely the stuff of moral philosophy. Instead, punishment is about the protection of rights. This can better capture a wider range of criminal offence categories, including strict liability where moral responsibility is lacking. Plus, this rights-based approach is more flexible: this is premised on an assumption that a pluralistic society may stand a better chance of finding agreement on its rights requiring protection than some shared morality to be expressed through retribution.

It is not claimed here that this alternative model is compelling, but rather that an alternative is possible to the dominant and problematic approach of retributivist expressivism. The point is clear: retributivist expressivists claim a strong link between the wrongness of the criminal mind and its punishment. This link should be rejected and another option is available grounded in a rights-based approach that has promise.

6 Conclusion

What is wrong about the criminal mind? Retributivist expressivists argue that the criminal mind is a state of mind that someone possesses at a particular time and which determines how much punishment is justified. Someone's retributivist desert – understood as a kind of moral responsibility for an immoral wrong – is all we require. This perspective has been rejected because it rests on an incorrect link between criminalisation and immorality. Moral responsibility may not feature in many, if not most, offences and it is unclear that it must play a role in other cases either. Instead, a rights-based alternative is available as one of

33 This understanding of rights is broadly consistent with a variety of different approaches to the philosophy of rights, including the capabilities approach. See Thom Brooks (ed), *Justice and the Capabilities Approach* (Ashgate 2012) and Martha C Nussbaum, *Women and Human Development: The Capabilities Approach* (CUP 2000).

34 T H Green, *Lectures on the Principles of Political Obligation* (Longman 1941) §197.

35 James Seth, *A Study of Ethical Principles* (William Blackwood & Sons 1898) 305. See H J W Hetherington and J H Muirhead (eds), *Social Purpose: A Contribution to a Philosophy of Civic Society* (George Allen & Unwin 1918) 129.

many possible options. This view suggests that what is wrong about the criminal mind is not its moral responsibility *per se*, but its connection to harms to our rights. The conclusion drawn is that we have compelling reasons to reject the influential approach of retributivist expressivism found in Feinberg, Duff and others and instead look elsewhere. This piece briefly sketched one such place that may prove fruitful. Better pastures can be found somewhere else.

NILQ 65(2): 153–66

The insanity defence in operation

R D Mackay

Professor of Criminal Policy and Mental Health, Leicester De Montfort Law School, De Montfort University

Little is known about how the M'Naghten Rules[1] operate in practice. This paper will attempt to redress this gap in knowledge by exploring available data on successful pleas of insanity over several decades. In doing so it will assess the value of such empirical studies and what they may bring to the reform agenda.

Other empirical studies

Although much has been written about the insanity defence and the M'Naghten Rules, very little of this work has considered how the rules operate in practice. In short, there have been very few empirical studies into the workings of the insanity defence in English law. Such studies as do exist have tended to focus on homicide with the result that other offences resulting in special verdicts have been excluded.[2] A notable exception, however, is the study by Professor Cheryl Thomas entitled 'Not Guilty by Reason of Insanity (NGRI) Verdicts (2006–2009)' commissioned by the Law Commission as part of its ongoing work into insanity and automatism.[3] This study covers the 28-month period of 1 October 2006 to 31 January 2009 and examines all Not Guilty by Reason of Insanity (NGRI) verdicts in relation to verdicts, defendants, cases and offences. In doing so it is stated that:

> This varied approach to analysing the data is important in order to present an accurate picture of Not Guilty by Reason of Insanity Verdicts. This is because it enables the analysis to take into account multiple charges and/or defendants, and therefore a single approach to analysing data can produce misleading results.[4]

Clearly, such a varied approach does add extra data and as such is a valuable addition to my own studies, which in turn I hope have not produced misleading results. In that connection, while it may be useful to have data on multiple verdicts, one must take care in how the overall figures are interpreted as it is stated that:

1 *R v M'Naghten* (1843) 10 Cl & F 200, 210; {1843–60] All ER Rep 229, 233.

2 See, for example, E Gibson and S Klein, *Murder 1957 to 1968: A Home Office Statistical Division Report on Murder in England and Wales* (The Stationery Office 1969); Nigel Walker, *Crime and Insanity in England* (Edinburgh University Press 1968); Matthew Large et al, 'Homicide Due to Mental Disorder in England and Wales over 50 years' (2008) 193 British Journal of Psychiatry 130.

3 See: Law Commission, *Insanity and Automatism: Supplementary Material to the Scoping Paper* (Law Com SP 2012) appendix B.

4 Ibid para B.2.

> In the 28-month time period covered by the CREST data, Not Guilty by Reason of Insanity verdicts (89) account for 0.3% of all Not Guilty jury verdicts reached by deliberation (33,865).[5]

While this may be so, it is important to note, as the study does in Table 1, that these 89 verdicts were the result of 40 cases in the sense that only 40 defendants were found NGRI during this research period. In essence, therefore, this study contains much of interest and is a welcome supplement to my own empirical studies as any additional data on the operation of the insanity defence are to be welcomed.

My older empirical studies

In my first study entitled 'Fact and Fiction about the Insanity Defence' published in 1990[6] and updated in 1995,[7] the legal position was that any successful insanity defence resulted in admission to hospital for an indefinite period of time with the result that the defence was rarely used. Despite this, the research revealed, firstly, that insanity was not confined to murder or attempted murder but was pleaded in respect of a wider variety of offences, particularly non-fatal offences; secondly, that the most common diagnosis used to support an insanity acquittal was that of schizophrenia, which in turn often led to a use of the 'wrongness' limb under the M'Naghten Rules; and, finally, that the majority of those acquitted on the grounds of insanity were being sent to local hospitals, where in some instances they were released within a mere matter of months. Taken together, these results suggested that the insanity defence was not quite as moribund as many had suggested. What followed was a policy paper from the then C3 Division of the Home Office favouring the introduction of flexibility of disposal for both insanity and unfitness to plead. This resulted in a Private Members' Bill which in turn resulted in the enactment of the Criminal Procedure (Insanity and Unfitness to Plead) Act 1991. The 1991 Act signalled the end of mandatory hospitalisation except where the charge was one of murder. It did so by introducing four additional disposals for both insanity and unfitness to plead, namely admission to hospital without restrictions, a guardianship order, a supervision and treatment order and an absolute discharge. By giving judges this new range of disposals it seemed likely that the use of the insanity defence would increase.

My second empirical study was designed to explore the first five years' operation of the 1991 Act from 1992 to 1996.[8] In doing so it found that the 1991 Act had resulted in an increase in the use of the insanity defence. Offences against the person continued to predominate but there had been a marked decrease in the number of cases of murder. The most common diagnosis used to support a defence of insanity continued to be schizophrenia. The 'wrongness limb' under the M'Naghten Rules continued to be more regularly used in psychiatric reports than the 'nature and quality' limb. Further, in the majority of cases the insanity defence was not disputed by the prosecution with the jury having no real deliberative role in the sense of being required to decide whether the accused was legally insane. The majority of those found NGRI were not sent to hospital (52.2%) but received community disposals, particularly supervision and treatment orders (47.7%). In short, judges were making full use of the disposal flexibility introduced by the 1991 Act. These findings were mirrored in my third empirical study which explored the second five-

5 Ibid para B.12.

6 R D Mackay, 'Fact and Fiction about the Insanity Defence' [1990] Criminal Law Review 247.

7 See R D Mackay, *Mental Condition Defences in the Criminal Law* (Oxford University Press 1995) 102–05.

8 R D Mackay and G Kearns, 'More Fact(s) about the Insanity Defence' [1999] Criminal Law Review 714.

year period of the operation of the 1991 Act from 1997 to 2001.[9] In particular, there was a continued but gradual increase in the use of the insanity defence, with a five-year total of 72 including a maximum of 17 special verdicts in 1999. Once again the majority of those found NGRI were not sent to hospital but received community disposals (52.8%, n=38), particularly supervision and treatment orders (41.7%, n=30), but with some increase in absolute discharges.

My most recent empirical study was commissioned by the Law Commission as part of its ongoing work in relation to the reform of the insanity defence. This study was published on 18 July 2012 in the Commission's 'Supplementary Material to its Scoping Paper on Insanity and Automatism'[10] and is summarised in my paper entitled 'Ten More Years of the Insanity Defence'.[11] Before I discuss this study I will make a few remarks about the Scoping Paper,[12] the primary purpose of which was 'to discover how in the criminal law of England and Wales the defences of insanity and automatism are working, if at all'.[13] In doing so the Commission emphasised the need for evidence about the use of insanity and automatism, describing existing data as 'very limited', and thus 'it is very difficult to make a meaningful assessment of the way the defences operate in practice'.[14]. In the hope of resolving this difficulty, the Commission invited responses to 76 questions which it posed. Question 9 asked for information about unsuccessful pleas of insanity together with evidence of how frequently insanity pleas are made. In its analysis of responses to this question the Commission stated that 'It is rarely pleaded unsuccessfully'[15] but could provide no data to support this conclusion. Question 10 asked for reasons why cases of successful insanity pleas might not have been recorded in the official data. The analysis of responses suggested, first, that some NGRIs appeared from Crown Prosecution Service data to be recorded as ordinary acquittals and, secondly, that in the view of the Criminal Bar Association fewer clerks allocated to courts made it less likely figures would be entered. While the latter is difficult to verify, with regard to the data I receive from the Ministry of Justice this suggests that, rather than being recorded as ordinary acquittals, there is a minimal number of such cases where the court had not recorded them using the correct verdict code; instead, they had entered it as an 'Other' verdict code.[16]

The Commission received only 20 written responses and one telephone response to its Scoping Paper.[17] In summarising the primary purpose of the paper as being 'to draw out evidence of how the defences work in practice', the Commission commented: 'Given that one of the difficulties we had identified was the lack of data, we were not surprised that little further data was provided. The nature of many of the responses was anecdotal.'[18] Despite this lack of data, the Commission drew from the responses the conclusion that insanity is 'little used . . . is rarely pleaded, and is pleaded only if it is likely to be a successful

9 R D Mackay, B J Mitchell and L Howe, 'Yet More Facts about the Insanity Defence' [2006] Criminal Law Review 399.

10 Law Commission (n 3) appendix E.

11 Ronnie Mackay, 'Ten More Years of the Insanity Defence' [2012] Criminal Law Review 946.

12 Law Commission, *Insanity and Automatism*, A Scoping Paper, 18 July 2012.

13 Ibid para 1.1.

14 Ibid para 1.23.

15 Law Commission, *Criminal Liability: Insanity and Automatism: A Discussion Paper* (Law Com DP 2013) appendix B, para B.22.

16 As a result, my updated study discussed below includes 7 additional NGRI cases giving a total for the 10-year period 2002–2011 of 230 cases compared to 223 cases in my Law Commission study.

17 Law Commission (n 15) para B.1.

18 Ibid para 1.80.

plea'.[19] This is certainly the view that I have long held, namely that my empirical studies of successful pleas of insanity capture the vast majority of such pleas. Of course, I cannot prove this but the failure of the Commission's Scoping Paper to produce data which suggests otherwise bolsters my view.

I now turn to the empirical study of the insanity defence that I conducted for the Commission which covered the ten-year period 2002 to 2011, which I have now updated to include the year 2012.

Recent research results

What follows is a study of verdicts (successful pleas) of NGRI during the 11-year period from 2002 to 2012 in order to assess the continued impact of flexibility of disposal together with the effect of the changes implemented by the Domestic Violence, Crime and Victims Act 2004. At the outset, however, the limitations of this current study need to be emphasised for, unlike my three earlier studies referred to above, on this occasion access to court files, and in particular relevant psychiatric reports, was unavailable. Despite this, however, the following research tries to give an up-to-date picture relating to insanity verdicts in England and Wales. Although the Statistics of Mentally Disordered Offenders continue to give the number of NGRI verdicts annually in relation to restricted patients,[20] no official statistics are published on the use of the insanity defence where other disposals are given. A final caveat, therefore, relates to the consistency of the data which were collected for this study using three statistical returns from the Ministry of Justice. Inevitably, although some disparity has been found in relation to these three sources, as complete a picture as seems possible of NGRI verdicts has emerged for the purpose of this research.[21]

The research findings

The number of NGRI findings

Table 1 gives the annual number of findings of NGRI for the last 5 years of the operation of the original 1964 Act, the first 5 years, the second 5 years and the third and fourth 5 years of the 1991 Act. The figures in brackets give the percentage increase in NGRI findings for each 5-year period of the 1991 Act. Although the picture is of a gradual but steady rise in the number of NGRI verdicts, it is noticeable that the overall percentage increase in NGRI findings has been in decline. Thus, in the fourth 5 years there was an annual average of 25.2 NGRI verdicts giving a 21.2% increase compared with an average of 20.8 (44.4% increase), 14.4 (63.6% increase) and 8.8 (120% increase) verdicts in the third, second and first 5-year periods respectively. This compares to an average of 4 from 1987–1991 (and 3.6 in the previous 5 years from 1982–1986, n=18) with an overall total for the first 20 years of the 1991 Act of 346, giving an annual average of 17.3 NGRI verdicts.

Table 2 gives the annual number of NGRI verdicts for the research period for this study, namely 2002 to 2012. The total of NGRI verdicts during this period was 260, giving an annual

19 Ibid 1.81.

20 See Ministry of Justice Offender Management Caseload Statistics 2012 annual tables at Table A6.5. It should also be noted that the Ministry of Justice figures are based on the date of the hospital warrants rather than the date of the finding. This may have led to minor inconsistency in relation to the actual number of annual findings. Thus, the total number of NGRI verdicts which resulted in hospital orders with restrictions recorded by the Ministry of Justice in the above table for the 11-year period 2002 to 2012 is 74 while the number contained in this study for the same 11 years is 73 (see Table 5 below).

21 I would like to acknowledge my gratitude to all the agencies and personnel involved for the generous assistance I received from them in carrying out this research.

1a 1964 Act Final 5years		1b 1991 Act 1st 5 years		1c 1991 Act 2nd 5 years		1c 1991 Act 3rd 5 years		1d 1991 Act 4th 5 years	
Year	Number	Year	Number	Year	Number	Year	Number	Year	Number
1987	2	1992	6	1997	10	2002	23	2007	13
1988	4	1993	5	1998	16	2003	17	2008	29
1989	3	1994	8	1999	17	2004	20	2009	27
1990	4	1995	12	2000	14	2005	20	2010	21
1991	7	1996	13	2001	15	2006	24	2011	36
Total	20	Total	44 (120%)	Total	72 (63.6%)	Total	104 (44.4%)	Total	126 (21.2%)

Table 1: Findings of NGRI by 5-year periods from 1987–2011

Year	Frequency	Per cent	Cumulative per cent
2002	23	8.8	8.8
2003	17	6.5	15.4
2004	20	7.7	23.1
2005	20	7.7	30.8
2006	24	9.2	40.0
2007	13	5.0	45.0
2008	29	11.2	56.2
2009	27	10.4	66.5
2010	21	8.1	74.6
2011	36	13.8	88.5
2012	30	11.5	100.0
Total	**260**	**100.0**	

Table 2: NGRI verdicts 2002–2012

Age range of accused	**Sex of accused**		**Total**
	male	female	
up to 15	1	0	1
15–19	10	0	10
20–29	78	6	84
30–39	74	13	87
40–49	35	9	44
50–59	24	2	26
60–69	6	1	7
70–79	1	0	1
Total	**229**	**31**	**260**

Table 3: Sex/age distribution

	Frequency	Per cent	Cumulative per cent
Murder	5	1.9	1.9
Attempted murder	45	17.3	19.2
Manslaughter	1	0.4	19.6
GBH	56	21.5	41.2
ABH	32	12.3	53.5
Arson	35	13.5	66.9
Criminal damage	7	2.7	69.6
Robbery	11	4.2	73.8
Burglary	8	3.1	76.9
Indecent/sexual assault	18	6.9	83.8
Threats to kill	2	0.8	84.6
Kidnap/child abduction	3	1.2	85.8
(Death by) dangerous driving	9	3.5	89.2
Possession/ importation/supply of drugs	1	0.4	89.6
Endangering aircraft	1	0.4	90.0
Breach restraining order	1	0.4	90.4
Affray	8	3.1	93.5
False imprisonment	2	0.8	94.2
Having article with blade	3	1.2	95.4
Theft	1	0.4	95.8
Racially aggravated assault	2	0.8	96.5
Bomb hoax	1	0.4	96.9
Child cruelty	1	0.4	97.3
Possession offensive weapon	2	0.8	98.1
Indecent exposure	2	0.8	98.8
Aid/abet reckless driving	1	0.4	99.2
Blackmail	1	0.4	99.6
Breach anti-social behaviour order	1	0.4	100.0
Total	**260**	**100.0**	

Table 4: Offences

average of 23.6. In essence, therefore, the annual average number of NGRI verdicts has now reached over 20 for the first time, with the totals for 2011 and 2012 now having exceeded 30.

Table 3 gives sex/age distribution of those found NGRI. It shows that the vast majority of those found NGRI continue to be males, at 88.1 per cent (n=229), compared to 11.9 per cent for females (n=31). The mean age at the time of the offence was 35.3 (range 15 to 74), with males having a mean age of 35, whilst females had a higher mean age of 37.4. The most prevalent age range for males is 20–29 (n=78) and for females 30–39 (n=13) with the vast majority of those found NGRI falling within the age ranges of 20–29 or 30–39 (n=171, 65.8%).

The offences charged

Table 4 gives the main offence charged that in each case led to a verdict of NGRI. It can be seen from this that there continues to be a wide spread of offences, the most prevalent of which are grievous bodily harm (GBH) (n=56, 21.5%) and attempted murder (n=45, 17.3 %). Once again, however, what is apparent is the very small number of murder charges: in the 1997–2001 study the number of such charges was 7 (9.7%), and this has now fallen to only 5 (1.9%).

As in previous studies, offences against the person (including robbery, kidnap/child abduction, false imprisonment and child cruelty) remain the most common type of offence with a total of 148 (56.9%) non-fatal and only 6 (2.3%) fatal offences. Overall, there has been an increase in GBH and actual bodily harm (ABH) combined from 27.8 per cent to 33.8 per cent when compared to the 5-year period 1997–2001 with a reduction in attempted murder from 22.2 per cent to 17.3 per cent.

The disposals

Previous studies of the insanity defence revealed that community-based disposals formed slightly over 50 per cent of all the disposals. In the 1997–2001 study, the figure was 52.8 per cent, although if the 7 mandatory disposals given in relation to the murder charges are ignored this total rises to 58.5 per cent. Table 5 gives the disposals for the current study. It

	Frequency	Per cent	Cumulative per cent
Restriction order without limit of time	73	28.1	28.1
Hospital order	50	19.2	47.3
Guardianship order	2	0.8	48.1
Supervision (and treatment) order – 2 years	71	27.3	75.4
Supervision (and treatment) order – under 2 years	20	7.7	83.1
Absolute discharge	43	16.5	99.6
Defendant discharged – hung jury	1	0.4	100.0
Total	**260**	**100.0**	

Table 5: Disposals

Main offence charged	Disposals							Total
	Restriction order without limit of time	Hospital order	Guardian-ship order	Supervision (and treatment) order – 2 years	Supervision (and treatment) order – under 2 years	Absolute discharge	Defendant discharged – hung jury	
Murder	4	0	0	1	0	0	0	5
Attempted murder	23	10	0	11	0	1	0	45
Manslaughter	1	0	0	0	0	0	0	1
GBH	20	17	0	10	3	5	1	56
ABH	7	3	0	9	3	10	0	32
Arson	8	10	0	9	5	3	0	35
Criminal damage	0	1	0	5	0	1	0	7
Robbery	1	0	0	5	0	5	0	11
Burglary	1	1	1	2	1	2	0	8
Indecent/ sexual assault	2	2	0	5	4	5	0	18
Threats to kill	0	0	0	2	0	0	0	2
Kidnap/child abduction	1	0	0	1	0	1	0	3
(Death by) dangerous driving	1	1	0	2	1	4	0	9
Possession/ importation/ supply of drugs	1	0	0	0	0	0	0	1
Endangering aircraft	1	0	0	0	0	0	0	1

Table 6: Main offence charged/disposals crosstabulation

Breach restraining order	0	0	0	1	0	0	0	1
Affray	2	3	1	1	1	0	0	8
False imprison-ment	0	0	0	2	0	0	0	2
Having article with blade	0	0	0	1	1	1	0	3
Theft	0	0	0	0	0	1	0	1
Racially aggravated assault	0	0	0	0	0	2	0	2
Bomb hoax	0	0	0	1	0	0	0	1
Child cruelty	0	0	0	1	0	0	0	1
Possession offensive weapon	0	0	0	0	1	1	0	2
Indecent exposure	0	2	0	0	0	0	0	2
Aid/abet reckless driving	0	0	0	1	0	0	0	1
Blackmail	0	0	0	1	0	0	0	1
Breach anti-social behaviour order	0	0	0	0	0	1	0	1
Total	73	50	2	71	20	43	1	260

Table 6 (continued): Main offence charged/disposals crosstabulation

Domestic Violence Act		Frequency	Per cent	Cumulative per cent
Pre-2004 Act	2002	23	33.3	33.3
	2003	17	24.6	58.0
	2004	20	29.0	87.0
	2005	9	13.0	100.0
	Total	**69**	**100.0**	
Post-2004 Act	2005	11	5.8	5.8
	2006	24	12.6	18.3
	2007	13	6.8	25.1
	2008	29	15.2	40.3
	2009	27	14.1	54.5
	2010	21	11.0	65.4
	2011	36	18.8	84.3
	2012	30	15.7	100.0
	Total	**191**	**100.0**	

Table 7: Year of decision

Domestic Violence Act		Frequency	Per cent	Cumulative per cent
Pre-2004 Act	Murder	2	2.9	2.9
	Attempted murder	16	23.2	26.1
	GBH	9	13.0	39.1
	ABH	5	7.2	46.4
	Arson	12	17.4	63.8
	Criminal damage	1	1.4	65.2
	Robbery	2	2.9	68.1
	Burglary	5	7.2	75.4
	Indecent/sexual assault	3	4.3	79.7
	Threats to kill	1	1.4	81.2
	Kidnap/child abduction	1	1.4	82.6
	(Death by)dangerous driving	2	2.9	85.5
	Possession/importation/supply of drugs	1	1.4	87.0
	Endangering aircraft	1	1.4	88.4
	Affray	2	2.9	91.3
	False imprisonment	1	1.4	92.8
	Having article with blade	1	1.4	94.2
	Theft	1	1.4	95.7
	Racially aggravated assault	1	1.4	97.1
	Bomb hoax	1	1.4	98.6
	Aid/abet reckless driving	1	1.4	100.0
	Total	**69**	**100.0**	

Table 8a: Main offence charged pre-2004 Act

Domestic Violence Act		Frequency	Per cent	Cumulative per cent
Post-2004 Act	Murder	3	1.6	1.6
	Attempted murder	29	15.2	16.8
	Manslaughter	1	.5	17.3
	GBH	47	24.6	41.9
	ABH	27	14.1	56.0
	Arson	23	12.0	68.1
	Criminal damage	6	3.1	71.2
	robbery	9	4.7	75.9
	Burglary	3	1.6	77.5
	Indecent/sexual assault	15	7.9	85.3
	Threats to kill	1	.5	85.9
	Kidnap/child abduction	2	1.0	86.9
	(Death by)dangerous driving	7	3.7	90.6
	Breach restraining order	1	.5	91.1
	Affray	6	3.1	94.2
	False imprisonment	1	.5	94.8
	Having article with blade	2	1.0	95.8
	Racially aggravated assault	1	.5	96.3
	Child cruelty	1	.5	96.9
	Possession offensive weapon	2	1.0	97.9
	Indecent exposure	2	1.0	99.0
	Blackmail	1	.5	99.5
	Breach anti-social behaviour order	1	.5	100.0
	Total	**191**	**100.0**	

Table 8b: Main offence charged post-2004 Act

can be seen from this that the number of hospital orders with and without restrictions was 123 (47.3%) which is similar to the total for the 1997–2001 study which was 47.2 per cent. However, the percentage of restriction orders has fallen from 37.5 per cent to 28.1 per cent with a marked increase in those without restrictions from 9.7 per cent to 19.2 per cent. It is also clear that community-based disposals continue to be well utilised, accounting for 52.3 per cent (n=136) of all disposals (ignoring the single case where the jury could not reach a verdict after insanity was pleaded and the defendant was discharged). Although the percentage of supervision (and treatment) orders has fallen from 41.7 per cent to 35 per cent (n=91), absolute discharges have risen from 9.7 per cent to 16.5 per cent (n=43). Overall, therefore, this figure of 52.3 per cent is similar to that in the 1997–2001 study of 52.8 per cent community disposals.

As in previous studies, Table 6 shows that community-based disposals continue to be given for serious offences including one case of murder (for the first time), attempted murder, including an absolute discharge (n=12), GBH (n=18), arson (n=17) and robbery (n=10).

Domestic Violence Act		Frequency	Per cent	Cumulative per cent
Pre-2004 Act	Restriction order without limit of time	26	37.7	37.7
	Hospital order	7	10.1	47.8
	Guardianship order	2	2.9	50.7
	Supervision (and treatment) order – 2 years	21	30.4	81.2
	Supervision (and treatment) order – under 2 years	6	8.7	89.9
	Absolute discharge	7	10.1	100.0
	Total	**69**	**100.0**	
Post-2004 Act	Restriction order without limit of time	47	24.6	24.6
	Hospital order	43	22.5	47.1
	Supervision (& treatment) order – 2 years	50	26.2	73.3
	Supervision (& treatment) order – under 2 years	14	7.3	80.6
	Absolute discharge	36	18.8	99.5
	Defendant discharged – hung jury	1	0.5	100.0
	Total	**191**	**100.0**	

Table 9: Disposal

The effect of the Domestic Violence, Crime and Victims Act 2004

The 2004 Act was implemented on March 31 2005. The Act reduced NGRI disposals to three, namely:

1. a hospital order (with or without a restriction order);[22]
2. a supervision order;
3. an order for an absolute discharge.

With regard to the present study which spans a period of 11 years, 39 (29.5%) months of the research period were prior to the implementation of the 2004 Act and 93 (70.5%) months post implementation.[23]

The following tables give a split of these two respective periods in order to show something of the impact of the 2004 disposal regime. Table 7 shows the numbers of NGRI cases involved pre- and post- the 2004 Act. It can be seen from this that 191 (73.5%) of the

22 The hospital order is now identical to one made under the Mental Health Act 1983 and, where the NGRI accused is charged with murder and the court has the power to make such an order, it must impose restrictions.

23 Only those defendants arraigned on or after 31 March 2005 are subject to the new disposal regime; see *R v Hussein* [2005] EWCA Crim 3556 [14]: 'The fact that the appellant was committed or sent to the Crown Court long before 31st March 2005 is nothing to the point.' Although this decision deals with a case of unfitness to plead, Schedule 12 para 8(2)(b) of the 2004 Act makes it clear that the same is true for the insanity defence.

NGRI cases fell to be dealt with under the 2004 Act, compared to 69 (26.5%) dealt with before the Act.

Tables 8a and 8b give a breakdown of the main offences charged in the periods before and after the enactment of the 2004 Act. It can be seen from this that the pattern of offences has remained fairly consistent. However, the percentage of cases of attempted murder has fallen in the post-2004 Act period by around one-third while cases of GBH have risen from 13 per cent to almost 25 per cent.

Table 9 gives the disposals for the two periods. What is of particular note is that, although the percentage of restriction orders has fallen, there has been an increase in the use of hospital orders from 10.1 per cent in the pre-2004 Act list to 22.5 per cent in the post-2004 Act list. Overall, however, the percentage of hospital-based disposals has fallen from 47.8 per cent under the pre-2004 Act period to 47.1 per cent under the post-2004 Act period, while the overall percentage of supervision (and treatment) orders has fallen from 39.1 per cent (42% if guardianship orders are included) to 33.5 per cent with a marked rise in absolute discharges by around 86 per cent.

Concluding remarks

As in my earlier studies, the number of verdicts of NGRI has continued to rise. The increase from a maximum of 17 findings in 1999 to a peak of 36 verdicts in 2011 certainly suggests that the legislative changes contained in the 1991 and 2004 Acts are having an ongoing effect. Overall, during the 11-year research period, hospital-based disposals have remained almost identical, 123 (47.3%) to the overall percentage for the 1997–2001 study which was 47.2 per cent. However, although community-based disposals accounted for 52.5 per cent (n=136) of NGRI cases, which is broadly similar to that in the 1997–2001 study of 52.8 per cent, absolute discharges have risen from 9.7 per cent (n=7) to 16.5 per cent (n=43).

With regard to the possible impact of the Domestic Violence, Crime and Victims Act 2004, the percentage of post-2004 Act non-hospital disposals has again remained virtually the same with a marginal increase from 52.1 per cent to 52.3 per cent post-2004 Act. Overall, therefore, the percentage of hospital-based disposals has fallen from 47.8 per cent under the pre-2004 Act period to 47.1 per cent under the post-2004 Act period.

Finally, although there has been this gradual increase in the number of NGRI verdicts, it remains the case, as the Law Commission emphasised in its Scoping Paper, that the number of such verdicts remains 'surprisingly low'.[24] In order to achieve a better understanding of how frequently the insanity defence is used, the Commission asked consultees to 'offer explanations as to why the number of special verdicts is so low'.[25] In its analysis of this response, the Commission suggested a number of possible reasons for the paucity of special verdicts, only one of which might impact on whether my empirical studies of successful pleas of insanity capture the vast majority of such pleas. Once again it is that there may be cases where: 'The defence is raised but unsuccessfully and the accused is convicted.'[26] As mentioned above, there may be such cases and I cannot prove otherwise. However, I am convinced that they are rare. Some tentative support for this conclusion is the fact that, in an earlier empirical study of 72 NGRI verdicts where we had access to psychiatric reports, it was found that out of a total of 161 such reports in only 6 of these reports was 'there clear evidence of contradictory opinion between the psychiatrists as to

24 Law Commission (n 12) para 1.48.

25 Ibid para 1.52.

26 Law Commission (n 15) para B.28.

whether a finding of NGRI could be supported'. From this we concluded that 'it seems likely that cases giving rise to this type of disagreement are rare',[27] in which case it would seem to follow that, if contested cases are rare, then the same may be true of failed cases. In any event, my empirical studies have never claimed to include failed cases and, as mentioned above, the Commission's analysis of responses to the questions posed in its Scoping Paper provided no real additional empirical data about the insanity defence. This supports the view that I have long held, namely that my empirical studies of successful pleas of insanity capture the vast majority of such pleas and continue to reveal valuable research data which in turn feed into the reform process.

Furthermore, this gradual increase in the number of NGRI verdicts needs to be tempered with the fact that, as Table 1 above reveals, the overall percentage increase in successive 5-year periods has slowed down and now stands at 21.2 per cent for the 5-year period 2007–2011. This may mean that any increase in NGRI verdicts is levelling out and that a further increase in numbers, if any, will be modest.[28] In short, until a new test for the insanity defence is implemented, it seems more than likely that the number of NGRI verdicts will remain very low. In that connection, the Law Commission's discussion paper, which provisionally proposes a new defence and special verdict,[29] is of real importance as it could at long last signal the end of the M'Naghten Rules together with the introduction of a defence which is more appropriate for the twenty-first century.

27 Mackay et al (n 9) 405.

28 I am continuing to monitor the number of NGRI verdicts and it will be interesting to discover whether my prediction is correct.

29 Law Commission (n 15) paras 1.86–1.96.

NILQ 65(2): 167–86

Automatism is never a defence

J J Child

Lecturer in Law and Co-chair of the Criminal Law, Criminal Justice and Criminology Stream, Centre for Responsibilities, Rights and the Law, Sussex Law School

and

Alan Reed

*Professor of Law and Associate Dean (Research and Innovation), Faculty of Business and Law, Northumbria University**

Introduction

In 2009, Andrew Simester's article 'Intoxication is Never a Defence' effectively highlighted a point he described as 'a simple one, and not entirely new'; that despite the longstanding (and enduring) description of the intoxication rules as a defence, this is not (and never has been) accurate.[1] The classification of intoxication as a defence is one of the criminal law's more peculiar self-delusions, not least because of the generally uncontroversial reasons for reaching the opposite conclusion. This is not a point of pedantry. As Simester explains, the appropriate classification of the intoxication rules as *inculpatory* leads us to evaluate those rules through a different set of norms and (as a logical conclusion) to question whether the law would be better served by a new voluntary intoxication-based offence.[2]

Following a similar pattern (albeit one that is likely, due to its relative novelty, to face greater resistance), it is our contention that the 'defence' of automatism is also incorrectly categorised.[3] A claim of automatism is a claim that D is not responsible for an (otherwise) criminal event because her acts were not voluntary. Therefore, automatism is never a defence; it is a description of an event that does not amount to an offence. Or, as we will see (in circumstances of prior fault), it provides a method of inculpation. Having explored and justified this interpretation, the article will discuss what this means for the development of the law when addressing longstanding debates surrounding automatism, such as the

* The authors thank Dr Tanya Palmer for comments on an earlier draft.

1 Andrew P Simester, 'Intoxication is Never a Defence' [2009] Criminal Law Review 3.

2 Ibid 14. A similar conclusion, based on similar reasoning, is reached by J C Smith and G Williams in the Criminal Law Revision Committee, *Offences Against the Person* (1980) 113–14. See also Rebecca Williams, 'Voluntary Intoxication—a Lost Cause?' (2013) 129 Law Quarterly Review 264; and John Child, 'Prior Fault: Blocking Defences or Constructing Crimes' in Alan Reed and Michael Bohlander (eds), *General Defences: Domestic and Comparative Perspectives* (Ashgate 2014).

3 See Law Commission, *Criminal Liability: Insanity and Automatism* Discussion Paper (July 2013) para 1.27: 'If a person totally lacked control of his or her body at the time of the offence, and that lack of control was not caused by his or her own prior fault, then he or she may plead not guilty and may be acquitted. This is referred to as the *defence* of automatism. It is a common law defence and it is available for all crimes.'; and see William Wilson, Irshaad Ibrahim, Peter Fenwick and Richard Marks, 'Violence, Sleepwalking and the Criminal Law (2) The Legal Aspects' [2005] Criminal Law Review 614, 618: '[A]utomatism floats relatively unchecked in the space between denials of capacity, denials of free choice and denials of bad character.' For a wider discussion of prior fault and the categorisation of offences and defences see Child (n 2).

necessary degree of involuntariness, as well as issues that emerge as a direct result, such as questions of fair labelling. As with the parallel analysis of intoxication, the logical conclusion of this debate is also the discussion of the potential for a new (prior) fault-based automatism offence.

Presentation of automatism in the current law

As with the intoxication rules, automatism is almost universally presented and discussed as a defence: defeating liability with the claim that D's acts were *in*voluntary. This is reflected in the presentation of automatism in textbooks, where the concept is often touched upon during early chapters on *actus reus* and *mens rea*, but then quickly referred to in a later and fuller discussion as a general defence. It is also an interpretation that appears consistently within the appellate courts, with the 'defence' only defeated by evidence of prior fault. In *Quick*,[4] Lawton LJ quotes with approval that:

> Automatism is a defence to a charge . . . provided that a person takes reasonable steps to prevent himself from acting involuntarily in a manner dangerous to the public. It must be caused by some factor which he could not reasonably foresee and not by a self-induced incapacity.

Similarly, Lord Justice Hughes has recently stated in *C*[5] that:

> . . . the defence of automatism is not available to a defendant who has induced an acute state of involuntary behaviour by his own fault.

As with similar statements relating to intoxication, these passages are not substantively wrong. They are misleading because they present the role of automatism in reverse: they present automatism as a defence capable of exculpating D from liability *unless* it is defeated by evidence of D culpably creating the conditions of her own defence.

The significance of automatism, as presently constructed, is to facilitate the individual actor with a means for raising a doubt as to whether she acted with the requisite culpability for the offence, and behaved voluntarily.[6] The common law, in broad terms, has made three classificatory distinctions related to perceived automatistic exculpation. First, the substantive law has demarcated automatism deriving from a disease of the mind (internal cause), and transmogrified such cases under insanity and mental defect provisions: mind referring herein to the 'ordinary sense of the mental faculties of reason, memory and understanding'.[7] A wide range of disposal powers has attached to the special verdict in this regard, viewed as essential for societal protection[8] and as a deterrent against recurrence of

4 [1973] QB 910.

5 [2013] EWCA Crim 223 [24].

6 See T H Jones, 'Insanity, Automatism and the Burden of Proof on the Accused' (1995) 111 Law Quarterly Review 475; and see generally, Stephen J Morse, 'Culpability and Control' (1994) 142 University of Pennsylvania Law Review 1587; William Wilson, 'Impaired Voluntariness: The Variable Standards' (2003) 6 Buffalo Law Review 1011; and R D Mackay and B J Mitchell, 'Sleepwalking, Automatism and Insanity' [2006] Criminal Law Review 901.

7 *Kemp* [1957] 1 QB 399, 407 (Devlin J): and see Patrick Healy, 'Automatism Confined' (2000) 45 McGill Law Journal 87 [22]: 'The crux of the approach is based on a double fiction: that automatistic involuntariness is presumptively internal in its origin, and that anything in the nature of an internal mental cause of automatism is presumptively mental disorder.'

8 *Sullivan* [1984] AC 156, 172 (Lord Diplock): 'The purpose of the legislation relating to the defence of insanity, ever since its origin in 1800, has been to protect society against recurrence of the dangerous conduct.'

violence.[9] Second, extant law has deontologically adduced involuntariness causally related to externally verifiable conditions, 'some external factor such as violence, drugs, including anaesthetics, alcohol and hypnotic influences'.[10] The prevalence of external factors, evidentially raised by the defendant and supported by medical evidence as to effect, may allow an absolute acquittal.

The binary divide created between internal/external causes has been problematic, and at times capricious,[11] in that it fails to make an appropriate classificatory division between physical and mental disorders, and both factorisations may operate simultaneously, for example, in relation to sleepwalking or hypnosis.[12] The defendant may be stereotyped in a particular taxonomy, despite acting in a similar involuntary manner, notably as diabetes is viewed as an internal factor whilst the administration of insulin is external.[13] Problems have also arisen over the correct ascription of disparate types of dissociative states.[14]

The classificatory system adopted has been vituperatively criticised as 'illogical',[15] 'little short of a disgrace'[16] and as 'making no sense',[17] and on occasions a judicial divining-rod has been needed for cause identification. A third ingredient is added to the mix in that culpability (prior fault) may operate to constitutively superimpose responsibility and deny exculpation, particularly evident in terms of driving offences and the intoxicated defendant. For example, in *C*,[18] a driving case, Moses LJ set out the orthodox view that D would 'have to provide an evidential basis for asserting that he could not reasonably have avoided the hypoglycaemic attack by advance testing'.[19] Prior fault principles, as stated, have been developed at common law for inculcated policy derivations to regulate the automatistic intoxicated 'offender'.

9 *Parks* [1992] SCR 871, 901 (La Forest J) in the Supreme Court of Canada: 'The continuing danger theory holds that any condition likely to present a recurring danger to the public should be treated as insanity . . . The two theories share a common concern for recurrence, the latter holding that an internal weakness is more likely to lead to recurrent violence than automatism brought on by some intervening external cause.'

10 *Quick* [1973] QB 910.

11 See *C* [2013] EWCA Crim 223 [20] (Hughes LJ): 'It is well known that the distinction drawn in *Quick* between external factors inducing a condition of the mind and internal factors which can properly be described as a disease can give rise to apparently strange results at the margin.'

12 Wilson et al (n 3) 617: 'The line drawn between sane and insane automatism can never make medical sense': It makes illogical, hair-splitting distinctions inevitable, allowing some 'an outright acquittal while condemning others to plead guilty or take the risk of a special verdict'.

13 See Andrew Ashworth and Jeremy Horder, *Principles of Criminal Law* 7th edn (Oxford University Press 2013) 94: 'There can be no sense in classifying hypoglycaemic states as automatism and hyperglycaemic states as insanity, when both states are so closely associated with a common condition as diabetes.'

14 See K Campbell, 'Psychological Blow Automatism: A Narrow Defence' (1980) 23 Criminal Law Quarterly 342; and B J Kormos, 'The Post-Traumatic Stress Defence in Canada: Reconnoitring the Old Lie' (2008) 54 Criminal Law Quarterly 189.

15 Law Commission (n 3) para 1.46 (Lord Justice Davis), referring to para 1.31 of the *Supplementary Material to the Scoping Paper*.

16 Ibid. See Ronnie Mackay and Markus Reuber, 'Epilepsy and the Defence of Insanity—Time for Change' [2007] Criminal Law Review 782, 791 stating this has led to the creation of 'a complex body of law which is manifestly unsatisfactory'.

17 Ashworth and Horder (n 13).

18 [2007] EWCA Crim 1862.

19 Ibid [35] and [38].

A similar presentation is also reflected explicitly in most common law jurisdictions.[20] The Supreme Court in Canada, for example, has consistently viewed automatism as a 'defence'.[21] *In Parks*[22] and *Stone*,[23] fundamental review of the parameters of this defence concluded that it is predicated on involuntariness constituting a complete lack of capacity to control one's conduct: unconsciousness, whether total or impaired is not supererogatory. Moreover, in *Stone* this defence has been deconstructed in reductionist terms: automatism is couched in a blanket of suspicion.[24] Trial judges, in light of the *Stone* decision, must weigh the adequacy of D's case to a balance of probabilities standardisation before the issue will even be put to normative fact-finders.[25] Policy-driven inculcations prevail in that it is believed that paternalistic considerations demand that it is necessary to protect the public from feigned claims of automatism. The judiciary should provide a bulwark against juries as moral arbiters who might be 'too quick to accept the story of an accused'.[26] Canadian courts have advocated a 'holistic' approach to dilemmatic choices presented and the adoption of a twin factorisation that embraces an internal cause test and the continuing danger test.[27] The former test invokes, as in English law, a bifurcation between internal and external causes of automatism, and in reality a dichotomous determination in cases where facts and circumstances typically reflect shades of grey, as in diabetes, sleepwalking and dissociative states. The latter test, as presaged by Lord Denning in *Bratty*,[28] determines that any condition of the defendant which is likely to recur and thereby present a danger to the public should be treated as a disease of the mind and subject to a wide range of disposal powers.

The presentation of automatism as a defence can also be seen, most recently, in the work of the Law Commission of England and Wales.[29] The recent Law Commission proposals, if adopted, would abrogate the schism that currently applies between internal and external causes of involuntary behaviour. A much broader template is suggested, creating a 'defence' predicated on a lack of capacity (total) arising from a recognised medical condition (RMC) embracing individuals with mental disorders and physical conditions, such as a person who suffers an epileptic seizure or who has a sleepwalking episode, or through a neurological defect such as Huntington's disease, but specifically excluding acute intoxication and where the condition was manifested solely or principally by abnormally aggressive or seriously irresponsible conduct.[30] The party seeking to raise the RMC defence must adduce evidence from at least two experts that at the time of the alleged offence they wholly lacked capacity: (i) rationally to form a judgment about the relevant conduct or circumstances; (ii) to understand the wrongfulness of what he or she is charged with having done; or (iii) to control his or her physical acts in relation to the relevant conduct or circumstances.[31]

20 See, for example, Scotland (*Ross v HM Advocate* (1991) JC 210); New Zealand (*Bannin* (1991) 2 NZLR 237); and Draft Criminal Code for England and Wales (1989) cl 33.

21 See Holly Phoenix, 'Automatism: A Fading Defence' (2010) 56 Criminal Law Quarterly 328; and Stanley Yeo, 'Clarifying Automatism' (2002) 25 International Journal of Law and Psychiatry 445.

22 [1992] 2 SCR 871, 75 CCC (3d) 287, 15 CR (4th) 289.

23 [1999] SCJ No 27, 134 CCC (3d) 353, 24 CR (5th) 1.

24 Ibid [180] (Bastarache J).

25 Ibid.

26 Ibid [29].

27 Phoenix (n 21) 352; and see Yeo (n 21) 449: 'Defendants pleading automatism are claiming that they are not criminally responsible for their conduct because they lacked the capacity to control such conduct.'

28 [1963] AC 386, 409.

29 Law Commission (n 3)

30 Ibid para 4.158–63.

31 Ibid para 4.160.

The broader gateway proposals for the new RMC defence are aligned with a more delimited role for automatism *per se*. The 'defence' of automatism would be available only where there is a total loss of capacity to control one's actions which is not caused by a recognised medical condition and for which the defendant was not culpably responsible.[32] An accused who successfully pleaded automatism would be simply acquitted. The Law Commission schema, consequently, restricts automatistic behaviour to automatic reflex reactions, or to transient states or circumstances, and only if an individual's condition persists and worsens it might then qualify as an RMC.[33] The difficulty, of course, as presented herein is the underlying premise of defence nomenclature, and the counterfactual assumption created thereby.

It is our view that this presentation of automatism does not conform to conventions relating to the division offences and defences; conventions (ironically) that have been consistently authorised by these same legal bodies.

Not a defence, even exceptionally

Offence elements are designed to target criminal wrongs; defining external (*actus reus*) and internal (*mens rea*) requirements in order to specify and isolate proscribed events. In contrast, criminal defences, strictly conceived, work in the opposite direction; defining certain circumstances where, despite committing the criminal offence, D's conduct should nevertheless be excused from liability.[34] For example, D may commit the offence of theft (satisfying both *actus reus* and *mens rea* elements) and yet be acquitted on the basis of a successful defence of duress: D is inculpated by her satisfaction of the offence elements, but then exculpated again by the defence. The distinction is a simple one, but it is also vitally important in order to make sense of the law in both substantive and moral terms.[35]

At the core of every criminal *offence* (including so-called strict or absolute liability offences) is the requirement that D's acts or omissions were performed voluntarily.[36] Criminal offences are generally constructed from a variety of external circumstances and results, but it is D's voluntary role within these elements that acts as a nexus of agency to hold them together: they become a single criminal event for which D may be held responsible. Thus, if D's conduct is involuntary (for example, D is unconscious or is being physically manipulated by X) then there is no nexus and D cannot have committed an offence. Automatism, as a denial of voluntary conduct, is therefore not a defence, it is a denial of this nexus and thus a denial of the offence itself. As Fletcher explains:

> Excuses arise in cases in which the actor's freedom of choice is constricted. His conduct is not strictly involuntary as if he suffered a seizure or if someone pushed his knife-holding hand down on the victim's throat. In these cases there is no act at all, no wrongdoing and therefore no need for an excuse.[37]

32 Law Commission (n 3) para 3.18; and see Andrew Ashworth, 'Insanity and Automatism: A Discussion Paper' [2013] Criminal Law Review 787.

33 Law Commission (n 3) para 5.110.

34 In the context of partial defences, their role is to block liability for murder (leading to liability for voluntary manslaughter instead).

35 For a discussion of this, see John Gardner, 'Fletcher on Offences and Defences', in *Offences and Defences: Selected Essays in the Philosophy of Criminal Law* (Oxford University Press 2007) 141, 144–46; Susan Dimock, '*Actio Libera in Causa*' (2013) 7 Criminal Law and Philosophy 549, 554; and William Wilson, 'The Structure of Criminal Defences' [2005] Criminal Law Review 108.

36 Ingrid Patient, 'Some Remarks about the Element of Voluntariness in Offences of Strict Liability' [1968] Criminal Law Review 23. The very rare exceptions to this rule have met with heavy criticism. See, for example, *Larsonneur* (1933) 24 Cr App R 74.

37 George Fletcher, *Rethinking Criminal Law* (Little Brown 2000) para. 10.3.2.

A similar point is made by Lord Denning in *Bratty*, although he goes on in this case to discuss automatism as a defence:

> No act is punishable if it is done involuntarily: and an involuntary act in this context—some people nowadays prefer to speak of it as 'automatism'—means an act which is done by the muscles without any control by the mind, such as a spasm, a reflex action or a convulsion. . . . [38]

Of course, talking loosely, even outside of automatism, it is possible to describe the denial of an offence in terms of a defence: this is common among barristers and particularly in civil law. Such descriptions are not consistent with the label 'defence' within the substantive law. For example, where D is brought to court on charges relating to burglary, she may claim that she was out of the country when the crime took place and therefore could not have been responsible. This is not a defence, it is an alibi, it is a denial that she completed the elements of the offence. The same is true with automatism.

It is worth noting here that there is some disagreement about which offence elements are denied when D claims to be acting in an automatic state. In Scotland, for example, involuntary conduct is described as a denial of *mens rea*: D may have acted in the sense of moving her body, but the movement was not internally willed.[39] This is also our preferred method of analysis.[40] Williams, in this regard, has identified that the capacity to act otherwise is constitutively the essence of voluntariness:[41] automatism is reviewed through a legal prism whereby it is an 'unnecessary refinement'[42] to view the doctrine as going beyond the denial of *mens rea*. Moreover, conduct is voluntary for the purposes of criminal responsibility, 'when the person could not have refrained from it if he had so willed; that is, he could have acted otherwise or kept still';[43] metaphorically, we should ask if D could have acted in a different fashion, if there had been 'a policeman at his shoulder'.[44] English courts[45] (and the Law Commission of England and Wales)[46] have generally described automatism as a denial of the *actus reus*: involuntary action not being considered as action at all. It may be that this uncertainty has encouraged use of the non-element specific terminology of automatism and perhaps thereby contributed to its presentation as a form of defence. However, this is mere speculation. What is important is that, whichever side of this debate one prefers, there remains the concession that the central role of automatism is

38 *Bratty* [1963] AC 386, 409.

39 *Ross v HM Advocate* (1991) JC 201, 213; and see, Pamela R Ferguson, 'The Limits of the Automatism Defence' (1991) 36 Journal of the Law Society of Scotland 446; Iain MacDougall, 'Automatism—Negation of *Mens Rea*' (1992) 37 Journal of the Law Society of Scotland 57; Pamela R Ferguson, 'Automatism—A Rejoinder' (1992) 37 Journal of the Law Society of Scotland 58; and Claire McDiarmid, 'How Do They Do That? Automatism, Coercion, Necessity and *Mens Rea* in Scots Criminal Law' in Reed and Bohlander (n 2).

40 As little turns on this debate for the current article, it will not be pursued in detail.

41 Glanville Williams, *Textbook of Criminal Law* 2nd edn (Stevens & Sons 1983) 148.

42 Ibid 663.

43 Ibid 148.

44 Ibid.

45 *Bratty* [1963] AC 386.

46 Law Commission (n 3) para 5.8. Some commentators have viewed automatism as a denial of either *actus reus* or *mens rea* simultaneously: see Emily Grant, 'While You Were Sleeping or Addicted: A Suggested Expansion of the Automatism Doctrine to Include an Addiction Defense' (2000) University of Illinois Law Review 997, 1002–03: 'Theoretically, the defense may be viewed from either standpoint, and thus it may be considered as relieving criminal liability either because the defendant lacks the mental state required for approval of a crime, or because the defendant has not engaged in an act—that is, involuntary bodily movement.'; and see Paul H Robinson, 'A Functional Analysis of *Mens Rea*' (1994) 88 Northwestern University Law Review 857, 896: 'Voluntariness might be thought to be more akin to *mens rea* than to *actus reus* elements.'

to deny something essential within the offence. Where all offence elements are satisfied, where we naturally move to consider defences, automatism has no role.[47]

In the US, the standpoint, in both the Model Penal Code (MPC) and across respective jurisdictions, has been that the demand that an act or omission be voluntary can be viewed as a preliminary requirement of culpability:[48] '[a] person is not guilty of an offense unless his liability is based on conduct which includes a voluntary act or the omission to perform an act of which he is physically capable'.[49] The MPC, although not specifically defining the term 'voluntary', provides instead four exemplars of acts that are not 'voluntary': (i) 'a reflex or convulsion'; (2) 'a bodily movement during unconsciousness or sleep'; (3) 'conduct during hypnosis or resulting from hypnotic suggestion'; and (iv) 'a bodily movement that is otherwise not a product of the effort or determination of the actor either conscious or habitual'.[50] The illustrations are detailed in the Commentaries as conduct that is not within the control of the actor,[51] but otherwise the template declines to offer a canonical formulation of the 'act' requirement, nor perform the determinative alchemy needed in terms of specificity for automatism vis á vis *mens rea* or *actus reus* elements of a crime.[52]

In exceptional circumstances, the automatism rules may have an alternative role within the law beyond a simple denial of offence elements; although, again, this is not as a defence. Rather, much like the intoxication rules, the automatism rules may function as the opposite of a defence, as a method of *in*culpation.[53] This arises where D's automatic state results from her own prior fault. In such cases, even though D does not satisfy the elements of the offence at the time it is committed (at T2), her earlier fault (at T1) substitutes for the missing elements at T2 to complete the offence as a form of constructive liability.[54] In the case of *Marison*,[55] for example, D was convicted of causing death by dangerous driving despite the fact that at the point of collision he was unconscious as a result of a hypoglycaemic episode. D had suffered such episodes before and so his prior fault in still deciding to drive (at T1) substituted for his lack of voluntariness when completing the other offence elements (at T2):

47 Gardner (n 35) 149.

48 See, generally, Deborah W Denno, 'A Mind to Blame: New Views on Involuntary Acts' (2003) 21 Behavioral Sciences and the Law 601; Kevin W Saunders, 'Voluntary Acts and the Criminal Law: Justifying Culpability Based on the Existence of Volition' (1998) 49 University of Pittsburgh Law Review 443; and Grant (n 46).

49 MPC, s 2.01(1). The philosophical theory behind the rule is explained further in the Commentaries in terms of free will and volition: 'That penal sanctions cannot be employed with justice unless these requirements are satisfied seems wholly clear. It is fundamental that a civilized society does not punish for thoughts alone. Beyond this, the law cannot hope to deter involuntary movement or to stimulate action that cannot physically be performed; the sense of personal security would be undermined in a society where such movement or inactivity could lead to formal social condemnation of the sort that a conviction necessarily entails. People whose involuntary movements threaten harm to others may present a public health or safety problem, calling for therapy or even custodial commitment; they do not present a problem of correction.'; see MPC and Commentaries, s. 2.01 cmt, at 214–15 (1985).

50 MPC and Commentaries s 2.01 (1985) 215; and see, generally, Grant (n 46).

51 MPC and Commentaries s 2.01 (1985) 215.

52 See, generally, Douglas Husak, 'Rethinking the Act Requirement' (2007) 28 Cardozo Law Journal 2437; and see Joshua Dressler, *Understanding Criminal Law* 4th edn (Lexis Nexis 2006) 101, formulating the act requirement in the following terms: 'A person is not guilty of an offense unless her conduct, which must include a voluntary act, and which must be accompanied by a culpable state of mind (the *mens rea* of the offense) is the actual and proximate cause of the social harm, as proscribed by the offense.'

53 Ronnie Mackay, 'Intoxication as a Factor in Automatism' [1982] Criminal Law Review 146, 146–48.

54 This should be distinguished from so-called grand-schemer cases, where D loses voluntary control in *order* to commit the offence. In such cases, it is contended, liability can be found simply through the appropriate use of the rules of causation: see Child (n 2).

55 [1997] RTR 457.

> It was argued before this court . . . that [D] was driving as an automaton and therefore cannot be guilty of the offence. In our judgment, automatism does not come into this case at all . . . Even if the appellant was in an automatic state for the last few seconds, he had already committed the offence by driving to that point, in circumstances which he knew were such that he might have a hypoglycaemic attack at any moment.[56]

In line with the judicial comments quoted in the first part of this article, the court in *Marison* presents the facts (D's prior fault) as blocking the defence of automatism: D cannot make use of the defence because he was at fault for creating the conditions that led to it.[57] As previously stated, this is to present the law in reverse. The correct analysis is that D did not complete the elements of the offence at T2 when death was caused; D was not acting voluntarily and so an essential element of his offence was missing. D's prior fault (choosing to drive and knowing of the possibility of losing consciousness in this manner) was used to substitute for that missing element in order for the offence to be completed. Referring to the 'defence of automatism', the court was right to say that it was irrelevant to this case. The rules of automatism as a constructer of liability, however, through the rules governing prior fault, played a crucial role.

The automatism rules, then, can operate in two ways. First, automatism can be a simple explanation (a shorthand) for D who does not commit an offence because her conduct is not voluntary. Secondly, where D lacks voluntary conduct as a result of prior fault, the automatism rules can be used to substitute for that lack of voluntariness to find liability. Automatism is never, even exceptionally, a defence.

Problems with the automatism rules

Having set out our central contention, that the automatism rules are inculpatory as opposed to exculpatory in function, it is useful to question what effect this might have on the application of those rules. To do so, we will explore two areas of debate that have been central to the automatism rules for some time, and then two further areas of debate that arise as a result of our analysis in this article. In this manner, we hope to demonstrate how the classification of automatism as a constructer of liability has important implications for the substance of those rules.

The first longstanding area of debate, relevant to all cases of potential involuntariness, concerns the threshold of capacity required for D's acts or omissions to be considered voluntary.[58] Discussed in the context of a 'defence' of automatism, the question is whether automatism requires D to lack all physical control of her conduct (for example, through unconsciousness or physical spasm), or whether it is sufficient that she lacks effective or rational control (for example, through dissociation short of full unconsciousness).[59] The debate has been a problematic one: lacking a medical consensus for the law to take reference

56 *Marison* [1997] RTR 457, 461 (McCowan LJ).

57 See John Rumbold and Martin Wasik, 'Diabetic Drivers, Hypoglycaemic Unawareness and Automatism' [2011] Criminal Law Review 863, 866: 'Usually the diabetic driver has been at fault in the management of their condition, and so any defence of automatism fails . . . [T]he condition of hypoglycaemic unawareness is highly relevant to this issue of fault, and is a factor to which lawyers involved in such cases should be alert.'

58 See A P Simester, J R Spencer, G R Sullivan and G J Virgo, *Criminal Law: Theory and Doctrine* 5th edn (Hart 2013) 112, delineating road traffic cases from others: 'It is noteworthy that the cases in which the most stringent demands are made all concern driving offences.'; and see further on the effective control requirement, Douglas Husak, 'The Alleged Act Requirement in Criminal Law' in *The Oxford Handbook of Philosophy of Criminal Law* (Oxford University Press 2011) 107.

59 A useful overview is provided in Law Commission (n 3) para. 5.22–32.

from[60] and prone to policy-based inconsistencies.[61] However, what is most interesting for present purposes, has been the presentation of this uncertainty within the courts; particularly in the last few years where they have shown a consistent preference for the narrower interpretation of automatism. For example, in the case of *C*,[62] Lord Justice Hughes comments:

> Automatism, if it occurs, results in a complete acquittal on the grounds that the act was not that of the defendant at all . . . *'Involuntary' is not the same as 'irrational'; indeed it needs sharply to be distinguished from it.*[63]

In contrast, commentators who favour a wider view of automatism have tended to focus on the related issue of moral responsibility. For example, Horder contends:

> The all-embracing explanatory claims of the voluntary conduct model . . . [is] where one finds an assumption about non-insane automatism that all that matters is physical capacity to engage in voluntary conduct, and that the question of whether one has control over conduct is the same thing as whether one is engaging in voluntary conduct at all.[64]

Whether we classify automatism as a defence does not determine the outcome of this debate, but it can play an important role. This is because, if automatism is (accurately) presented as a simple shorthand for an incomplete offence, then we are forced to consider what is required in order to form a complete offence. This focuses on questions of sufficient moral responsibility and the required nexus of agency between D's conduct and surrounding offence elements: exactly the focus that leads Horder and others to advocate a narrower view of voluntariness (i.e. a broader 'defence' of automatism). In contrast, the dominant view of automatism as a defence encourages the courts to begin from the opposing premise; asking whether D's lack of control was sufficient to excuse her from liability for an existing criminal wrong. This approach encourages the courts, as we have seen, to think in terms of maintaining sensible limits on a *defence* that can lead to a complete acquittal: a focus that inevitably leads one to a narrower conception of the 'defence' (i.e. a broader notion of voluntariness).[65] Demonstrating that automatism is never a defence, we hope that this debate can be set on the appropriate foundations: questioning whether impaired or dissociative mental control should be considered sufficient to construct and tie together criminal wrongs.

The second longstanding debate affected by our classification discussion relates specifically to prior fault[66] and the inculpatory role of the automatism rules: questioning

60 McLeod et al, 'Automatism and Dissociation: Disturbances of Consciousness and Volition from a Psychological Perspective' (2004) International Journal of Law and Psychiatry 471; D Bell, 'Judgements Revisited: Falconer' (2011) Aus JFS 313; and Irshaad Ibrahim et al, 'Violence, Sleepwalking and the Criminal Law: Part 1: The Medical Aspects' [2005] Criminal Law Review 601.

61 For example, the developments in case law that seem to apply different standards to different categories of cases, particularly driving cases and post-traumatic stress disorders: and see, Kormos (n 14)

62 [2013] EWCA Crim 223.

63 Ibid [22] (emphasis added).

64 Jeremy Horder, 'Pleading Involuntary Lack of Capacity' (1993) Cambridge Law Journal 298, 312; and see G R Sullivan, 'Making Excuses' in A P Simester and A T H Smith (eds), *Harm and Culpability* (Oxford University Press 1996) 131.

65 A similar dynamic can be seen in the discussion of consent in the House of Lords case of *Brown* [1994] 1 AC 212, where the majority advocated for a narrow 'defence' of consent and the minority for a narrow 'offence element' of non-consent.

66 See Ashworth and Horder (n 13) 93: 'The aim of the doctrine of prior fault is to prevent D taking advantage of a condition if it arose through D's own fault.' Automatism is conceived as a denial of 'authorship' on the part of D (ibid).

whether the notion of prior fault requires subjective or objective foresight, and foresight of what? Case law engaging these questions has moved inconsistently between various options, with cases such as *Quick*[67] suggesting that the automatism 'defence' would be defeated where D 'could have *reasonably* foreseen [the criminal harms] as a result of either doing, or omitting to do, something',[68] whereas others, such as *Bailey*,[69] have suggested that it would only be defeated where D subjectively foresaw the possibility of future harms.[70] As above, we do not believe that the classification of automatism as a defence is determinative of this debate, but again, it must have a role. This is because, if our argument is accepted that the automatism rules (as they relate to prior fault) are inculpatory in function, that they are replacing missing elements of an offence, then the debate must hinge on what construction of prior fault is required for culpability equivalence with the offence elements they are seeking to replace (the lack of voluntariness at T2). With this in mind, it becomes very difficult to maintain that negligently[71] failing to foresee the potential for future dangerousness or simple future involuntariness (i.e. objective prior fault at T1) is equivalent to voluntary movement and awareness of circumstances at T2. In fact, it is even difficult to accept an equivalence between voluntariness and awareness at T2 with some manner of subjective foresight of involuntariness or a possible future risk (i.e. subjective foresight at T1), but this will be discussed further below. Again, we have a longstanding debate of vital importance, but one that is currently being conducted on faulty terms.

The debate has been enervated in recent times by a number of US and Canadian commentators, who have suggested that a recategorisation of *actio libera in causa* principles should broadly apply to formulate an appropriate prior fault and intoxication doctrine.[72] By parity of reason, a similar reformulation is propounded within the purview of automatism. The *actio libera* doctrine, as previously constructed, operates to disallow D relying on exculpation (defence) at T2, the conditions for which she has culpably created at T1.[73] As stated, when properly deconstructed, a reverse nexus may apply, not in terms of defence nomenclature, but rather as a predicate of liability for the morally culpable automatistic individual. Dimock categorises the principle, however, in another distinctive hue, as a derivative of 'imputation' not of inculpation: '[I]f . . . we think such conduct can, despite being voluntary, reveal the relevant attitudes of the agent, it seems we must be looking to the prior conduct of the agent in creating the conditions of involuntariness to make the connection.'[74] The practical reality, viewed either through a kaleidoscope of imputation or inculpation, is that the criminal responsibility and fault of the actor at T1 must be traced through in a causal sense to harm commission at T2: a requirement Robinson has stated of

67 [1973] QB 910.

68 Ibid (Lawton LJ) (emphasis added).

69 [1983] 2 All ER 503.

70 See Rumbold and Wasik (n 57); and Law Commission (n 3) para 6.12–28.

71 See, by way of comparison, the definition of negligence in MPC s. 2.02 (4)(d): 'A person acts negligently with respect to a material element of an offense when he should be aware of a substantial and unjustifiable risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that the actor's failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, involves a gross deviation from the standard of care that a reasonable person would observe in the model's situation.'

72 Dimock (n 35); Leo Katz, 'Entrapment through the Lens, of the *Actio Libera in Causa*' (2013) 7 Criminal Law and Philosophy 587; Larry Alexander, 'Causing the Conditions of One's Defence: A Theoretical Non-Problem' (2013) 7 Criminal Law and Philosophy 623; and Douglas Husak, 'Intoxication and Culpability' (2012) 6 Criminal Law and Philosophy 363.

73 Dimock (n 35) 551.

74 Ibid 560.

'a strong causal connection with the imputed objective element, culpability as to the causal connection itself, and the culpability required by the substantive offence'.[75]

In constitutive effect, prior fault automatistic 'conduct' engages a conflagration of criminal responsibility, blameworthiness and tracing principles, aligned together to focus potentiate culpability at the temporal individuation point where D may questionably have had 'guidance-control' over voluntary action, as Fischer and Ravizza have cogently articulated:

> When one acts from reasons—responsiveness mechanism at T1, and one can reasonably be expected to know that so acting will (or may) lead to acting from an unresponsive mechanism at some later time T2, one can be held responsible for so acting at T2.[76]

The responsibility-tracing-fault nexus, as propounded for prior fault automatistic individuation, was vividly exemplified in a straightforward categorisation by the High Court of Australia in *Ryan v R*.[77] D, in the course of a robbery, had threatened a service station cashier with a sawn-off rifle; the rifle was loaded and the safety-catch had been deliberately removed at T1 time-frame. Ryan attempted to tie up the cashier with one hand while pointing the rifle at him with the other. Unfortunately, the cashier made a sudden movement and D shot him dead. D asserted that, 'startled' by V at whom his gun was pointed, his finger depressed the trigger as a truly involuntary 'reflex action'. The majority in the High Court of Australia, in contradistinction, adopted the perspective that Ryan had voluntarily (culpably at T1) placed himself in a situation where he might need to make a split-second decision and the fact that he so responded by pulling the trigger did not make that act an involuntary act in the nature of an act done in a convulsion or epileptic seizure. Chief Justice Barwick, in the minority, but legitimately on the facts, determined that D's account of the events engaged in pulling the trigger, if true, did embody a reflex action in the sense of being unwilled: Ryan's squeezing of the trigger was more akin to an act done in a convulsion or epileptic seizure than it is to that of a tennis player retrieving a difficult shot where the action is a willed muscular movement albeit that the decision to make it is made in a split second.[78] Literal involuntariness may standardise Ryan's pressing of the trigger, but prior fault applied in releasing the safety-catch of the weapon and, similarly in *Commonwealth v Fain*,[79] where D, a sleep-pattern disordered individual (that made him violent when aroused from sleep) was criminally responsible at T1 for going to sleep in a public room of a hotel with a deadly weapon on his person. Inculpation is derived from prior fault at T1 temporal individuation for which the individual is criminally responsible and not automatistic '*in*voluntariness' at T2. Pithy realism should apply to our consideration of prior fault and the *actio libera* doctrine attached to automatism as well as intoxication, arguably standardising the criminalisation of behaviour in this sphere derivatively from harmful moral agency.[80]

---

75 Paul H Robinson, 'Imputed Criminal Liability' (1984) 93 Yale Law Journal 609, 638–39.

76 John Martin Fischer and Mark Ravizza, *Responsibility and Control* (Cambridge University Press 1988) 50.

77 (1967) 121 CLR 205, High Court of Australia.

78 David Ormerod, *Smith and Hogan: Criminal Law* 12th edn (Oxford University Press 2008) 56, referring to I D Elliott, 'Responsibility for Involuntary Acts: *Ryan v the Queen*' (1968) 41 Australian Law Journal 497.

79 79 Ky 183 (1879); and see further Paul H Robinson, 'Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine' (1985) 71 Virginia Law Review 1, 33, asserting that: '[B]ecause he no doubt [knew] his propensity to do acts of violence when aroused from sleep, he could probably have been held liable for reckless homicide based on his earlier conduct of going to sleep in a public place with a gun in his lap.'

80 See, generally, Alan Reed and Nicola Wake, 'Potentiate Liability and Prevening Fault Attribution: The Intoxicated "Offender" and Anglo-American Dépecage Standardisations' (2014) John Marshall Law Review, Chicago.

Automatism and intoxication: the construction of liability

The next two areas of debate, although not entirely foreign to the discussion of automatism, are more commonly associated with the intoxication rules. However, once it is accepted that, like the intoxication rules, automatism is never a defence (and may act as a constructer of liability), then they become central to this area as well. These debates both relate to the role of prior fault and the construction of liability, first, to the specific/basic intent distinction and, secondly, to the appropriate labelling of offenders.

The distinction between basic and specific intent offences is crucial to the operation of the *intoxication* rules. The distinction relates to the *mens rea* required as to any circumstance or result elements within the offence charged. As well as substituting for a lack of voluntary conduct at T2,[81] D's intoxication will substitute for a lack of *mens rea* as to these elements where the offence is one of basic intent (constructing liability), but not for offences of specific intent (failing to construct liability).[82] For example, if D attacks V causing grievous bodily harm (GBH), but lacks all *mens rea* and even acts involuntarily as a result of voluntary intoxication, she cannot be liable for an offence of causing GBH with intent[83] (specific intent offence), but will be liable for a recklessness-based GBH offence (basic intent offence).[84] Following our interpretation of prior fault automatism in line with the intoxication rules, the question now is whether the same distinction applies to automatism? It is clear that prior fault automatism is capable of substituting for more than solely a lack of voluntariness because otherwise it could only operate to construct liability for strict liability offences: where D lacks control of her body, she is very unlikely to be acting with any subjective *mens rea* as to associated circumstances or results.[85] But in what manner (if at all) is the potential for constructing liability in this context restricted?

The early case law on prior fault and automatism did not recognise a basic/specific intent distinction; implying a very broad potential for the substitution of missing *mens rea* elements. In *Quick*,[86] for example, Lawton LJ states:

> A self-induced incapacity will not excuse . . . nor will one which could have been reasonably foreseen as a result of either doing, or omitting to do something, as, for example, taking alcohol against medical advice after using certain prescribed drugs, or failing to have regular meals while taking insulin.[87]

Although the offence in *Quick* was one of basic intent, the broad language of this statement (potentially embracing both basic and specific intent offences) led Mackay to highlight the potential for an indefensible inconsistency between the intoxication rules and automatism.

> . . . a defendant like *Quick*, had he been prosecuted for a crime of 'specific intent,' would have been convicted of that offence had he been found not to have followed his doctor's instructions, whereas his intoxicated counterpart would only have been convicted of a crime of 'basic intent.'[88]

81 *Lipman* [1970] 1 QB.

82 *DPP v Majewski* [1977] AC 443.

83 Non-Fatal Offences Against the Person Act 1861, s. 18.

84 Ibid s. 20

85 It should be remembered, however, that unlike the intoxication rules, automatism will only come into play where there has *also* been a lack of voluntary movement: where D's non-intoxicated prior fault results in a lack of *mens rea*, but not a lack of voluntary movement, the automatism rules will play no part. This is discussed further below.

86 [1972] QB 910.

87 Ibid 922.

88 Mackay (n 53) 155.

The potential for this inconsistency, and the role for a basic/specific intent distinction in automatism cases, finally arose in the case of *Bailey*:[89] a case involving the specific intent offence of wounding or causing GBH with intent. However, this case does more to confuse the law than to clarify it. The court highlighted (in line with Mackay) that *Quick* should *not* be interpreted to allow non-intoxicated prior fault automatism to substitute for missing *mens rea* elements in crimes of specific intent.[90] The court then goes further to cast doubt on its ability to substitute for similar elements in crimes of *basic* intent as well:

> In our judgment, self-induced automatism, other than that due to intoxication from alcohol or drugs, may provide a defence to crimes of basic intent. The question in each case will be whether the prosecution have proved the necessary element of recklessness. In cases of assault, if the accused knows that his actions or inaction are likely to make him aggressive, unpredictable or uncontrolled with the result that he may cause some injury to others and he persists in the action or takes no remedial action when he knows it is required, it will be open to the jury to find that he was reckless.[91]

In this statement, Griffiths LJ is essentially undermining any role that prior fault could play in the construction of liability, regardless of the offence charged. This is because he would restrict the automatism prior fault rules to cases where D foresees not only that her conduct might lead to involuntariness, but also that that involuntariness might lead to relevant harms or be performed in relevant circumstances. As discussed elsewhere, these cases do not require a substitution of missing *mens rea* elements and are better dealt with through the rules of causation.[92] Where a substitution *is* required to construct liability, where D foresees possible involuntariness (is at fault) but does not foresee risks of harm, the court in *Bailey* would not find liability.

We are left with two areas of confusion. First, does the basic/specific intent distinction have a role in automatism cases? And, secondly, if prior fault automatism can construct liability for at least basic intent offences, what must D foresee at T1 to be considered at fault? These questions have been touched upon in a recent flurry of automatism cases in the Court of Appeal,[93] but received very little specific consideration. Importantly, however, the Law Commission (in its recent discussion paper)[94] has provided some analysis on these questions and has attempted to draw principles from the case law: principles that may well encourage greater consistency in future cases. For the Commission, despite its reservations as to the specific/basic intent distinction, there is a useful recognition that prior fault for automatism should be consistent with prior fault for intoxication. Thus, contrary to *Bailey*, prior fault automatism should be able to replace a lack of *mens rea* for basic intent offences, even where D merely foresees (at T1) a potential loss of voluntariness as opposed to future results or circumstances.[95] The Commission also concludes, with reference to our discussion above, that subjective (as opposed to objective) foresight should be required as to that loss of voluntariness.[96] Such conclusions are useful from the point of view of

89 [1983] 1 WLR 760.

90 Ibid 764.

91 *Bailey* [1983] 1 WLR 760, 765 (Griffiths LJ).

92 See Child (n 2).

93 *C; M; H* [2013] EWCA Crim 223.

94 Law Commission (n 3).

95 The Law Commission is not, perhaps, as clear on this point as it could have been in its discussion. However, it is a necessary conclusion from its very helpful flow chart: ibid 95 and para 6.77.

96 Ibid 6.12–28.

certainty, however, whether they are correctly made is likely to have been influenced by their categorisation of automatism as a defence (as discussed above).

The Law Commission has suggested a limited classificatory exception, to cover the scenario where a defendant becomes intoxicated through taking drugs in accordance with a medical prescription, or reasonably in the circumstances and, through no fault of his own, suffers a reaction which causes a loss of capacity.[97] The individual in this scenario, contrary to earlier perspectives, would not fit within involuntary intoxication provisions, but rather the newly articulated RMC defence.[98] This formulation is cognisant, to a degree, of the American Law Institute acceptance of pathological intoxication providing an affirmative defence:[99] pathological intoxication defined as the rapid onset of acute intoxication following consumption of intoxicants, which is insufficient to cause intoxication in most people.[100] The condition is standardised as 'involuntary' as the actor was unaware that the substance would intoxicate her to the extent it did, and, moreover, is distinguished from self-induced incapacity as contextually it takes the individual 'by surprise':[101]

> [A] provision was required because of a concern that bizarre behaviour caused in part by an abnormal bodily condition (in some cases, in others the atypical intoxication can be related to mental disturbance) . . . would not seem to fall under s. 4.01 [the insanity defence].[102]

The final area of debate is one that has emerged in recent years as the defining issue for the intoxication rules, but is rarely discussed in the context of automatism. It asks, if D's prior fault results in liability for a basic intent offence, is D fairly labelled and criminalised for that offence? To answer this question, we must weigh the blameworthiness of D's prior fault (at T1) against the blameworthiness of the missing offence elements (at T2), looking for equivalence in order to justify the potential for substitution. In the context of the intoxication rules, the traditional interpretation of intoxication as a defence had supplied a useful mechanism for attempts to avoid the burden of establishing this equivalence: as a defence, you simply need to consider what should be required to excuse a (fictionally completed) criminal wrong. However, as more commentators have accepted (in line with Simester) that intoxication is never a defence, it is clear that such equivalence must be demonstrated to justify liability. That is, equivalence between voluntary intoxication at T1 and the missing basic intent elements such as recklessness as to specific circumstances and/or results at T2 (and, potentially, missing voluntary conduct).

Still focusing on the equivalence thesis in the context of intoxication, there have been several attempts to demonstrate such equivalence that are useful to our discussion of automatism. In each case, the argument is made that the effects of alcohol or other dangerous drugs are sufficiently well know within the population, that when becoming intoxicated at T1, we can assume that D (at least subconsciously) was aware of creating a danger at T2. As Horder states:

97 Law Commission (n 3) 6.51. The new defence presupposes that D suffers from an RMC.

98 Ibid.

99 See Lawrence P Tiffany, 'Pathological Intoxication and the Model Penal Code' (1990) 69 Nebraska Law Review 763; and Lawrence P Tiffany and Mary Tiffany, 'Nosologic Objections to the Criminal Defense of Pathological Intoxication: What do the Doubters Doubt?' (1990) 13 International Journal of Law and Psychiatry 49.

100 MPC s 2.08(5); and see Reed and Wake (n 80).

101 Tiffany (n 99) 768.

102 MPC and Commentaries ss 2.08, 7, 12 (Tentative Draft No 9 1959).

> . . . my blundering into harm can hardly be said to be spontaneous or unexpected if I have knowingly done that which is—as is taken to be common knowledge—liable to make me blunder.[103]

Such logic is often repeated in the courts, for example:

> . . . there is nothing unreasonable or illogical in the law holding that a mind rendered self-inducedly insensible . . . through drink or drugs, to the nature of a prohibited act or to its probable consequences is as wrongful a mind as one which consciously contemplates the prohibited act and foresees its probable consequences (or is reckless as to whether they ensue).[104]

And, recently, by the Law Commission:

> Given the culpability associated with knowingly and voluntarily becoming intoxicated, and the associated increase in the known risk of aggressive behaviour, there is a compelling argument for imposing criminal liability to the extent reflected by that culpability.[105]

The problem with this logic, as highlighted by Simester and others, is that demonstrating some general fault in becoming intoxicated (at T1) is not the same as a specific foresight of a specific risk at the point that the *actus reus* of the offence is committed (at T2).[106] Thus, the question of *equivalence* between the two types of fault remains open.

If we are correct that the structure of prior fault automatism mirrors that of intoxication as a potentially inculpatory tool, then we would expect a similar debate to arise here as well. This has not been the case. It is our contention that this debate is missing from automatism chiefly because of the three areas of uncertainty discussed above. How can we measure the equivalent blameworthiness of D's prior fault automatism if we do not know the required degree of D's involuntariness, whether that prior fault is subjective or objective, and what exactly D must foresee? And what can this fault be balanced against if we do not know whether the prior fault rules apply to basic and/or specific intent offences? However, our criticism here is not only that legal uncertainties have beguiled the debate, but that, as soon as these uncertainties are resolved, the incoherence of the automatism rules within the current law becomes fully apparent.

The Law Commission's search for consistent interpretations of the current law (introduced above) allow us to set out a similar equivalence debate for prior fault automatism as exists for the intoxication rules. Thus, for the Commission, consistently with intoxication, prior fault automatism will apply (constructing liability) where D subjectively foresees the possibility of becoming automatic at T1, and this is *equivalent* to missing basic intent *mens rea* (recklessness, negligence etc) and voluntariness at T2. Although this logic is essential to the justifiable operation of the current law, it is an equivalence that few commentators would be willing to accept. Indeed, even those advocating equivalence between missing basic intent *mens rea* and intoxication are likely to have a problem here. The court in *Bailey*, for example, has stated that:

---

103 Jeremy Horder, 'Sobering Up? The Law Commission on Criminal Intoxication' (1995) Modern Law Review 534, 537.

104 *DPP v Majewski* [1977] AC 433, 479.

105 Law Commission, *Intoxication and Criminal Liability* (Law Com No 314 2008) para 1.55. It should be noted that the Commission rejected this approach as it applied to offences, but endorsed it in relation to individual offence elements; and see, John Child, 'Drink, Drugs and Law Reform: A Review of Law Commission Report No 314' [2009] Criminal Law Review 488.

106 See Simester (n 1); and Williams (n 2).

> . . . it seems to us that there may be material distinctions between a man who consumes alcohol or takes dangerous drugs [prior fault intoxication] and one who fails to take sufficient food after insulin to avert hypoglycaemia [prior fault automatism]. It is common knowledge that those who take alcohol to excess or certain sorts of drugs may become aggressive or do dangerous or unpredictable things, they may be able to foresee the risks of causing harm to others but nevertheless persist in their conduct. But the same cannot be said without more of a man who fails to take food after an insulin injection.[107]

Even if one is persuaded by the (we believe, unconvincing) case for equivalence between intoxication and basic intent *mens rea*, such equivalence is even more difficult to demonstrate in the context of prior fault automatism. It is clear that the current law relating to prior fault automatism must be relying on such equivalence in order to justify the use of prior fault to replace missing offence elements and construct liability. Therefore, if no such equivalence is present, any defendant convicted and labelled as an offender under this doctrine has been unfairly treated.

Looking forward

As with the inculpatory rules on intoxication, there are three main options for automatism moving forward: attempting to justify something similar to the status quo; abolishing the prior fault rules and acquitting D whenever there is evidence of involuntariness; or creating a bespoke prior fault-based offence. As with intoxication, it is the last of these options that seems most attractive.

The first option, then, is to seek equivalence between prior fault automatism (at T1) and missing basic intent *mens rea* and voluntariness (at T2). This route, of course, has been the one employed by the Law Commission for the intoxication rules.[108] It is contended that this option (and thus any thought of maintaining the current law) lacks viability. Even within the courts and those advocating the equivalence thesis in the context of intoxication, there is very little support for its automatism equivalent: there is simply no public consensus (as is claimed for intoxication) that foresight of involuntariness equates to a general foresight of danger to others or property. The law could require more of D at T1 to create something closer to comparable fault, and this was what we meant above when we suggested that old debates relating to degrees of involuntariness and subjective versus objective foresight were recast by the recognition of prior fault automatism as an inculpatory tool. Without additional foresight of future circumstances or results at T1, as per *Bailey*, any prospect of equivalence seems unlikely. But following *Bailey* is also unattractive: where D satisfies the *mens rea* of the offence at T1 (is reckless as to future results, for example) then there is no need for rules of prior fault, it is enough to apply the general rules of causation. We could require some non-specific foresight of danger (as well as involuntariness) at T1 as a middle ground between *Bailey* and mere foresight of involuntariness. Even with this additional fault at T1, it would be difficult to demonstrate equivalence with missing *mens rea* as to specific offence elements and as to voluntariness at T2. In short, the current law, and even adjustments based on the current structure of the law, appear undesirable.

The second option moving forward concedes that the current law is operating on the basis of an unjustifiable equivalence, and in the absence of such equivalence no liability (at all) should be found. This option is exemplified in the quotation from *Bailey* above, maintaining that D should only be liable where she acts at T1 with (essentially) the full *mens*

107 *Bailey* [1983] 1 WLR 760, 764–65 (Griffiths LJ). For academic support of this position, see Horder (n 103) 544–45.

108 Law Commission No 314 (n 105).

rea required for the principal offence. As stated, liability here does not require constructed liability through prior fault, but simply the standard rules of causation: D's acts at T1 become the conduct element of the *actus reus* (coinciding with her *mens rea*), and the question is then whether these acts caused the result at T2. Where D lacks *mens rea* at T1, even where she foresees future involuntariness (exhibits some prior fault), the court in *Bailey* and this option moving forward would find no liability. In this manner, missing basic intent offence elements could never be substituted for by evidence of non-intoxicated prior fault automatism. In our view, unlike the first option, this option has an obvious attraction.

The final option will be sketched here, but will require considerably more work before it becomes truly viable: a new prior fault automatism offence. In line with option two, the first premise of this approach is that prior fault automatism is not equivalent to missing offence elements and should never be able to reconstruct or substitute for such elements. Moving beyond the second option, a new offence would recognise that certain cases of prior fault automatism that lead to harms are deserving of criminalisation: not through a fiction that D has completed offence elements that are missing (as with the current law and option one), but by creating an offence that accurately labels and punishes that which D *has* done. This approach has been advocated in the context of intoxication for many years, seeking to criminalise and label D in line with her actual conduct: voluntarily becoming intoxicated and causing harm.[109]

In this regard, Williams[110] has recently advocated the creation of a new alternative intoxication offence, facilitating inculpation where a defendant 'commits [the *actus reus* of offence X] while intoxicated'.[111] The essence of this suggestion is that the combined simulacrum of intoxication and harm, despite absence of *mens rea*, is inculpatory in circumstances where the harm *per se* would not be criminalised without *mens rea* or intoxication. The bespoke offence template has also been identified by Loughnan[112] as the way forward for intoxication prior fault doctrine, arguing that this approach would 'make overt the connection between intoxication and criminal liability, sabotaging the myth that intoxication is some kind of "defence" to a criminal charge'.[113] This mirrors the standardisation adopted in German criminal law, identifying a specific offence to detail the inculpatory nature of prevening fault for intoxication:

> Whosoever intentionally or negligently puts himself into a drunken state by consuming alcoholic beverages or other intoxicants shall be liable to imprisonment of not more than five years or a fine if he commits an unlawful act while in this state . . . The penalty must not be more severe than the penalty provided for the offence which was committed while he was in the drunken state.[114]

When considering a new offence of prior fault automatism causing harm, one of the main advantages is that we can explicitly and accurately consider what is required for criminalisation in a manner that is hopelessly confused within the current law. Returning to

109 See Smith and Williams in the Criminal Law Revision Committee (n 2); and see Child (n 2).

110 Williams (n 2) 277.

111 Ibid 277.

112 Arlie Loughnan, *Manifest Madness: Mental Incapacity in the Criminal Law* (Oxford University Press 2012) 314–16.

113 Ibid 315; and see further, Arlie Loughnan, 'Mental Incapacity Doctrines in Criminal Law' (2012) 15 New Criminal Law Review 1; Douglas Husak, 'Intoxication and Culpability' (2012) 6 Criminal Law and Philosophy 363; Susan Dimock, 'What are Intoxicated Offenders Responsible for? The "Intoxication Defense" Re-examined' (2011) 5 Criminal Law and Philosophy 1; and Gideon Yaffe, 'Intoxication, Recklessness and Negligence' (2012) 9 Ohio State Journal of Criminal Law 545.

114 Die Übersetzung [German Criminal Code], 13 November 1998, BGB1 [Federal Law Gazette] s 323a, amended by Article 3 of the Law of 2 October 2009.

first principles, our task is to identify a public wrong (if there is one) deserving of criminalisation. The first question will be to consider and define the fault required of D at T1. Assuming that a new offence of this kind is not simply criminalising those who are likely to lose control and be a danger,[115] but rather those who are at fault in some way for their future involuntariness, it is likely that we would require some subjective foresight of future involuntariness and also foresight of a non-specific danger to others associated with that involuntariness. The preference for *subjective* foresight, and foresight of *danger* as well as simple involuntariness, is designed to target only those who lose control at T1 with a similar level of blameworthiness as the voluntarily intoxicated (where such foresight is, perhaps, more easily assumed). Although often neglected in debates on a possible intoxication offence, we would also need to consider what types of harms would be required at T2.[116] For example, although we may conclude that a criminal wrong is completed where D's prior fault automatism results in bodily harm to a victim or property damage (harms that may be linked to D's general foresight at T1 of future danger to people and property), it may be inappropriate to criminalise D where the same prior fault leads to an unpredicted harm (for example, to property rights). Indeed, even within the categories of bodily or property harm, we may conclude that D's wrong is only deserving of criminalisation where that harm is of a particularly serious degree.

Having established a potential offence of prior fault automatism to work alongside existing offences, it is then appropriate to discuss the degree of involuntariness required for D to be classed as automatic. The extremes are clear: complete voluntariness would be dealt with under existing offences, and total involuntariness (i.e. unconsciousness) would be considered under the new offence. The space between these extremes gives rise to an interesting debate. In view of our previous discussions above, we would broadly support the idea that D is not sufficiently competent to create the nexus of agency required for existing offences unless she acts with a high level of voluntariness. Therefore, if we cannot ascribe those later harms to D through traditional means, it only seems appropriate to blame D for those harms where she acted with prior fault. Our view is that the potential new offence should therefore operate across a considerably wider class of cases than the 'automatism defence' does under the current law. Under the current law, where D causes harm in a partially involuntary (or dissociative) state, she is likely to be convicted for existing offences: failing to meet the high threshold of the automatism 'defence'. Under the approach mooted, in contrast, D would either be liable for the new offence in circumstances of prior fault (accurately labelling what D has done), or D would be acquitted.

A presented danger, of course, in creating a new prior fault automatism offence, as with intoxication, is that it creates the potentiality of split juries across the bifurcatory offence particularisations.[117] It is submitted, however, that this concern may be more apparent than real, as fact-finders have been commendably robust in ascription of liability for murder/manslaughter and respective gradations, contextualising gross negligence manslaughter, and substitute alternative verdicts for s 18, s 20 or s 47 Offences Against the Person Act 1861 thresholds of harm. Moreover, as Williams asserts in the province of a prior fault intoxication offence, difficulties may be abrogated by utilisation of majority verdict precepts[118] and the Law Commission, although reticent over creation of prior fault

115 For discussion of automatism and dangerousness (as opposed to fault), see Mackay and Mitchell (n 6) 905; and see Nicola Padfield, 'Exploring a Quagmore: Insanity and Automatism' (1989) Cambridge Law Journal 354, 357.

116 See Child (n 2).

117 Williams (n 2) 282.

118 Ibid.

inculpatory offences, nonetheless acknowledges that jurors may be 'as capable of handling the decision between the two offences [in this context] as they are of handling the many other cases where, at present, a defendant may be convicted of alternative offences'.[119]

A final point that should be considered as to the potential new offence relates to its consistency with intoxication where D lacks *mens rea* as a result of her prior fault, but is not automatic. It is an interesting point of comparison that, where D lacks *mens rea* as a result of prior fault but still acts voluntarily, if D is intoxicated, we may find liability, but if D is not intoxicated, we would not: even under the prior fault automatism offence discussed here, a lack of voluntariness at T2 is an essential part of the offence. For example, D1 and D2 both knock over V causing bodily harm, both are acting voluntarily but did not notice V in their way. D1 does not notice V because she is voluntarily intoxicated and will therefore be liable for an offence under the current law (constructing liability), or for a proposed intoxication offence. D2 does not notice V because she is in a daze having negligently not eaten after taking insulin for her diabetes and will therefore not be liable for any offence as she lacks *mens rea*. Some of this inconsistency may be corrected by our wider view of involuntariness within the new offence (leading to liability in cases that would not have done within the current law), but there will still be cases that do not fall within even our inflated definition of involuntariness. It may be that these cases do not warrant criminalisation and the inconsistency with intoxication is simply a reflection of the different wrongs involved in each route to liability. We are generally minded towards this conclusion. However, the debate is important and emerges most clearly with the recognition that both intoxication and automatism are playing inculpatory roles.

Conclusion

The central aim of this article has been to set out and justify the contention that automatism is never a defence. Where D is not at fault for her lack of voluntariness, the term 'automatism' is simply a shorthand explanation that D does not satisfy an essential element of every offence: voluntary conduct. Where D is at fault for her lack of voluntariness, the automatism rules (within the current law) become an inculpatory tool through which to substitute for missing offence elements and construct liability.

Having recognised that automatism plays an inculpatory role within the law, we have then analysed this role and concluded that it is defective: prior fault automatism lacks the equivalent blameworthiness necessary to fairly substitute for even missing basic intent offence elements. It is from here that we have discussed the possibility of a new automatism offence, to recognise the criminal blameworthiness of D's conduct in certain cases, but to do so in a coherent manner that appropriately criminalises and labels the defendant. Looking at the outline of the potential new offence, we are in a much better position to evaluate the future role of automatism in the criminal law. If we do *not* believe that such an offence is deserving of criminalisation, then the current law must be changed to prevent prior fault automatism constructing liability under any circumstances. If we *do* believe that such an offence has a place within the criminal law, then the current law should be changed to reflect this more clearly, and we must focus on exactly how it should be defined.

The obvious next step will involve a similar examination of insanity. In line with intoxication and automatism, prior fault insanity (recognised by the Law Commission)[120] may also have a role in the construction of offences, but the case law here is

119 Law Commission, *Intoxication and Criminal Liability* (Law Com No 127 1993) para 6.27.

120 Law Commission (n 3) para 5.36–76 and 6.30–31.

undeveloped.[121] If D negligently or recklessly *fails* to take medication resulting in involuntariness or dissociation and ultimately harm to the person or property, it would seem strange if they avoided liability where an intoxicated D, or D who actively medicated improperly, would not. Indeed, it may be that such liability is best served through a new or combined offence alongside voluntary intoxication and prior fault automatism causing harm. However, insanity will be a more difficult case. After all, unlike automatism and intoxication, where D satisfies both the *actus reus* and *mens rea* of an offence, insanity may still be a relevant consideration: it is, sometimes, a defence.

121 See discussion of *Harris* in Ronnie Mackay, 'Case Comment: *R v Coley*; *R v McGhee*; *R v Harris*: Insanity—Distinction between Voluntary Intoxication and Disease of the Mind Caused by Voluntary Intoxication' [2012] Criminal Law Review 923; and see also Edward W Mitchell, 'Culpability for Inducing Mental States: The Insanity Defence of Dr Jekyll' (2004) Journal of the American Academy of Psychiatry Law 63.

NILQ 65(2): 187–204

A new partial defence for the mercy killer: revisiting loss of control

BEN LIVINGS[1]

Senior Lecturer in Law, University of New England, Australia

Introduction

> Murder, as every practitioner of the law knows, though often described as one of the utmost heinousness, is not in fact necessarily so, but consists in a whole bundle of offences of vastly differing degrees of culpability, ranging from brutal cynical and repeated offences like the so called Moors murders to the almost venial, if objectively immoral, 'mercy killing' of a beloved partner.[2]

It is a truism to assert that mercy killing amounts to murder under the criminal law of England and Wales,[3] but this does not reflect reality insofar as a murder conviction is a very unlikely outcome, even where cases are pursued through the criminal justice system.[4] Of what he terms the 'mystery of the disappearing murderers', Huxtable writes: 'the law is geared towards ensuring that mercy killers reappear in most cases as either assistants in suicide or manslaughterers'.[5] Each of these routes allows for a more flexible sentencing response than the mandatory life sentence imposed for murder, and avoidance of the imposition of an offence label deemed inappropriate. Whilst they may achieve similar ends,

1 The author wishes to thank Nicola Wake and Amy Purvis for their comments on a draft of this article, and attendees at the 2013 Socio-Legal Studies Association and Society of Legal Scholars annual conferences for helpful discussion of some of the ideas presented here.

2 *R v Howe* [1987] 1 AC 417, 433 (Lord Hailsham).

3 As Lord Goff asserts: 'if I kill you from the motive of compassion (so-called mercy killing) I nevertheless intend to kill you and the crime is one of murder' (Lord Goff, 'The Mental Element in the Crime of Murder' (1988) 104 LQR 30).

4 Richard Huxtable, *Euthanasia, Ethics and the Law: From Conflict to Compromise* (Routledge Cavendish 2007) 36. Until the recent case of *R v Inglis* [2011] 1 WLR 1110, *R v Cocker* [1989] Crim LR 740 appears to be the only recent example of a murder conviction.

5 Huxtable (n 4). In the recent case of Kay Gilderdale, the judge was highly critical of the Crown Prosecution Service's decision to bring a case for attempted murder <www.solicitorsjournal.com /news/public/care/prosecuting-mother-attempted-murder-public-interest-dpp-insists> accessed 10 February 2014.

however, they comprise very different offences,[6] and my focus here is on the availability of voluntary manslaughter. I shall not engage in any meaningful way with the normative arguments about whether or not mercy killing should be lawful, or indeed broader questions of policy and the criminal law in relation to the end of life, beyond making the relatively uncontentious assumption that murderer is an inappropriate label for a mercy killer in the vast majority of cases, especially in light of the continuing existence of the concomitant mandatory life sentence.

My argument proceeds in three parts. In part 1, I offer a short account of the partial defences; their historical role and some of the perceived inadequacies that fed the changes brought about by the Coroners and Justice Act (CJA) 2009. From here, I look in more detail at the adoption of diminished responsibility as a *de facto* partial defence for the mercy killer under the law that existed before the inception of the CJA 2009, which stands as a demonstration of the way in which the partial defences serve to protect the defendant from an inappropriate murder conviction.

Part 2 looks at the changes to the partial defences implemented by the CJA 2009 and their potential impact. Much of the discussion here is necessarily speculative, insofar as the appellate courts have yet fully to consider the finer details of the pleas. However, plausible claims have been made in relation to a narrowing effect when it comes to diminished responsibility, which may now exclude mercy killing cases. In light of this possibility, I look to the changes brought in by the new partial defence of loss of control, the impact of which is far less straightforward.

In part 3, I dissect the new plea of loss of control, and challenge the view that it amounts to a more restrictive provision than its predecessor. In so doing, I make two main assertions: firstly, that the wording and structure of loss of control admit the possibility of its successful application to a case of mercy killing; and, secondly, that this is arguably preferable to the situation that had operated under diminished responsibility. Where the benign conspiracy isolated defendants by applying a spurious pathologisation in order to afford them a partial defence, loss of control invariably demands consideration of the social context in which defendants act, and effectively asks jurors whether they might react in the same way, if faced with the circumstances of the defendant. This more overt confrontation of the realities of homicide brings a social calibration that is wholly in keeping with the philosophy underlying the partial defences, and may assist in opening up debates about end of life issues, and thus contribute to resolving how the law should best respond.

1 The partial defences pre-CJA 2009 and the impetus for change

The partial defences of 'diminished responsibility' and 'loss of control' (formerly 'provocation') operate to transmute murder into voluntary manslaughter. Aside from the nomenclature, the principal effect of this is in sentencing; instead of the mandatory life sentence afforded to convictions for murder, the judge will have discretion as to the sentence

6 Under s 2 of the Suicide Act 1961, a person found guilty of assisting suicide is liable to imprisonment for up to 14 years. Much has recently been written on the issuance of guidelines by the Director of Public Prosecutions in the wake of the House of Lords' judgment in *R (on the application of Purdy) v DPP* [2010] 1 AC 345. For discussion, see: Ben Livings, 'A Right to Assist? Assisted Dying and the Interim Policy' (2010) 74 Journal of Criminal Law 31; Penney Lewis, 'Informal Legal Change on Assisted Suicide: the Policy for Prosecutors' (2011) 31(1) Legal Studies 119.

to pass.[7] The partial defences are therefore inextricably linked to the current structure of the law of homicide, as 'instances of murder where the application of the mandatory life sentence appears too draconic in comparison to the blameworthiness of the defendant's act'.[8] In addition to their utility to the defendants who successfully plead them, the partial defences serve a symbolic role; as Horder notes, they are intended to further a 'liberal common good . . . the preservation of the humanity of the criminal law'.[9] In this respect, the Law Commission acknowledges that they are inevitably susceptible to changes in social mores and public opinion, and that this relationship should be symbiotic; neither wholly dependent nor wholly divorced from such forces.[10] Given their role in mitigating its potential injustice, the Commission observes that, absent 're-examination of the surrounding law of murder', the partial defences are subject to inevitable expansionary pressures.[11]

The reform to the partial defences enacted through the CJA 2009 is largely built upon work by the Law Commission,[12] which points to longstanding dissatisfaction with their operation.[13] A major spur to reform, prominent in the work of the Commission and addressed in the legislation, was a perceived gender-bias widely held to subsist under the plea of provocation. This was felt to disadvantage, in particular, women who killed their abusive partners, in respect of whom diminished responsibility was perceived as inappropriate,[14] and provocation problematic. Here, the typically male response of sudden, reactive anger was privileged over a more typically female 'slow-burn' response born of fear.[15] Thus, as Norrie explains, 'the abused woman was rather shoe-horned into the defence of provocation, and this led to difficulties in defending her'.[16] In an attempt to make it more accessible and appropriate to such a defendant, a number of obstacles were

7 Norrie points to differentiation within the mandatory life sentence brought about by the sentencing provisions contained in s 269 of the Criminal Justice Act 2003 and argues that this has brought about 'three degrees of offence seriousness' when it comes to murder. For Norrie, 'the 2003 Act has in symbolic terms chipped away at the notion of one, uniquely serious, crime possessing one, uniquely serious, penalty': Alan Norrie, 'Between Orthodox Subjectivism and Moral Contextualism: Intention and the Consultation Paper' [2006] Criminal Law Review 486.

8 David Ormerod, *Smith and Hogan's Criminal Law* 13th edn (Oxford University Press 2011) 505. As Reed and Bohlander assert, 'cases of manslaughter because of loss of control and diminished responsibility are in fact nothing but instances of murder where the application of the mandatory life sentence appears too draconic in comparison to the blameworthiness of the defendant's act': Alan Reed and Michael Bohlander, *Loss of Control and Diminished Responsibility: Domestic, Comparative and International Perspectives* (Ashgate 2011) 1.

9 Jeremy Horder, *Homicide and the Politics of Law Reform* (Oxford University Press 2012) 202.

10 Law Commission, *Partial Defences to Murder* (Law Com Report No 290 2004) [3.146]. Any study of the history of the defence of provocation shows that it has changed as public values have changed, and that the change of social attitudes is a gradual process. Public opinion should not necessarily decide what the law should be, for public opinion may not be carefully thought out and the law may itself help to shape public opinion, but it should properly be taken into account.

11 Ibid [2.8].

12 Particularly: ibid; Law Commission, *Murder, Manslaughter and Infanticide* (Law Com Report No 304 2006).

13 The Law Commission stated the following: '[T]he terms of section 3 are now, in large measure, effectively ignored and scarcely anyone has a good word for it': Law Com No 290 (n 10) [2.8]. Ormerod describes the partial defences as 'subsist[ing] unsatisfactorily in order to avoid the mandatory life sentence for murder': Ormerod (n 8) 505.

14 As Miles notes, there had long been objections to 'a law which requires such defendants to plead their own mental illness to obtain mitigation, rather than offering them a defence which narrates a story of self-defence against unjustified violence': Jo Miles, 'The Coroners and Justice Act 2009: A "Dog's Breakfast" of Homicide Reform' (2009) 10 Archbold News 6.

15 See, for example, D Nicholson and R Sanghvi, 'Battered Women and Provocation' [1993] Criminal Law Review 728.

16 Alan Norrie, 'The Coroners and Justice Act 2009—Partial Defences to Murder (1) Loss of Control' [2010] Criminal Law Review 275.

removed and a specific limb to loss of control developed. An apparently unintended, and yet arguably inevitable, consequence of this has been to expand the possible application of the plea when compared to its predecessor in ways that are explored below.

In eschewing broader and more radical reform, the overriding aims of the criminal law reforms contained in the CJA 2009 are set out in the Explanatory Notes to the Bill, which allude to a desire to 'improve its clarity, fairness and effectiveness',[17] echoing some of the goals of the Law Commission.[18] The legislation has kept the basic structure of the partial defences intact, a move criticised by Horder, who suggests that the CJA 2009 amounts to a missed opportunity.[19] For Horder, the government's 'heavily doctrinal' approach to homicide reform 'mired the law in a bog of ever-thickening legal complexity from which, following the coming into force of the Coroners and Justice Act 2009, there is now little hope of escape'.[20]

Beyond those suggestions which went on to form the substance of the revised partial defences, as enacted in the CJA 2009, the Law Commission considered there to be 'other circumstances which may significantly extenuate moral responsibility for homicide', and offered as an example 'the genuine case of mercy killing', but considered them to be 'outside the terms of our present review'.[21] Ashworth notes the 'pragmatism' of the Law Commission in this respect, writing: '[a]ny attempt at homicide law reform that includes this topic is likely to meet acute controversy that may well derail the whole project of reform'.[22] In House of Lords debates during the passage of the CJA 2009, Lord Lloyd suggested a broad, discretionary approach to the founding of a partial defence (that murder should be mitigated where there are 'extenuating circumstances'),[23] and referred specifically to mercy killing as a spur to his proposal. In his view, a partial defence deriving from extenuating circumstances 'would provide an answer to the case of mercy killing'.[24] Lord Lloyd's proposal gained support within the House of Lords, but was not to form part of the CJA

17 Coroners and Justice Bill, Explanatory Notes [15] <www.publications.parliament.uk/pa/ld200809/ldbills/033/en/09033x—.htm> accessed 10 February 2014.

18 The Law Commission emphasised bringing 'order, fairness and clarity to the law of homicide': Law Com No 304 (n 12) [2.4].

19 For Horder, 'the 2009 Act wrongly retained the partial defence structure left in place by the doctrinal path to reform ushered in by the 1957 Act': Horder (n 9) 199–200).

20 Ibid viii–ix.

21 Law Com No 290 (n 10) [3.63].

22 Andrew Ashworth, 'Principles, pragmatism and the Law Commission's recommendations on homicide law reform' [2007] Criminal Law Review 333.

23 Lord Lloyd's proposed amendment was as follows: 'Murder: extenuating circumstances: (1) In a trial for murder the trial judge may in the course of his summing up direct the jury that if they are satisfied that the defendant is guilty of murder, but are of the opinion that there were extenuating circumstances, they may on returning their verdict add a rider to that effect. (2) The judge may not give such a direction unless there is evidence on which a reasonable jury might so find. (3) Where the jury has so found, the judge shall not be obliged to pass a sentence of life imprisonment but may pass such other sentence as he considers appropriate having regard to any extenuating circumstances found by the jury. (4) If it appears to the Attorney General that the sentence so passed is unduly lenient he may refer it to the Court of Appeal under section 36 of the Criminal Justice Act 1988 (c 33) (reviews of sentencing).': (HL Deb 30 June 2009, vol 99, col 150.

24 Lord Lloyd also cited: 'difficult cases . . . where a soldier or a policeman fires and kills in the agony of the moment'; 'the case of the battered wife'; and 'awkward cases on the edge of provocation and of diminished responsibility'; he acknowledged that there were probably 'many others': HL Deb 30 June 2009, vol 99, col 152.

2009.[25] As a result of the narrow approach to reform, the legislation makes no reference to mercy killing, but the subject resonates in spite of the silence.

A 'BENIGN CONSPIRACY' AND DIMINISHED RESPONSIBILITY AS A *DE FACTO* HOME FOR THE MERCY KILLER

Before the coming into force of the CJA 2009, the success of a diminished responsibility plea hinged on being able to establish that the defendant was 'suffering from such abnormality of mind . . . as substantially impaired his mental responsibility for his acts and omissions in doing or being a party to the killing'.[26] The strict application of this to the mercy killer can be problematic; as Dell notes, mercy killers tend to display a 'total lack of mental disorder'.[27] However, the obscure language of s 2 enabled a degree of interpretation that allowed a 'benign conspiracy' to operate between 'psychiatrists, lawyers and judges', in order to bring the mercy killer within the definition.[28] Thus, diminished responsibility became a *de facto* partial defence for the mercy killer, percolating through the criminal process and diverting the defendant from a murder conviction. As Miles notes, the 'benign conspiracy' extended into aspects of the criminal justice system beyond the application of the substantive criminal law, affecting the practices of the Crown Prosecution Service in 'pleas of guilty to manslaughter on grounds of diminished responsibility'.[29] Ost describes a 'desire to treat mercy killers with compassion' and points out that 'medicalisation of the mercy killer's behaviour . . . provides this sympathetic legal response with a cloak of respectability'.[30] The effective systematisation of this is evident in the advice given in *Blackstone's Criminal Practice*: '"Mercy killing" can . . . be dealt with as manslaughter, where the dilemma which has caused the accused to kill can be said to have given rise to depression or some other medically recognised disorder which can be said to be the cause of an abnormality of mind.'[31]

25 Lord Pannick said: 'I support this amendment for all the reasons that have been given by the noble and learned Lord, Lord Lloyd of Berwick, and by so many other noble Lords. It is striking indeed that so many lawyers should agree about anything. I would add one further argument in favour of the amendment. We all know that there are occasions when the jury is reluctant to convict, despite compelling evidence that the defendant is guilty of murder. The jury is reluctant to convict because of its concern that on the facts of the case a mandatory life sentence is simply inappropriate. One of the great attractions of the amendment is that it would involve the jury in the assessment of whether there are extenuating circumstances.': HL Deb 30 June 2009, vol 99, col 157.

26 Homicide Act 1957, s 2. This could have arisen 'from a condition of arrested or retarded development of mind or any inherent causes or induced by disease or injury'. In *Byrne* [1960] 2 QB 396, Lord Parker CJ contrasted 'abnormality of mind' with the expression 'defect of reason' for the purposes of the M'Naghten Rules. The court held that an 'abnormality of mind' was wide enough to cover 'the mind's activities in all its aspects, not only the perception of physical acts and matters, and the ability to form a rational judgement whether an act is right or wrong, but also the ability to exercise willpower to control physical acts in accordance with that rational judgement'.

27 S Dell, 'The Mandatory Sentence and section 2' (1986) 12 Journal of Medical Ethics 28–31, 30. See also: S Dell, *Murder into Manslaughter: The Diminished Responsibility Defence in Practice* (Oxford University Press 1984).

28 Edward Griew, 'The Future of Diminished Responsibility' [1988] Criminal Law Review 75, 79–80.

29 Although central to the operation of the law and an omnipresent concern in all matters relating to the practical role and function of the criminal law, the wide-reaching influence of prosecutorial discretion is largely outside the remit of this article.

30 Suzanne Ost, 'The De-Medicalisation of Assisted Dying: Is a Less Medicalised Model the Way Forward?' (2010) 18 Medical Law Review 497.

31 David Ormerod et al (eds), *Blackstone's Criminal Practice 2012* (Oxford University Press 2011) 186. It should be noted that this advice has not changed since the advent of the CJA 2009; Huxtable cites the same passage which appears in the 2007 edition (Huxtable (n 4) 41).

The Court of Appeal's recent judgment in *R v Inglis* is an illustration of the extent to which this approach had become ingrained.[32] Frances Inglis had killed her son in what she perceived to be an act of mercy, but refused to plead diminished responsibility and was convicted of murder. There are indications in the Court of Appeal's judgment that a diminished responsibility plea may have been accepted, and hints at some consternation that it was not submitted as a partial defence.[33] To this end, principles derived from diminished responsibility were employed in consideration of sentence: 'In our view her mental responsibility for her actions, driven as she was by a compulsive obsession, was diminished if not sufficiently for the purposes of the defence of diminished responsibility, certainly to an extent that reduced her culpability.'[34]

Somewhat unsurprisingly, the operation of the benign conspiracy has prompted a mixed response from commentators; whilst Huxtable refers to the outcome as 'often appropriate',[35] he describes it as a 're-casting' which is 'neither accurate nor fair'.[36] For Griew, an obvious problem with such an improvised approach is the potential for inconsistency in its application, dependent as it is upon the 'right combination of professionals'.[37] He identifies a number of sources from which this inconsistency may stem:

> the less robust or sophisticated psychiatrist . . . may not appreciate that the law in its present formal condition is in practice as flexible as it is; so the case may not, pre-trial, be identified as a possible section 2 case. Or the prosecutor or judge may be less sympathetic than some of his colleagues to such liberal use of the section.[38]

Baroness Mallalieu also criticised the operation of the benign conspiracy during debates which preceded the passing of the CJA 2009: 'To watch the judge, the prosecutor and the defence trying to find a way to achieve the right result is not the way justice should be administered. Consistency is needed but so are honesty, openness, clarity and, above all, justice.'[39]

2 Reform under the CJA 2009

In this part, I look at the changes to the partial defences brought about by the CJA 2009. There is good reason to suppose that the reformulated diminished responsibility is narrower in application than its predecessor, and that this may preclude the operation of the benign conspiracy detailed above. However, I suggest that the same cannot be said of the move from provocation to loss of control. Thus, whilst diminished responsibility may no longer be available, there is scope to accommodate the mercy killer within the ambiguous and potentially more expansive loss of control provisions.

In its 2006 report, the Law Commission advocated the modernisation of diminished responsibility, in order to render it 'both clearer and better able to accommodate developments in expert diagnostic practice',[40] and this is reflected in the CJA 2009. To this end, s 52 made changes to s 2 of the Homicide Act 1957,[41] and a successful plea now requires the defendant to prove, on the balance of probabilities, that at the time of the

32 [2010] EWCA Crim 2637. This case was decided according to the situation pre-CJA 2009.

33 Ibid [42] (Lord Judge CJ).

34 Ibid [58] (Lord Judge CJ).

35 Huxtable (n 4) 53.

36 Ibid xv.

37 Griew (n 28) 80.

38 Ibid.

39 HL Deb 30 June 2009, vol 99, col 157 (Baroness Mallalieu).

40 Law Com No 304 (n 12) [5.107].

41 Ss 2(1), (1A) and (1B) of the Homicide Act 1957 were inserted under s 52(1) of the CJA 2009.

killing he was suffering from an 'abnormality of mental functioning' arising from a 'recognised medical condition',[42] and which provides an explanation for the killing.[43]

There has been broad agreement that the reformulation is narrower in scope and applicability than its predecessor.[44] Satisfying the demand for 'clarity' in relation to the excusatory ambit of diminished responsibility has the inevitable effect of reducing the discretion available to the courts, and the wider criminal process, to administer justice according to moral standards. Although the practical import of this remains fully to be seen, the regime as it existed under s 2 appears to have moved from being a morally nuanced question (it had, under the previous formulation, been characterised as 'a moral question of degree and essentially one for the jury')[45] to one of straightforward medical fact. Fortson and Ormerod agree that this change emphasises the role of the expert over broader value judgment,[46] and that this will 'create the opportunity for experts to have even greater influence over the outcome'.[47]

Diminished responsibility retains its name, but Mackay argues that the advent of the CJA 2009 amounts to more than a mere modernisation; that it has brought the potential for real change to the doctrine. For Mackay, this 'radical departure from its former self'[48] has resulted in something 'entirely new',[49] and that one effect of this could be to override the benign conspiracy and remove the partial defence from the reach of the mercy killer:

> While it is difficult to know whether these differing 'abnormality' requirements will exclude any conditions which might formerly have qualified for a diminished responsibility plea, there is a concern that because 'recognised medical condition' focuses exclusively on the need for a defined and demonstrable clinical condition which is medically recognised, it may fail to include those 'mercy killing' cases which in the past qualified for a s 2 plea.[50]

The future availability of diminished responsibility to mercy killers is unclear. Although Miles speculates that the reforms may preclude the use of the plea, 'resulting in more trials and in more murder convictions',[51] it must be assumed that the desire to avoid murder convictions that manifested in the benign conspiracy outlined above is unlikely to dissipate simply because of the enactment of the CJA 2009, and it may be that the changes therefore have little effect. As Kennefick notes:

---

42 Homicide Act 1957, s 2, as amended by CJA 2009, s 52.

43 Homicide Act 1957, s 2(1B) as amended by CJA 2009, s 52.

44 Ormerod (n 8), goes on to say: 'There is no doubt that the new law is stricter than the original s 2.' 530 Although cf Nicola Wake, 'Recognising Acute Intoxication as Diminished Responsibility? A Comparative Analysis' (2012) 76 Journal of Criminal Law 71–98. Wake examines the possible implications under the new formulation of diminished responsibility for the intoxicated defendant and argues that there may be a widening in application here.

45 Ormerod (n 8) 511.

46 Rudi Fortson, 'The Modern Partial Defence of Diminished Responsibility' in Reed and Bohlander (n 8) 37.

47 Ormerod (n 8) 529.

48 Ronnie Mackay, 'The New Diminished Responsibility Plea: More than Mere Modernisation?' in Reed and Bohlander (n 8) 9.

49 Ronnie Mackay, 'The Coroners and Justice Act 2009—Partial Defences to Murder (2) The New Diminished Responsibility Plea' (2010) Criminal Law Review 290. The requirement for a 'recognised medical condition' was regarded as necessary 'to accommodate future developments in diagnostic practice and encourage defences to be grounded in a valid medical diagnosis linked to the accepted classificatory systems which together encompass the recognised physical, psychiatric and psychological conditions': Ministry of Justice, *Murder, Manslaughter and Infanticide* (MoJ CP No 19 2008) [49].

50 Mackay (n 48) 16.

51 Although central to the operation of the criminal justice system, the wide-reaching influence of prosecutorial discretion is largely outside the remit of this article.

Time will tell how the system responds to this particular circumstance in practice. It may well be the case that the courts will likewise stretch their interpretation of the term 'recognised medical condition', or the experts their diagnoses, in order to enable the 'benign conspiracy' to continue, albeit in a slightly different guise.[52]

What follows is predicated upon the possibility that some or all of those mercy killers to whom diminished responsibility was effectively available may be adversely affected by the changes the CJA 2009 made to the plea. In light of this, I suggest that an alternative may lie in the potential breadth of the loss of control provisions brought in under ss 54 and 55 of the CJA 2009 and, further, that this may address some of the complaints that arose in relation to the benign conspiracy, insofar as it would not rely upon the spurious pathologisation of defendants by way of professional collusion. Thus, what the CJA 2009 has taken away, it may simultaneously have reintroduced in a more appropriate form.

From provocation to loss of control

Alongside reform of diminished responsibility, s 56 of the CJA 2009 abolished the partial defence of provocation, which is replaced by a new plea of loss of control.[53] Throughout its history, the doctrine of provocation had exhibited a somewhat unpredictable development, its political and legal foundations articulated over decades of shifting case law.[54] At the core of the plea was the relatively stable concept of a 'loss of self-control', as encapsulated in the summing-up in *Duffy*: 'a sudden and temporary loss of self-control rendering the accused so subject to passion as to make him for the moment not master of his mind'.[55] The law was codified around this core concept in s 3 of the Homicide Act 1957,[56] and its legacy lives on in the new plea. Thus, the provisions of the CJA 2009 retain provocation's bipartite structure, comprising what Norrie succinctly refers to as 'a factual and a regulative aspect';[57] much of the terminology is also retained.

Despite these similarities, the move to loss of control has brought potentially far-reaching change. The majority of commentators agree that the plea of loss of control is more restrictive than its predecessor,[58] and the Explanatory Notes certainly suggest that

52 Louise Kennefick, 'Introducing a New Diminished Responsibility Defence for England and Wales' (2011) 74 Modern Law Review 750.

53 S 56(1) of the Act; s 56(2)(a) provides that s 3 of the Homicide Act 1957, which modified the common law on provocation, ceases to have effect.

54 This is true also of other jurisdictions in which the plea, or a variation of it, plays a role. Of the situation in the US, Nourse writes: 'ideas of provocation have vacillated greatly within criminal law theory': Victoria Nourse 'Reconceptualizing Criminal Law Defenses' (2003) 151 University of Pennsylvania Law Review 1691, 1716.

55 *R v Duffy* [1945] 1 All ER 932 (Devlin J).

56 'Where on a charge of murder there is evidence on which the jury can find that the person charged was provoked (whether by things done or by things said or by both together) to lose his self-control, the question whether the provocation was enough to make a reasonable man do as he did shall be left to be determined by the jury; and in determining that question the jury shall take into account everything both done and said according to the effect which, in their opinion, it would have on a reasonable man.'

57 Alan Norrie, *Law and the Beautiful Soul* (Glasshouse 2005) 121. As Mitchell notes: 'The new loss of self-control plea is similar to the old common law which it replaces in that it requires defendants to comply with both subjective (viz. a loss of self-control triggered by—in crude terms—fear or anger) and objective (the person of normal tolerance etc.) tests.': Barry Mitchell, 'Loss of Self-Control under the Coroners and Justice Act 2009: Oh No!' in Reed and Bohlander (n 8) 40.

58 For Ormerod, the reforms are designed to be 'a much narrower defence than at common law and under s 3 of the 1957 Act': (n 8) 508. See also: Norrie (n 16); Carol Withey, 'Loss of Control, Loss of Opportunity?' [2011] Criminal Law Review 263.

this was the legislators' intention.[59] However, the government described the effect of the legislation as 'raising the bar of the availability [of the partial defence] and extending it to cover those who kill in fear of serious violence as well as those who kill in anger'.[60] This means that the legislation necessarily pulls in two directions, attempting to constrain whilst at the same time broadening the basis of the partial defence.

Navigating the provisions is not a straightforward exercise; as Lord Judge CJ stated, 'the circumstances in which it is available do not exactly jump off the legislative page'.[61] Whilst the finer details of loss of control have yet fully to be tested, the case of *R v Clinton*[62] provides an indication of the way in which the courts may approach the plea.[63] *Clinton* was not concerned with mercy killing,[64] but two aspects of the Court of Appeal's judgment provide insight of general application: firstly, the court emphasised that the reforms should be viewed as standalone provisions, divorced from their historical context and the legacy of the moribund doctrine of provocation.[65] The court warned against any temptation to rely on the extensive case law that attended the erstwhile plea, stating that the provisions of the CJA 2009 relating to loss of control are 'self-contained'. Describing its 'common law heritage' as 'irrelevant' and a part of 'legal history', the court stressed that 'the full ambit of the defence is encompassed within [the new] statutory provisions'.[66]

The second aspect of *Clinton* that is of interest to my present argument is its treatment of legislative provisions that sought to narrow the availability of the partial defence by excluding 'sexual infidelity' as a qualifying trigger for loss of control. The court chooses to construe this exclusion restrictively and justifies this by referring to contextual factors affecting the gravity of the circumstances. For the court, seeking overly to 'compartmentalise' would be 'difficult' and 'unrealistic' and 'carr[y] with it the potential for injustice'.[67] This amounts to an expansive approach which effectively broadens the availability of loss of control; on this, Wake remarks: 'Judicial law-making appears to have triumphed over the provisions of the statute in this regard.'[68]

The approach taken in *Clinton*, juxtaposed with the inescapable similarities between provocation and loss of control, casts the authorities in an interesting light. Together, the pronouncements of the Court of Appeal and the potentially far-reaching reforms brought in by the CJA 2009 give the opportunity for the courts to depart radically from the law

59 Explanatory Note 347 states: 'Subsection (4) therefore sets a very high threshold for the circumstances in which a partial defence is available where a person loses self-control in response to words or actions. The effect is to substantially narrow the potential availability of a partial defence in cases where a loss of control is attributable to things done or said compared to the current partial defence of provocation (where no threshold exists in relation to the provoking circumstances).'

60 A curious statement of the government's intentions in reformulating the partial defences: 'What we therefore sought to do in respect of the change to a provocation defence is to raise the threshold generally, so that those who kill in anger can succeed in having their conviction reduced to manslaughter only in exceptional circumstances. So, we are raising the bar of the availability of that defence and extending it to cover those who kill in fear of serious violence as well as those who kill in anger.': Maria Eagle, Parliamentary Under-Secretary of State for Justice, Hansard Public Bill Committee, Tuesday 3 February 2009.

61 *R v Dawes, Hatter and Bowyer* [2013] EWCA Crim 322 [48] (Lord Judge CJ).

62 *R v Clinton and others* [2012] 1 Cr App R 26.

63 Ibid.

64 The Court of Appeal looked *inter alia* at the 'extent of the prohibition against "sexual infidelity" as a qualifying trigger for the purposes of the loss of control defence': ibid (Lord Judge CJ). Sexual infidelity is precluded by virtue of s 55(6)(c) of the CJA 2009.

65 The new law can be found in ss 54–55 of the CJA 2009.

66 *R v Clinton and others* [2012] 1 Cr App R 26, 364 (Lord Judge CJ).

67 Ibid.

68 See: Nicola Wake in Reed and Bohlander (n 8).

under provocation. And yet, Clinton also suggests that the appellate courts will act trenchantly in order to ensure that the partial defences are able to mitigate the inflexibility of the law of murder.

3 Possibilities for the mercy killer under loss of control

The structure of loss of control largely follows that of provocation, comprising factual and regulatory aspects. In what follows, I analyse the constituent elements of ss 54 and 55 of the CJA 2009, which detail the new plea, and point to the ambiguity they have created. Whilst there are strong indications that the legislators intended for the partial defence to be more narrowly construed than provocation, there is a lack of conceptual certainty in the demand for a loss of self-control and in the 'qualifying triggers' that were introduced ostensibly as a means by which to limit the applicability of the doctrine. Beyond this, there is also ample scope for an empathic response on the part of jurors, who are invited to take a holistic view of the circumstances of the case. These factors combine to make the plea more amenable to a defendant such as the mercy killer. Further, the means by which this is achieved, whilst not optimal, are arguably preferable to the situation as it existed under the benign conspiracy.

The factual question: what *is* a loss of control?

As was the case with provocation, and contrary to the advice of the Law Commission,[69] success under the new plea demands that the killing is as a result of the defendant's 'loss of self-control'. At first sight, this requirement presents a significant obstacle for the mercy killer,[70] as it may appear the very antithesis of their conduct. This point was made recently by the Court of Appeal in *Inglis*;[71] whilst this case was decided under the doctrine of provocation, the relevant part of the judgment is worth quoting at length:

> There is no doubt at all that the appellant was subjected to great stress and anguish, but dealing with it briefly and starkly, there was, as our analysis of the evidence underlines, *not a scintilla of evidence that when the appellant injected the fatal dose of heroin into her son she had lost her self-control.* Rather, it was to the contrary: all the evidence demonstrated that *the appellant applied her mind to her objective, which was to kill her son, and that she did so with scrupulous and meticulous care, and that in doing so she fulfilled her long-standing objective.* Of course, we accept that the appellant is a decent woman, of positive good character, and that acts of violence of any kind, let alone fatal or potentially fatal actions, were quite outside her normal character. However, in relation to her son and his injuries, she was resolved that she should relieve him of his suffering. *When she did so, she knew exactly what she was doing, and why she was doing it, and how it was to be done, and how it was imperative that its success should be assured. Far from lacking or losing self-control (an essential ingredient for the defence of provocation) the appellant was completely in control of herself.*[72]

The court here formulates a particular view of the exercise of self-control, characterising it as the ability to deliberate and to act in accordance with preconceived intentions; but this

69 Law Com No 304 (n 12) [5.11] and [5.20].

70 The requirements of loss of control are set out in s 54(1): 'Where a person ("D") kills or is a party to the killing of another ("V"), D is not to be convicted of murder if — (a) D's acts and omissions in doing or being a party to the killing resulted from D's loss of self-control; (b) the loss of self-control had a qualifying trigger, and; (c) a person of D's sex and age, with a normal degree of tolerance and self-restraint and in the circumstances of D, might have reacted in the same or in a similar way to D.'

71 [2010] EWCA Crim 2637.

72 Ibid [38] (Lord Judge CJ) [emphasis added]. For discussion of this, see: David Thomas 'Sentencing: Murder—Mercy Killing' [2011] Criminal Law Review 243, 244–45.

is not the only possible construction. What is meant by a 'loss of self-control', and the means by which this is measured, were debated in Parliament during the passage of the CJA 2009. Lord Thomas considered the term to be 'ambiguous because it could denote either a failure to exercise self-control or an inability to exercise self-control'.[73] Herring makes a broader point when he writes: 'An alternative vision of loss of self-control would see it as involving a lack of self-restraint, a lack of "moral check" over one's actions.'[74] In this way, a loss of self-control is seen in a social context, as lacking the ability to behave as society expects and demands.

The changes brought in during the transformation from provocation to loss of control reflect in large part a political will to redress the gender disbalance alluded to above, and this is addressed at various points in the legislation.[75] One of the principal steps taken in order to achieve parity, and accommodate the reality of a more typically female response, was the removal of the requirement that the loss of self-control be 'sudden'.[76] The Ministry of Justice explained that this 'would allow for situations where the defendant's reaction has been delayed or builds gradually'.[77] The practical importance of this change has been contested, insofar as it might be difficult to conceive of a loss of control that is not sudden.[78] However, such criticisms assume that the plea is structured around anger, and are thus rooted in a conventional conception of a loss of self-control (one that is challenged below). For Withey, the change has a much greater significance: 'Separating "loss of self-control" from the concept of "suddenness" and dispensing with the latter requirement could be explained by interpreting loss of self-control as a feeling of being unable to refrain from killing.'[79] Thus, the removal of the suddenness requirement allows for a more expansive conception of a loss of self-control and feeds the ambiguity to which Lord Thomas and Herring allude.

Should the fact of a loss of self-control be established, the defendant's conduct is then subjected to a number of regulatory requirements: it must be as a result of a 'qualifying trigger';[80] and it must be the case that 'a person of D's sex and age, with a normal degree of tolerance and self-restraint and in the circumstances of D, might have reacted in the same or in a similar way to D'.[81] As such, these regulatory considerations act to limit both the causes of the loss of self-control that will qualify for the plea, and to calibrate the reaction of the defendant against an expected societal standard.

---

73 HL Deb, 7 July 2009, vol 712, col 572.

74 Jonathan Herring, *Criminal Law* 6th edn (Palgrave Macmillan 2009) 160. See also R Holton and S Shute, 'Self-Control in Modern Provocation Defence' (2007) 27 Oxford Journal of Legal Studies 49.

75 For the most part, this has been effected through two elements of the plea: the implementation of the so-called 'fear trigger' and the removal of the requirement that the loss of control be sudden (contained in ss 55(3) and 54(2) respectively).

76 The deliberate nature of this absence is specifically referred to in s 54(2).

77 Ministry of Justice (n 49) [37]. Explaining the reason for resiling from the Law Commission's recommendation that the loss of self-control be done away with entirely, the MoJ said of its proposed change: 'We think this strikes the right balance between addressing the problems identified with the current law whilst not creating new ones.'

78 Norrie (n 16) is sceptical, suggesting that a loss of control will rarely be anything other than 'sudden' ; Ormerod makes a similar point, arguing that the removal of the requirement for 'suddenness' may make the defence 'difficult to get off the ground': (n 8) 511–12.

79 Withey (n 58).

80 S 54(1)(b).

81 S 54(1)(c).

Ambiguity and the qualifying triggers: a widening, a narrowing or a shift of focus?

A prominent means by which the CJA 2009 seeks to regulate the scope of loss of control is by specifying the causes that will qualify. In order to satisfy a case where the plea is raised, the defendant's conduct must be as a result of a 'qualifying trigger', found in s 55(3)–(5). The first of these provides that the defendant's loss of self-control must be 'attributable to D's fear of serious violence from V against D or another identified person'.[82] The second applies where the loss of self-control is 'attributable to a thing or things done or said (or both) which (a) constituted circumstances of an extremely grave character, and (b) caused D to have a justifiable sense of being seriously wronged'.[83]

The approach taken in delimiting the 'qualifying triggers' is deliberately and self-evidently exclusionary, insofar as it seeks to exclude any cause of the loss of control that falls outside of their purview. The 'fear trigger' laid out in s 55(3) is narrowly drawn, and in place to avail a defendant such as the abused spouse. The more interesting trigger for the present purpose, and the one that I will argue offers a greater possibility for the expansion of the loss of control plea, and its use by the mercy killer, is that found in s 55(4), a provision with an ambiguity that has thus far been underappreciated.[84]

Section 55(4) has been characterised as the 'anger trigger',[85] but there is no reference to anger in the loss of control provisions,[86] and to construe it as such is to view it, not unreasonably, through the lens of the erstwhile provocation.[87] However, even before the inception of the CJA 2009, commentators had pontificated on the possibility of a range of emotions that might qualify for consideration under the old provocation defence. On this, Herring observed: ' . . . there is nothing in s 3 which would limit the defence to those who lose their self-control through anger'.[88] Of the mercy killing case of *Cocker*, he writes: 'the alleged loss of self-control was through compassion, but this was not mentioned as a reason why the defence was unsuccessful'.[89]

Whether or not the plea of provocation was centred on a response born of anger, changes brought in by the CJA 2009 may facilitate a shift to a broader view of the qualifying emotions. In what follows, I will demonstrate the extent to which s 55(4) can be decoupled from a narrow anger requirement, and thus overcome one of the perceived obstacles to the success of loss of control claims in mercy-killing cases. In light of this, I shall not refer to the trigger under s 55(4) as the 'anger trigger', but rather as the 'seriously wronged trigger', so as to make room for a more expansive conception of its scope.

Under the reformed plea, the test for establishing the seriously wronged trigger comprises four elements, requiring: (1) 'a thing or things done or said (or both)'; (2) which constitute 'circumstances of an extremely grave character'; (3) that this has caused the

82 S 55(3).

83 S 55(4). S 55(5) allows for a combination of ss 55(3) and 55(4).

84 Since the inception of the CJA 2009, and indeed dating back to the Law Commission proposals that inspired them, the majority of attention has been focused on the first of the qualifying triggers that may legitimately have caused the 'loss of self-control'.

85 For example, Edwards refers to it as 'the anger defence' (S Edwards, 'Loss of Self-Control: When His Anger is Worth More than Her Fear' in Reed and Bohlander (n 8).

86 This is true both of the legislation and its accompanying Explanatory Notes.

87 In restricting his discussion to fear and anger, Norrie notes the absence of the word 'anger' but makes the reasonable point that 'anger or something like it—lost temper, outrage—has always been the link between the provoking conduct and the loss of control': Norrie (n 16).

88 Herring (n 74) 160.

89 Ibid 160.

defendant to have a 'justifiable sense of being seriously wronged'; and (4) that the defendant's loss of self-control is 'attributable' to the above. Section 55(4) is a direct descendant of the old provocation doctrine, and also draws heavily on the recommendations for reform contained in the work of the Law Commission.

The first of the requirements under the seriously wronged trigger is that the conduct of the defendant must be as a result of a 'thing or things done or said (or both)'. The legacy of provocation is clear in this capacious requirement,[90] which the Law Commission considered had come to mean 'no more than "caused"'.[91] Writing whether such a requirement might be satisfied in the case of the mercy killer, Taylor is unequivocal: 'Surely the answer must be "yes." The illness of the "victim" often manifests itself in numerous acts, often seen by the defendant as being humiliating for the victim. This is further supported if requests are made by the victim to have his/her life ended'.[92] Should the circumstances of a mercy killing be held to amount to 'things said or done', the second demand of s 55(4) is that this must 'constitute circumstances of an extremely grave character'. Whilst this has a narrowing effect, it is difficult to imagine a situation involving a mercy killing in which this would not be the case.[93] Insofar as this stands to be appraised by a jury, it will be examined in further detail below.

The third element (a 'justifiable sense of being seriously wronged') amounts to a dual requirement: that the defendant has been 'wronged' and that this sense of being wronged is 'justifiable'. In setting this out, the statute gives no further detail. The ambiguity, moreover, is carried over into the Explanatory Notes.[94] There is much semantic argument that could be engaged in exploring the exact meaning of 'wronged', and the extent to which this could correlate to the circumstances surrounding a mercy killing. Insofar as the requirement is to be applied to the case of the mercy killer, an intuitive response may be that it cannot, for how is this to be interpreted other than by reference to an aggressor and a person who feels slighted, either by their attitude or actions towards them? A defence that was constrained in such a way would offer little by way of an avenue for the mercy killer, but I suggest that this narrow interpretation is not the only explanation that could find space within this concept. In essence, I suggest that a person can have a sense of being wronged as a result of circumstance, or as a result of decisions taken by people or bodies, and that this can both satisfy the broad requirement of 'things said or done' and induce a reaction that will fulfil the loss of control provisions.

By way of judicial support for this contention, a brief survey of recent appellate cases reveals situations in which the courts have described those who have not been afforded equal treatment under anti-discrimination legislation as having been 'wronged', and reveals

90 S 3 of the Homicide Act 1957 required that 'the person charged was provoked (whether by things done or by things said or by both together)'.

91 Law Com No 290 (n 10) [3.25]. In provocation, it had been established that the things done and/or things said did not need to have emanated from the victim (*R v Davies* [1975] QB 691), nor to have been directed at the defendant (*R v Pearson* [1992] Criminal Law Review 193; *R v Baillie* [1995] Criminal Law Review 739).

92 Paul Taylor, 'Provocation and Mercy Killing' [1991] Criminal Law Review 111.

93 If it could be said that a purported mercy killing was taking place absent 'circumstances of an extremely grave character', it is surely the case that either this is not a mercy killing, or that the alternative partial defence of diminished responsibility, or even a plea of insanity, would be more appropriate.

94 Explanatory Note 346 states: 'Subsection (4) sets out when a thing or things done or said (or both) can amount to a qualifying trigger for the loss of self-control. The thing(s) done or said must amount to circumstances of an extremely grave character and cause the defendant to have a justifiable sense of being seriously wronged. Whether a defendant's sense of being seriously wronged is justifiable will be an objective question for a jury to determine (assuming that there is sufficient evidence for the defence to be left to the jury).'

an assortment of entities and sources from which this may have emanated.[95] Further examples include the wrongfully convicted; could it not be argued that the wrongful imprisonment of the Birmingham Six led to their feeling 'wronged'?[96] Consider further the situation of a parent whose child is denied treatment for a rare condition, the treatment for which is expensive and beyond the scope of the medical authorities; could they not also feel 'wronged'? If such grounds for the sense of feeling wronged are accepted, why not the wrong that is felt when a loved one is suffering and wishes to die?[97] In all of these situations, it may or may not be possible to point to a person or persons who have brought about the emotion, but such considerations may do little to alter the strength or validity of the emotion experienced, and do not appear to be precluded by the wording of the Act. Should the 'sense of being seriously wronged' extend effectively to cover the invidious situation of the mercy killer, this must also be 'justifiable'. This requirement amounts to one of a number of 'objective' regulatory controls on the plea, to which I will now turn.

Objective calibration: the regulation of loss of control

The exclusionary effect of the qualifying triggers is one means by which the loss of control plea can be regulated. Allied to this, the legislation also precludes use of the plea where the defendant had 'acted in a considered desire for revenge',[98] or had behaved in a way so as to provide 'an excuse to use violence',[99] or where the trigger 'constituted sexual infidelity'.[100] The legislation does not expressly preclude mercy killing.[101]

These specific exclusions are buttressed by a number of tests that aim to calibrate the response of the defendant according to societal standards and which are woven into the provisions at various points. The first two of these occur in s 55(4) and were alluded to above: the requirements that the 'thing or things done or said . . . *constituted circumstances of an extremely grave character*';[102] and that the sense of being seriously wronged is 'justifiable'.[103] Of the latter requirement, the Explanatory Notes to the CJA 2009 stipulate that this is to be 'an objective question for a jury to determine'.[104] Whilst it is not specified of the former requirement, it is submitted that this must also amount to an objective appraisal on the part of the jury.

95 *Kelly v National University of Ireland (University College, Dublin)* (Case C-104/10) [2011] 3 CMLR 36; *Tariq v Home Office Supreme Court* [2011] UKSC 35, [2011] 3 WLR 322; *Donna Marie Arrowsmith v Nottingham Trent University* Case No A2/2010/2289 [2011] EWCA Civ 797 2011 WL 2039891 ('She believes she has been seriously wronged by Nottingham.' [28]); *The Queen on the Application of Mjemer v Secretary of State for the Home Department* CO/13206/2010 [2011] EWHC 1514 (Admin) 2011 WL 1151605 (in a report referred to in this case, it is said that individuals may feel 'consistently wronged by the system' [130]); *R (on the application of Tate & Lyle Industries Ltd) v Secretary of State for Energy and Climate Change* Court of Appeal (Civil Division) 3 June 2011 [2011] EWCA Civ 664, [2011] ACD 92 ('wronged' by the actions of a public body (279–80)).

96 <www.independent.co.uk/news/bond-that-unites-two-wronged-men-men-1164526.html>; <www.independent.co.uk/opinion/leading-article-wronged-people-still-in-the-dock-1371666.html> accessed 10 February 2014.

97 The recent case of Kay Gilderdale provides a good example in this respect <www.solicitorsjournal.com/news/public/care/prosecuting-mother-attempted-murder-public-interest-dpp-insists> accessed 10 February 2014.

98 S 54(4).

99 S 55(6)(a) and (b).

100 S 55(6)(c). See: *R v Clinton and others* [2012] 1 Cr App R 26.

101 This 'oversight' is presumably born of an unwillingness to acknowledge the possible relevance of the practice, or an assumption that the provisions exclude it.

102 S 55(4)(a) [emphasis added].

103 S 55(4)(b).

104 Explanatory Note 346. This is subject to the following caveat: 'assuming that there is sufficient evidence for the defence to be left to the jury'.

Supplementing these requirements is the demand that, in the face of the seriously wronged trigger, a person of a 'normal degree of tolerance and self-restraint . . . might have reacted in the same way' as the defendant.[105] This ostensibly objective test is tempered by some concessions to the particularities of the defendant, insofar as the loss of control should be viewed according to what might be expected of a 'person of D's sex and age' and 'in the circumstances of D'.

In a piece that is strongly condemnatory of the judgment in *Clinton*,[106] Baker and Zhao express the view that the loss of control plea should be construed narrowly, drawing parallels between its inception and the complete abolition of provocation in New Zealand. They assert that the overlapping objective tests should be seen as constrictive, and could be summarised as amounting to a straightforward question for the jury: 'Was it reasonable for the defendant to lose control and kill V?'[107] This analysis is problematic; as Fletcher famously wrote, 'the reasonable person does not kill at all, even under provocation'.[108] In *R v Smith*, Lord Millett specifically warned against using this standard:

> [I]t can never be reasonable to react to provocation by killing the person responsible. Nor by pleading provocation does the accused claim to have acted reasonably. His case is that he acted unreasonably but only because he was provoked. But while this may not be reasonable it may be understandable, for even normally reasonable people may lose their self-control and react unreasonably if sufficiently provoked.[109]

In suggesting that the defendant's behaviour should be judged against a standard of reasonableness, Baker and Zhao misunderstand the basis of the plea. Where they are more likely to be correct, however, is in their appraisal of the overlap between the different regulatory aspects of loss of control outlined above. These combine to reinforce the idea that a *prima facie* instance of a loss of self-control as a result of a qualifying trigger should be submitted to further control, comprising a moral question for the jury to assess by reference to 'contemporary society's norms and values . . . the normative standards of a normal person communally situated in Britain'.[110]

Any plea of loss of control may founder on judicial discretion, as s 54(6) implements an important recommendation by the Law Commission, namely that the judge should instruct the jury as to whether there is evidence for the defence to stand and to be considered. As Norrie points out, this 'empower[s] judges to make the moral and political call',[111] but the significance of this potentially onerous requirement is easily overstated. In *R v Dawes, Hatter and Bowyer* and in *Clinton*, the Court of Appeal took a permissive approach to the provision and addressed s 54(6) as primarily a question of evidential sufficiency.[112] On a conceptual level, it could be said that the test is somewhat redundant. The judge is asked to leave to the jury a determination of loss of control where 'a jury, properly directed,

105 S 54(1)(c).

106 *R v Clinton and others* [2012] 1 Cr App R 26.

107 Dennis Baker and Lucy Zhao, 'Contributory Qualifying and Non-Qualifying Triggers in the Loss of Control Defence: A Wrong Turn on Sexual Infidelity' (2012) 76 Journal of Criminal Law 254.

108 George P Fletcher, *Rethinking Criminal Law* (Little, Brown & Co 1978) 247.

109 [2000] 3 WLR 654, HL, 712.

110 Baker and Zhao (n 107).

111 Norrie (n 16).

112 In *R v Dawes, Hatter and Bowyer* [2013] EWCA Crim 322 [53], Lord Judge CJ stated: 'This requires a commonsense judgment based on an analysis of all the evidence . . . the judge should not reject disputed evidence which the jury might choose to believe.' See also: *R v Clinton and others* [2012] 1 Cr App R 26 [45] and [46].

could reasonably conclude that the defence might apply'. Whilst this should rule out fanciful claims, it is of limited assistance insofar as it is tautologically dependent upon the scope of the provisions to which it applies. When it comes to trial judges, the impact of this provision is of course difficult to predict, but their role in the effective operation of the benign conspiracy described above should be borne in mind.[113]

A feature of the loss of control provisions that might give the opportunity for the sympathetic jury to allow the plea lies in its asking whether 'a person of D's sex and age, with a normal degree of tolerance and self-restraint *might have* reacted in the same or in a similar way to D'.[114] The question of whether to use the word 'might' in this clause was one upon which the Law Commission equivocated, ruling out the word 'would' on the grounds that a test that incorporated the term would be 'near to impossible' to satisfy. Withey concludes that the use of 'might' has brought a potential expansion, as it makes the test 'easier to satisfy than before': 'Even if most other people would not have reacted in the same or a similar way, if D's reaction was *within* the response range of other people, unusual and disproportionate responses can be accommodated.'[115]

In light of the above, it is somewhat strange that Baker and Zhao downplay the importance of the concessions to subjectivity in the loss of control plea; they assert: 'Considering the defendant's personal circumstances is not likely to add much, but such a consideration might tip the balance in a borderline case.'[116] Since they concede the importance of the jury's normative view of the defendant's conduct, an invitation to appraise that against a context that may make an enormous difference to how that conduct is perceived is surely of far greater significance than Baker and Zhao entertain. Writing in response to proposals to repeal the partial defence of provocation in New Zealand, Tolmie cites its successful use by a man who 'snapped in response to his mother begging for relief from the pain of the final stages of her terminal bowel cancer and sped up her eventual death';[117] Tolmie points to the availability of provocation as 'a concession to human frailty—a recognition that everyone, if the circumstances were extreme enough, could potentially snap and kill'.[118] Where the jury is allowed to factor in a consideration of the 'personal circumstances' of the defendant, this may prove a powerful factor in considering the extremely grave character of the circumstances, the justifiability of the sense of being seriously wronged, and whether a whether a person of a 'normal degree of tolerance and self-restraint . . . might have reacted in the same way' as the defendant.[119] Writing of *Cocker*,[120] in which the erstwhile plea of provocation was unsuccessfully pleaded, Taylor paints a plausible picture of a jury's probable reasoning when confronted with a mercy-killing case:

113 Though it may not be representative of the views of all trial judges, in such a case, Judge Alan Goldsack, QC, is reported to have said: 'My professional life is spent dealing with criminals. I do not recognise you as a criminal. You pose no risk to anyone. You are not in need of treatment. Despite what you thought on New Year's Day, you still have an important role to play as father and grandfather within your extended family. You should be allowed to get on with it.' (C Brooke, 'Mercy for Husband who Killed Alzheimer's wife "Neglected by Hospital"' *The Daily Mail*, 27 April 2006, 29).

114 S 54(1)(c) [emphasis added].

115 Withey (n 58) 278.

116 Baker and Zhao (n 107) 271.

117 J Tolmie, 'Defence of provocation has its place', *New Zealand Herald*, 7 September 2009 <www.nzherald.co.nz/justice-system/news/article.cfm?c_id=240&objectid=10595632> accessed 10 February 2014. The partial defence of provocation (s 169 of the Crime Act 1961) was subsequently repealed, on 8 December 2009, by s 4 of the Crimes (Provocation Repeal) Amendment Act 2009 (2009 No 64).

118 Tolmie (n 117).

119 S 54(1)(c).

120 [1989] Criminal Law Review 740.

> The practicalities of mercy killing cases would probably mean that quite simply the jury, if sufficiently sympathetic to the defendant, would find that he acted as a reasonable man would in the circumstances, and if they are unsympathetic they would not. Much will depend on the facts of the individual case and the last sentence does not necessarily imply that the jury will apply the provocation test incorrectly; it simply means that, as with most cases, it will not be the only thing that the jury take into account in reaching their verdict.[121]

Conclusion

The legal response to mercy killing is of greater societal importance than can be measured by the quantity of cases appearing before the courts; that this piece has not sought to engage more fully with normative moral, political and social questions reflects the disengaged, yet perversely straightforward, status of mercy killing under the criminal law. Here, it occupies an unresolved space, formally classified as murder and yet rarely prosecuted as such. This invisibility to the substantive law is largely the result of a lack of political engagement that forces the criminal justice system into an improvised response, and the facilitation of the benign conspiracy is symptomatic of the disjunction between the substantive law and the social practice of mercy killing. As Huxtable observes, neither murder nor the application of diminished responsibility 'quite captures the lived, and legal, reality of mercy killing',[122] and both comprise unsatisfactory responses. It is difficult to disagree with Ashworth's indictment of the legislative stasis that attends the issue: 'The unwillingness of governments to confront this category of cases leaves the law in a state of impoverishment and duplicity.'[123]

It may be that the advent of the CJA 2009 forces a change to the prevailing response; the requirement for a 'recognised medical condition', buttressed by the causal requirement also introduced by s 52 of the CJA 2009, may foreclose the use of diminished responsibility. Insofar as this model is unsatisfactory, this is to be welcomed, but it would bring injustice if no other means were available to obviate the often grossly disproportionate effect of a murder conviction. The need for an alternative should prompt a closer examination of the newly introduced plea of loss of control, which might be found to have broader application than has thus far been appreciated.

In asserting the availability of loss of control, there are three hurdles to negotiate, subject to the leave of the judge: firstly, establishing a loss of control; secondly, establishing that the particular circumstances satisfy the qualifying trigger; and, thirdly, the regulatory demands, which appeal to social standards of conduct. Whilst most commentators have looked to its roots in provocation, and the ostensible intentions of the legislature in enacting the loss of control provisions, I have sought to decouple them from the political background and the historical basis of the plea. In so doing, I have shown that both the idea of a loss of self-control and the relevant qualifying trigger under the CJA 2009 can be construed so as to render loss of control potentially broader in effect than its predecessor, and that it may avail the mercy killer.

That is not to advocate an expansive view of loss of control as the best way forward for society or the criminal law. Keating and Bridgeman write of the current state of the partial defences: 'As matters stand this is not a blueprint for a law which responds

121 Taylor (n 92).

122 Huxtable (n 4) 53.

123 Andrew Ashworth 'Sentencing: Murder—Mercy Killing' [2011] Criminal Law Review 243, 248.

appropriately to compassionate killing.'[124] It is hard to disagree with this point; mercy killing should be considered openly as part of a broader consideration of the important moral, social and legal questions that pertain to issues at the end of life. The adoption of loss of control, which necessitates examination of the defendant's actions against societal standards, may increase the politico-legal visibility of the subject and promote an open debate that is not facilitated by ad hoc and shadowy arrangements between legal and medical practitioners. In the meantime, the partial defences exist to ameliorate the harsher aspects of the law of murder; to exculpate in part those who are deemed less worthy of sanction. Until the subject can be addressed in a more honest and thorough way, their use can and should include those who have committed the sort of 'almost venial' homicide that may be personified by mercy killing.[125]

124 Heather Keating and Jo Bridgeman, 'Compassionate Killings: The Case for a Partial Defence' (2012) 75(5) Modern Law Review 697, 711.

125 *R v Howe* [1987] 1 AC 417, 433 (Lord Hailsham).

NILQ 65(2): 205–14

Insanity and automatism: notes from over the border and across the boundary

James Chalmers*

University of Glasgow

Introduction

The English Law Commission's recent discussion paper on insanity and automatism[1] follows only a few years after the defence of insanity was replaced with a plea of mental disorder excluding criminal responsibility in Scots law, which itself followed from work by the Scottish Law Commission. The proximity of two law reform projects in this area, in jurisdictions which have taken a similar but not identical approach to this area of the law in the past, offers a useful opportunity for contrasts to be drawn. As there is surprisingly little written on the Scottish law of insanity and automatism,[2] this paper begins with an account of the development of these two defences in Scots law, demonstrating their close relationship with the English McNaghten Rules, before drawing comparisons with the English Law Commission's proposals, raising concern in particular about the proposed boundary between automatism and the proposed new defence of 'not criminally responsible by reason of recognised medical condition'.

The development of the Scots law of insanity: a brief history

The relationship between the Scottish and English laws of insanity is a curious one, alternating between acceptance and rejection. There are no significant reported cases on insanity in Scots law prior to the McNaghten Rules,[3] and in the first significant post-McNaghten case in Scotland, the trial judge quoted the rules in his charge to the jury.[4] In the same set of reports in which that charge appears, the editor saw fit to insert the Rules as an appendix given that they were of 'such general importance'.[5]

* I am grateful for the support of the Leverhulme Trust.

1 Law Commission, *Criminal Liability: Insanity and Automatism* Discussion Paper (Law Com DP 2013).

2 Aside from detailed discussion in two textbooks: see G H Gordon, *The Criminal Law of Scotland* 3rd edn by M G A Christie, vol 1 (W Green 2000) ch 10; James Chalmers and Fiona Leverick, *Criminal Defences and Pleas in Bar of Trial* (W Green 2006) ch 7.

3 The closest is *Eugene Whelps* (1842) 1 Broun 278, but this is no more than a brief account of a charge to a jury.

4 *James Gibson* (1844) 2 Broun 332, where Lord Justice-Clerk Hope said (355) that the rules 'expresse[d] the law of Scotland, as well as of England, upon the matter'.

5 See the editor's footnote to *James Gibson* ibid, 355.

There is nothing especially surprising about this. The older Scottish writers had formulated their discussions of insanity in terms which were strikingly similar to the rules.[6] The Rules themselves added little to Scots law, but as a contemporary source, they were of use to judges who wished to repel contentions that modern medical knowledge should lead to a change in the courts' approach to the defence.[7]

The courts were not wholly blind to medical writing, and some nineteenth-century judges did pay heed to writers who argued that cognitive approaches to insanity were unduly narrow,[8] directing juries that they could find insanity established if the accused knew that his actions were wrong but had acted under an irresistible impulse.[9] Such judges found themselves in a minority,[10] although the absence of any appeal court prior to 1926[11] meant that there was no opportunity for the divergence of judicial opinion to be addressed.

Even after the creation of that court, reported decisions on insanity invariably came in the form of charges to juries by trial judges. Most of the cases reported in the late nineteenth century were the work of one judge, Lord Moncrieff – 'an eccentric in this branch of the law'[12] – and involved almost abandoning any attempt at definition whatsoever, leaving it to the jury to decide simply whether or not an accused was 'responsible'[13] but without meaningful guidance on how to approach this task. One result of this was an increasing inconsistency in the approach of judges to charging juries,[14] so that when the Royal Commission on Capital Punishment carried out its work between 1949 and 1953, it was difficult for the judicial witnesses to the Royal Commission to say 'just what the Scots law on insanity was',[15] although with a sense of where the wind might be blowing, the most senior Scottish judge (Lord Justice-General Cooper) downplayed the extent to which the McNaghten Rules might be regarded as part of Scots law.[16]

When, therefore, the Royal Commission reported that the McNaghten Rules could not be 'defended in the light of modern medical knowledge and modern penal views',[17] it was easy for Scots judges to adjust course and proclaim that they formed no part of Scots law.[18] In *Brennan v HM Advocate* in 1977, the appeal court finally produced an authoritative statement

6 A Alison, *Principles of the Criminal Law of Scotland* (1833) 645; D Hume, *Commentaries on the Law of Scotland, Respecting Crimes* 4th edn by B R Bell (1844) vol i, 37. Although the fourth edition of Hume's text was published after the McNaghten Rules were formulated, the text remained unchanged from previous editions; the editor (Bell) had restricted his work to adding supplemental notes on recent cases.

7 In *Gibson* (n 4), defence counsel had referred to Issac Ray's *A Treatise on the Medical Jurisprudence of Insanity* (1838). Lord Justice-Clerk Hope instructed the jury (356) to disregard Ray's 'fantastic and shadowy definitions'.

8 See e.g. G F Blandford, *Insanity and its Treatment* (1884) 360.

9 *James Denny Scott* (1853) 1 Irv 132; *John McFadyen* (1860) 3 Irv 650. See also *Isabella Blyth* (1852) J Shaw 567.

10 See Chalmers and Leverick (n 2) para 7.13.

11 Criminal Appeal (Scotland) Act 1926.

12 Gordon (n 2) para 10.35.

13 *Archibald Miller* (1874) 3 Coup 16, 17. See further Chalmers and Leverick, (n 2) paras 7.14–15.

14 Chalmers and Leverick (n 2) para 7.17.

15 Gordon (n 2) para 10.39.

16 *Minutes of Evidence Taken Before the Royal Commission on Capital Punishment* (1950) para 5465. For criticism of Lord Cooper's evidence, see Chalmers and Leverick (n 2) para 7.10.

17 *Report of the Royal Commission on Capital Punishment* (Cmd 8932 1953) para 291.

18 See, for example, *Mackenzie v Mackenzie* 1960 SC 322, 325 (Lord Walker); *HM Advocate v Kidd* 1960 JC 61, 71 (Lord Strachan); *Breen v Breen* 1961 SC 158, 185 (Lord Patrick); *Brennan v HM Advocate* 1977 JC 38, 46 (Lord Justice-General Emslie).

on what was meant by insanity in Scots law: 'total alienation of reason in relation to the act charged as the result of mental illness, mental disease or defect or unsoundness of mind'.[19]

Automatism's emergence in Scots law

Despite this McNaghten scepticism, the relatively recent recognition of automatism as a defence in Scots law brought McNaghten back to the fore insofar as the insanity plea was concerned. The Scots law of automatism is sometimes traced back to the 1926 decision in *HM Advocate v Ritchie*,[20] where a motorist who hit and killed a pedestrian was charged with culpable homicide, and pled not guilty on the basis of 'temporary mental dissociation due to toxic exhaustive factors'. The judge allowed this defence to go to the jury, who acquitted Ritchie. In the late twentieth century, the case was almost invariably read as one where a driver had been overcome by car exhaust fumes, and therefore an obvious candidate for an automatism plea.[21] In fact, that was not what had happened at all – the 'toxic exhaustive factors' were poison entering Ritchie's blood from an abscess in his lungs[22] – but the misreading of *Ritchie* provided a false foundation for the modern recognition of the automatism defence, meaning that identifying its origins in 1926 is simultaneously both erroneous and correct.

Building on *Ritchie*, the defence of automatism was formally recognised in the 1991 case of *Ross v HM Advocate*,[23] where the accused's drink had been 'spiked' with temazepam and LSD, causing him to behave uncontrollably and violently. The court identified three requirements for the defence: first, there must be an external factor; secondly, it must be neither self-induced nor something which the accused was bound to foresee; thirdly, it must result in a total alienation of reason.[24]

The use of 'total alienation of reason' as one of the criteria for the defence aligns it directly with insanity, something which is made clear in the parallel drawn between the two defences by the court.[25] In Scots law, therefore, the two common law defences are not simply closely related but two sides of the same coin. Both are based on a total alienation of reason, and are distinguished by reference to the cause of that alienation, with the further caveat that the automatism defence is unavailable where the alienation has been caused by the prior fault of the accused.

Subsequently, the appeal court offered further guidance on the meaning of 'total alienation of reason' in *Cardle v Mulrainey*,[26] where an accused whose drink had been spiked with amphetamine attempted to steal a number of motor cars. Mulrainey's defence of automatism – on the basis that he had been aware of his actions and of their wrongful nature, but that he had been unable to stop himself committing them – was rejected by the court, which held that a 'total alienation of reason' was not made out where an accused 'knew what he was doing and was aware of the nature and quality of his acts and that what he was doing was wrong'.[27] Because the concept of total alienation of reason is core to

19 1977 JC 38, 45 (Lord Justice-General Emslie).

20 1926 JC 45.

21 See, for example, *Ross v HM Advocate* 1991 JC 210, 215 (Lord Justice-General Hope); Gordon (n 2) para 3.20.

22 Chalmers and Leverick (n 2) para 7.34; J Ross, 'A Long Motor Run on a Dark Night: Reconstructing *HM Advocate v Ritchie*' (2010) 18 Edinburgh Law Review 193.

23 1991 JC 210, overruling *HM Advocate v Cunningham* 1963 JC 80, which had denied the existence of any such defence.

24 See *Ross* (n 21) 222 (Lord Justice-General Hope).

25 Ibid 213–14 (Lord Justice-General Hope).

26 1992 SLT 1152.

27 *Cardle v Mulrainey* 1160 (Lord Justice-General Hope).

both insanity and automatism, this language, which is practically identical to that found in the McNaghten Rules, effectively (re)incorporated the core of the rules into the Scottish law of insanity.

Statutory reform of the Scots law of insanity

The Scottish law of insanity was reformed in 2010,[28] following on from a review by the Scottish Law Commission.[29] The relevant Scottish project was concerned only with insanity and diminished responsibility, in contrast to the English Law Commission's recent work. This is unfortunate: as the English Law Commission observed, the defence of automatism 'is so closely related to that of insanity that reform of one entails reform of the other'.[30]

As the above account of Scots law makes clear, that is – or should be – *a fortiori* the case in Scots law, given the interrelationship between the two defences. Sadly, that was not recognised as part of the Scottish reform project. It had originated in the work of the Millan Committee, which had reviewed mental health legislation in Scotland.[31] In its report, the committee had noted dissatisfaction and difficulty amongst psychiatrists who had to consider the legal tests applicable where insanity or diminished responsibility was raised.[32] Insanity 'depend[ed] on terms and definitions which [were] largely meaningless to those with the responsibility of giving expert evidence to the court',[33] while the definition of diminished responsibility was 'obscure, and difficult to apply in individual cases'.[34] Accordingly, the committee recommended that the matter be referred to the Scottish Law Commission, and this was taken up. Although it is unsurprising that psychiatrists did not recognise the relationship between the defences of insanity and automatism, it is regrettable that this was not identified at a subsequent point in the law reform process.

Before the issue was even formally referred to the Scottish Law Commission, the practical difficulties with the defence of diminished responsibility were addressed in the case of *Galbraith v HM Advocate*.[35] *Galbraith* largely aligned the Scottish law of diminished responsibility with the statutory definition then found in English law.[36] While the appeal court was far from explicit about this parallel,[37] it was hardly objectionable given that the English statute had been drafted specifically to 'introduce into English law the Scottish doctrine of diminished responsibility'.[38] In the light of *Galbraith*, the Scottish Law Commission's proposals – subsequently enacted in legislation[39] – were essentially a statutory restatement of the common law as formulated in that case.[40]

28 Criminal Justice and Licensing (Scotland) Act 2010, s 168.

29 Scottish Law Commission, *Report on Insanity and Diminished Responsibility* (Scot Law Com No 195 2004).

30 Law Com DP (n 1) para 1.29.

31 Scottish Executive, *New Directions: Review of the Mental Health (Scotland) Act 1984* (SE/2001/56 2001).

32 Also, and perhaps particularly, in relation to unfitness to plead: ibid para 29.29.

33 Ibid para 29.43.

34 Ibid para 29.55.

35 2002 JC 1. The judgment in *Galbraith* was issued in July 2001; the reference to the Scottish Law Commission was made in October of the same year. See Scot Law Com No 195 (n 29) para 1.1.

36 Homicide Act 1957, s 2, subsequently amended by the Coroners and Justice Act 2009, s 52.

37 On which see James Chalmers, 'Abnormality and Anglicisation: First Thoughts on *Galbraith v HM Advocate (No 2)*' (2002) 6 Edinburgh Law Review 108.

38 HC Debs 27 November 1956, col 318 (statement of the Attorney-General). See also the Home Secretary's statement at HC Debs 15 November 1956, col 1153.

39 Criminal Procedure (Scotland) Act 1995, s 51B, inserted by the Criminal Justice and Licensing (Scotland) Act 2010, s 168.

40 Gerry Maher, 'The New Mental Condition Defences: Some Comments' 2013 SLT (News) 1, 3. Maher was formerly the commissioner responsible for this law reform project.

Nor were the Scottish Law Commission's proposals on insanity particularly radical. Gone was the earlier Scottish scepticism about McNaghten; the Scottish Law Commission was impressed by Finbarr McAuley's remark that if the McNaghten Rules did not exist, 'it would be necessary to invent something like them'.[41] The test suggested by the Scottish Law Commission, and subsequently enacted by the Scottish Parliament, is concise: a person is not criminally responsible for their conduct 'if the person was at the time of the conduct unable by reason of mental disorder to appreciate the nature or wrongfulness of the conduct'.[42] Although this abandons the terminology of 'total alienation of reason', it simply replaces that phrase with the definition offered in *Cardle v Mulrainey*[43] and represents the most recent stage in McNaghten's rise, fall, and rise in Scots law.

This does not mean that Scots and English law are identical. There is one particularly important distinction, which is that the approach to 'wrongfulness' taken in the two jurisdictions differs. The English courts have held that 'wrongfulness' in this context means *legally* wrong, so that a man who knows his actions are prohibited legally but, because of mental disorder, believes them to be morally justified, has no recourse to the plea.[44] The Scottish courts have previously accepted that the plea can be based on a failure to appreciate moral wrongfulness, allowing a plea of insanity in a case where a mentally ill man had 'formed the idea' that he had to kill the two youngest members of his family to 'ease the burden' on his wife, regarding this as a 'solemn sacrifice which he was called upon to make' despite appreciating the legal penalty which would follow from it.[45] The Scottish approach is consistent with the view taken in other jurisdictions,[46] and the Scottish Law Commission recommended that it should continue.[47] This broader approach to wrongfulness now also has the support of the English Law Commission.[48]

Although the legislation following on from the Scottish Law Commission's work is in many respects a restatement of the common law position, there is one particularly important if inelegantly executed change. The defence is no longer called insanity. This is welcome: the term 'insanity' has long been regarded as outdated and inappropriate.[49] What is unfortunate is that the Scottish Law Commission was able to decide only what the defence should *not* be called. Having rejected the name 'mental disorder' as confusing and inaccurate, it decided it was sufficient that the relevant section be headed 'Criminal responsibility of persons with mental disorder'.[50] Regrettably, this has led to the defence being referred to as simply 'mental disorder',[51] despite the Scottish Law Commission rightly identifying that as inadequate. A formulation such as 'mental disorder excluding criminal responsibility', while perhaps slightly cumbersome, would have avoided the difficulties identified by the Scottish Law Commission.

41 See Scot Law Com No 195 (n 29) para 2.44, quoting F McAuley, *Insanity, Psychiatry and Criminal Responsibility* (Round Hall Press 1993) 25.

42 Scot Law Com No 195 (n 29) 78 (s 1(1) of Draft Bill); Criminal Procedure (Scotland) Act 1995, s 51A, inserted by the Criminal Justice and Licensing (Scotland) Act 2010, s 168.

43 1992 SLT 1152, discussed at text to n 26 above.

44 *R v Windle* [1952] 2 QB 826.

45 *HM Advocate v Sharp* 1927 JC 66, 69 (Lord Constable).

46 See, for example, *Stapleton v R* (1952) 86 CLR 358; *R v Chaulk* [1990] 3 SCR 1303.

47 Scot Law Com No 195 (n 29) paras 2.46–.51. The Scottish Law Commission makes no reference to *Sharp*, the significance of which has often been overlooked. See Chalmers and Leverick (n 2) para 7.41.

48 Law Com DP (n 1) para 4.33.

49 See e.g. *Report of the Committee on Mentally Abnormal Offenders* (Cmnd 6244 1975) para 18.18.

50 Scot Law Com No 195 (n 29) paras 2.21–23. The Commissioner responsible has since described the failure to name the defence as 'unfortunate': Maher (n 40) 1.

51 See e.g. T H Jones and M G A Christie, *Criminal Law* 5th edn (W Green 2012) para 8.11.

Some other consequences of the Scottish Law Commission's review are worthy of note. First, the legislation now expressly provides that the plea cannot be based on psychopathy,[52] something which was probably already true at common law but had not been clearly established.[53] Secondly, the defence can only be raised by the person charged,[54] contrary to the common law position.[55] That is unfortunate. Where an accused's mental disorder is such as to exclude *mens rea*, the Scottish position now seems to be, at least in theory, that it is open to him to deny *mens rea* on the basis of that condition without pleading insanity and receive an unqualified acquittal.[56] The English Law Commission's proposal that the prosecution should have a limited power to itself raise the defence in such cases is surely preferable.[57]

Thirdly, while the Scottish Law Commission initially proposed that the common law rule placing the burden of proof on the accused should be reversed,[58] so that the Crown would bear the burden of disproving a mental disorder-based defence beyond reasonable doubt if it were raised, it was persuaded to abandon this suggestion. This proposal, particularly the Scottish Law Commission's basis for it, was strongly criticised,[59] and the Scottish Law Commission changed its position, accepting that the reverse burden was not contrary to the European Convention on Human Rights and that requiring the Crown to prove sanity would pose considerable practical difficulties.[60] On this issue, the Law Commission for England and Wales has taken a different view, suggesting that the European Commission on Human Rights may have fallen into error in holding the reverse burden compatible with the European Convention on Human Rights,[61] and proposing instead that the accused should bear an 'elevated evidential burden'.[62] The difference in practice between an 'elevated evidential burden' and a persuasive one may, of course, be rather narrow.

Fourthly, the Scottish Law Commission rejected proposals that the defence of insanity should include an 'irresistible impulse' or volitional component, noting that the mental health experts they met with were 'virtually unanimous in rejecting a category of mental disorder which was purely volitional in nature and which had no impact on cognitive functions'[63] and that none of their consultees could identify a case for which such a defence would be necessary and appropriate.[64] The Law Commission for England and Wales does not identify such a case either, although it does refer to the example of a 'compulsive

52 Criminal Procedure (Scotland) Act 1995, s 51A(2), referring to 'a personality disorder which is characterised solely or principally by abnormally aggressive or seriously irresponsible conduct'.

53 See Maher (n 40) 2.

54 Criminal Procedure (Scotland) Act 1995, s 51A(4).

55 *HM Advocate v Harrison* (1968) 32 JCL 119, where the Crown was permitted to argue that the basis for H's plea of diminished responsibility in fact amounted to insanity.

56 A difficulty recognised in *R v Cottle* [1958] NZLR 999, 1027 per North J. See Chalmers and Leverick (n 2) para 7.05.

57 Law Com DP (n 1) para 4.131.

58 Scottish Law Commission, *Discussion Paper on Insanity and Diminished Responsibility* (Scot Law Com DP No 122 2003) para 5.19.

59 James Chalmers, 'Reforming the Pleas of Insanity and Diminished Responsibility: Some Aspects of the Scottish Law Commission's Discussion Paper' (2003) 8 SLPQ 79.

60 Scot Law Com No 195 (n 29) paras 5.2–28. See now Criminal Procedure (Scotland) Act 1995, s 51A(4).

61 Law Com DP (n 1) paras 8.20–21, discussing *H v UK* App no 15023/89, unreported.

62 Law Com DP (n 1) para 8.50.

63 Scot Law Com No 195 (n 29) para 2.54.

64 Ibid.

hoarder' as someone who lacks self-control but cannot be said to lack rationality.[65] Surprisingly, the English Law Commission notes the practical difficulties associated with such a plea,[66] briefly quotes Mackay as saying that these are 'perhaps too negative'[67] (without further elaboration on the reasons for this view), notes further practical difficulties,[68] and then abruptly concludes that such a defence should be allowed.[69]

Boundary issues

One of the most significant issues in the English Law Commission's discussion paper, and a useful point of contrast between Scots and English law, is the boundary between the defence of automatism and an alternative 'mental disorder' or 'medical condition' defence. This is where the deficiencies in the Scottish project of reviewing insanity in isolation from automatism become clearly evident.

Prior to this, the Scottish courts had adopted a noticeably flexible approach to distinguishing between the defences. Automatism, it was said, was based on an 'external cause';[70] insanity was based on 'mental illness, mental disease or defect or unsoundness of mind'.[71] Neither of these formulations had, however, been developed further by the courts, and various decisions made it clear that they were unlikely to be applied rigorously. For example, it was accepted that sleepwalking should be dealt with as automatism despite the absence of an external cause,[72] while the courts' approach to physical illnesses varied, treating 'psychic epilepsy' as the basis for a plea of insanity in one case[73] and internal injuries causing blood poisoning leading to 'temporary mental dissociation' as the basis for an unqualified acquittal in another.[74] There has been no discussion in the Scottish cases of whether the concept of 'mental illness, mental disease or defect or unsoundness of mind' can be interpreted broadly to include conditions which *affect* the mind even if they are 'physical' rather than 'mental' in nature.[75]

The statutory formulation of the new Scottish defence of mental disorder excluding criminal responsibility is such that 'mental disorder' is restricted to mental illness, personality disorders and learning disabilities, a test which was adopted from existing mental health legislation apparently without analysis on the Scottish Law Commission's part.[76] The logical consequence of this seems to be that any other condition which leads to a 'total alienation of reason' – which remains the test for automatism – must be entitled to an acquittal. Although the automatism defence must in theory be based on an 'external cause', that was simply a mechanism for distinguishing it from the cognate plea of insanity. The

65 Law Com DP (n 1) para 4.48.

66 Ibid paras 4.50–51.

67 Ibid para 4.51, quoting R D Mackay, *Mental Condition Defences in the Criminal Law* (Oxford University Press 1995) 116.

68 Law Com DP (n 1) para 4.52.

69 Ibid para 4.53.

70 *Ross* (n 21) 214 (Lord Justice-General Hope).

71 *Brennan v HM Advocate* 1977 JC 38, 46 (Lord Justice-General Emslie).

72 *Finegan v Heywood* 2000 JC 444. In this case, F's sleepwalking was regarded as externally caused (by alcohol consumption) and the defence was excluded on the basis of prior fault, but the court proceeded on the basis that this external cause excluded what would otherwise have been a valid defence.

73 *HM Advocate v Mitchell* 1951 JC 53.

74 *HM Advocate v Ritchie* 1926 JC 45.

75 As in *R v Kemp* [1957] 1 QB 399.

76 Criminal Procedure (Scotland) Act 1995, s 307; Mental Health (Care and Treatment) (Scotland) Act 2003, s 328(1).

courts have made it clear that a total alienation of reason is an absence of *mens rea*,[77] which in itself entitles the accused to an acquittal, and that would be true where the cause of the alienation is internal but does not amount to mental disorder. The only exception to this is that, if they are at fault in bringing about that alienation, *mens rea* will be presumed by the application of a legal fiction.[78]

The effect of the Scottish legislative reforms, therefore, is to reposition moderately the boundary between the defences of insanity/mental disorder and automatism, marginally broadening the latter. This is crucially important because, of course, the distinction has significant consequences: first, in terms of stigma, which may attach to the mental disorder defence but is unlikely to do so where automatism is concerned, and, secondly, because the former defence does not result in an unqualified acquittal but leaves the accused subject to the coercive power of the state.

Here, the approach of the Law Commission for England and Wales is a radically different one. It recommends a new defence of 'not criminally responsible by reason of recognised medical condition'. Automatism would remain, but as a wholly residual defence. This is brought out vividly by a table in the English Law Commission's paper which 'illustrates how existing cases would be decided were they to be tried under the law contained in our proposals'.[79] The table lists 16 cases decided by the courts between 1955 and 2007, alongside a fictional example – a swarm of bees entering a car and causing the driver to swerve – mentioned in a 1945 decision.[80] One of the cases would be determined under the rules on self-induced intoxication and the other 15 would be considered for the defence of recognised medical condition, although it would not succeed in all of them. It is only the fictional example of a swarm of bees where automatism would apply. The fact that the only case the English Law Commission can identify which would be dealt with as automatism under its proposed scheme is a fictional example is a clear demonstration of just how practically irrelevant automatism would become on its approach.

Does this matter? In one respect, the boundary change is a hugely welcome one, because it does more to combat stigma than a mere renaming of the insanity plea ever could.[81] Insofar as the verdict is a statement about the culpability of the defendant, avoiding any distinction between 'mental' and 'physical' illnesses is to be welcomed.

However, a second consequence of the change is a far less welcome one. Automatism results in an unqualified acquittal. The recognised medical condition defence would not. This is not an accident; the English Law Commission has consciously reached the view that the recognised medical condition defence provides 'more appropriate disposal powers', suggesting that it is unnecessary to be concerned about the net-widening involved in this proposal because the verdict is 'non-stigmatising'.[82]

Coercive powers following acquittal: is prior fault an alternative approach?

An acquittal, simply stated, is a decision that someone has not been established to be morally blameworthy, and it follows from this that the state is not entitled to inflict

77 As is clear from *Ross* (n 21). It is thought that, despite this approach, the defence would remain available in respect of offences of strict liability: see Chalmers and Leverick (n 2) para 7.05.

78 See *Ross* (n 21) 215 (Lord Justice-General Hope). This is a problematically harsh rule: see further Chalmers and Leverick (n 2) paras 7.44–45.

79 Law Com DP (n 1) para 4.169.

80 *Kay v Butterworth* (1945) 61 TLR 452.

81 See Law Com DP (n 1) para 4.57.

82 Ibid para 6.47.

punishment on them. For a variety of well-understood reasons, it may be reasonable for someone who is acquitted on the ground of mental disorder to be subjected to coercive measures, particularly because it may be unreasonable to expect them voluntarily to take the necessary steps to address their condition. What the English Law Commission suggests, however, is rather different. Under its proposals, someone who *is* competent to deal with a medical condition, but was not at fault in failing to do so, would be subject to the coercive power of the state despite being neither at fault nor in some way lacking in competence. In justifying one aspect of the breadth of the recognised medical condition defence, the English Law Commission suggests that it resolves an inconsistency in the law relating to those with diabetes:[83]

> Under the present law the outcomes in cases involving diabetic defendants who plead a lack of capacity are inconsistent. A diabetic who fails to take insulin and then commits an allegedly criminal act while totally incapacitated will be found not guilty by reason of insanity. This result flows from the fact that the incapacity had an 'internal cause' (the diabetes). If, on the other hand, she took insulin in accordance with a medical prescription, but was unable to take it with food or had an unexpected reaction to it through no fault of her own, and committed an allegedly criminal act while lacking capacity, she would be entitled to a verdict of not guilty for all crimes since the loss of capacity was involuntary. We think that it is illogical that one blameless defendant should be entitled to a complete acquittal while the other is labelled insane, for the same reasons as explained above. Under our proposal, the verdict in both cases would be not criminally responsible by reason of a recognised medical condition.

The English Law Commission is correct to identify an illogicality here, but it does not follow from this that *both* individuals should be denied complete acquittals. The illogicality could equally be avoided by affording a complete acquittal to both. The state, after all, has no power to take coercive measures against a competent person simply by virtue of their being diabetic; it is doubtful that any necessity for such a power has ever been suggested. It is, therefore, unclear why it should obtain such a power because a person with diabetes is found not guilty of a criminal offence.

The English Law Commission emphasises that the recognised medical condition defence will be excluded where prior fault exists.[84] The prior fault rule, however, may be sufficient on its own to deal with such cases. If an individual is at fault in managing a medical condition, so that they lose capacity and act in a blameworthy fashion, they will not be entitled to the defence. While the existence of this rule can be justified simply as a matter of culpability, it also allows the courts to protect the public against individuals who are unable or unwilling to manage their own medical conditions. However, in a case where the defendant was not at fault, what is the justification for treating them as a person who must be subjected to coercion in order to ensure that they manage their condition in future? If such a justification exists, why should it be contingent on a criminal prosecution?

To take the example of a person with diabetes, if a person is clearly entitled to the recognised medical condition defence on this basis, it would not be appropriate to prosecute them, because there would be no reasonable prospect of conviction.[85] In such cases there would be no civil mechanism whereby coercive measures might be sought (in contrast to those available under mental health legislation). Coercive measures of any sort would therefore be ruled out in cases where the defence was clearly made out on the facts as

83 Law Com DP (n 1) para 6.50.

84 Ibid, para 3.19; the principle of prior fault is discussed in more detail throughout the paper.

85 Crown Prosecution Service, Code for Crown Prosecutors (CPS 2013) para 4.4.

known to the prosecution prior to trial, and there is no obvious principled basis for holding that they should be available (and only available) where there was some doubt about the defence which justified a prosecution. The likelihood, of course, is that coercive measures would rarely if ever be applied by the courts in cases not involving mental disorder, if only because there seems to be no clearly identified need for them in such cases. To that extent, the English Law Commission's proposals would be unlikely to lead to practical difficulties, but they represent a surprisingly relaxed approach to the extent to which it is appropriate for the state to claim coercive powers to restrict the liberty of the individual.

NILQ 65(2): 215–28

'Special hearings' under New Zealand's Criminal Procedure (Mentally Impaired Persons) Act 2003

WARREN BROOKBANKS

University of Auckland

1 Introduction

In 2003 the New Zealand government passed legislation aimed, *inter alia*, at reforming the law on unfitness to stand trial and making provision for appropriate options for the detention, assessment and care of defendants and offenders with an intellectual disability.[1] The Intellectual Disability (Compulsory Care and Rehabilitation) Act 2003 (IDCCR) was novel legislation making express statutory provision for the first time for intellectually disabled offenders.[2] However, that statute is not considered further in the present context. One of the innovations achieved by the Criminal Procedure (Mentally Impaired Persons) Act 2003 (NZ) (CPMIP) was the creation, in s9 of the Act, of a new procedure for testing 'evidential sufficiency' prior to a determination of unfitness to stand trial. The 'evidential sufficiency' hearing, referred to as a 'special hearing' in the Act, and sometimes referred to as an 'involvement hearing', was designed to determine whether the offender had 'caused' the acts or omissions constituting the *actus reus* elements of the offence he or she was charged with before they were at risk of a finding of unfitness to stand trial. The procedure is broadly modelled after the English 'trial of the facts' procedure although, unlike its English counterpart, it occurs *prior* to a determination of unfitness. If a negative finding was made by the court, the offender would be immediately discharged, although a discharge did not amount to an acquittal,[3] a matter I will refer to later in this article. If a positive finding was made, the matter would proceed to a fitness to plead hearing, as specified in the Act.[4]

The apparent statutory purpose of the s9 hearing was to divert mentally impaired offenders from the criminal justice system where the evidence was insufficient to sustain a criminal charge. In practice, however, the wording of the section and its interpretation and application by the courts have rendered it a contentious and intensely litigated provision.

1 See the Criminal Procedure (Mentally Impaired Persons) Act 2003 (NZ) and the Intellectual Disability (Compulsory Care and Rehabilitation) Act 2003 (NZ).

2 For a detailed history of this legislation and some of the problems that have emerged since its enactment, see Warren Brookbanks, 'New Zealand's Intellectual Disability (Compulsory Care) Legislation' in K Diesfeld and I Freckelton (eds), *Involuntary Detention and Therapeutic Jurisprudence: International Perspectives on Civil Commitment* (Ashgate 2003); Warren Brookbanks, 'Managing the Challenges and Protecting the Rights of Intellectually Disabled Offenders' in B McSherry and I Freckelton (eds), *Coercive Care Rights, Law and Policy* (Routledge 2013).

3 See CPMIP, s 13(3).

4 See CPMIP, s 14.

This has not, in my view, fulfilled the expectations of the legislature. Ten years after its enactment, case law devoted substantially to analysing and determining the operation of the section continues to flood from the courts. Yet it would seem that in New Zealand we are no closer to a settled understanding of the purpose and scope of the section than we were when it was first enacted.

In this article I examine some of the more controversial aspects of the 'evidential sufficiency' hearing, with a view to highlighting the complexities the section has given rise to, and suggesting directions for possible future reform.

2 Evidential sufficiency

Section 9 of the CPMIP states:

> A court may not make a finding as to whether a defendant is unfit to stand trial unless the court is satisfied, on the balance of probabilities, that the evidence against the defendant is sufficient to establish that the defendant caused the act or omission that forms the basis of the offence with which the defendant is charged.

The term 'evidential sufficiency' was coined by judges to describe the purpose of the s9 hearing. It is concerned with ascertaining whether there is enough evidence to prove the offender *caused* the *actus reus* of the offence(s) charged. However, the statute gives no indication as to what might constitute evidential sufficiency and it has been left to the courts to determine the nature, scope and limits of available evidence for s9 purposes. Essentially, the s9 hearing is designed to determine whether the defendant committed the *actus reus* elements of the 'offence'.

It is clear that in drafting the provision to refer to 'offence' in the singular, the legislature did not anticipate *multiple* offences or differing degrees of complicity in determining the s9 issue. This has raised the question of what must be proved by the prosecution to meet the evidentiary threshold and whether the physical elements of *all* offences charged in a particular prosecution must be proved.

In order to resolve this problem the courts have held that, where multiple offences are alleged, it is only necessary to prove the *actus reus* of the most serious charge(s) in the indictment. It is not necessary for the court to make findings on all counts in the indictment.[5] The legislation has had to be made to work despite inadequate guidance as to types of evidence that may be adduced or the manner of its proof.[6]

At the heart of this inquiry is the issue of whether the defendant 'caused the act or omission that forms the basis of the offence charged'. The policy intent was that the involvement hearing would be a 'simple, short process' in which the prosecution proves that the defendant was responsible for the *physical elements* of the offence charged.[7] The conventional view of the courts was that, because the defendant's mental capacity was in doubt, it was not appropriate for the court to consider *mens rea* issues at a s9 hearing. It was thought that such an approach would considerably simplify a s9 hearing. Simplicity, however, is a luxury in this domain.

5 See *Barton v Police* DC Palmerston North CRI 2008–054–003750, 14 November 2011 [3] (Garland DCJ).

6 See *R v Komene* HC Auckland CRI-2012–090–002641, 7 June 2013 (Asher J). There the High Court interpreted s 7(1) of the CPMIP, which defines when a finding of unfitness to stand trial may be made, as applying when, after the commencement of proceedings, the option is to enter a guilty plea. It held that such an approach is consistent with the 'robust' approach to jurisdiction adopted by the Court of Appeal in *McKay v R* [2010] 1 NZLR 441 [92] where it was stated that the courts will fill a gap to make the legislation work.

7 Ministries of Justice and Health, *Improving the Criminal Justice Process for Mentally Impaired Offenders* (Issues Paper, CJS-16–07–05, 2011) 11.

Whatever elements *must* be proved, creating a burden of proof on a balance of probabilities suggests that the legislature's concern was that trial judge only needed to be satisfied that the offender *probably* did the act. There was no requirement, on face of the statute, for strict evidential proof. Hence, my characterisation of the s9 hearing as a 'relaxed evidential enquiry in which any evidence (in any form) which assists the court in making a determination on the core issue ought to be admissible, subject to the requirements of natural justice and the ability to test any evidence which may be inherently unreliable'.[8] This analysis of the nature of the s9 inquiry was described by the New Zealand Court of Appeal as 'probably right' in *R v McKay*.[9]

However, experience has since shown that s9 inquiry has been anything but relaxed. Furthermore, while the apparent intention of the legislature was to eliminate any inquiry into *mens rea* issues, in practice it has proven impossible to exclude *mens rea* completely. This is an aspect of the legislation that will be explored more fully as this article proceeds.

3 Elements of offence to be proved in determining whether accused caused the act or omission

A continuing area of controversy with s9 concerns the scope of the expression 'caused the act or omission' in the section. Because of the inherent ambiguity in this expression, a number of other issues have arisen, requiring clarification by the courts. These include: the scope of *mens rea* in s9 hearings; the availability of defences; and the meaning of 'objective evidence'.

The *mens rea* issue constitutes one of the major testing grounds for the operation of s9 hearing. It has shown that s9 is more difficult to apply in practice than was first envisaged. A question that the courts were forced to address at an early stage was whether it is possible to separate *actus reus* elements from the mental element in those crimes where the *actus reus* includes a *mens rea* element. Various solutions have been proposed, including excluding *mens rea* altogether from the inquiry, requiring proof of *all elements* (including *mens rea*) and a hybrid approach, which allows the introduction of evidence negating the unlawfulness of an act but not requiring proof of the full mental element. These approaches are reflected in various common law decisions and in statutory models adopted in different jurisdictions.[10]

The issue was considered at length by the New Zealand High Court in *R v Cumming*.[11] Counsel for the prosecution in *Cumming* argued that the s9 inquiry was limited to the *actus reus* of the offences charged, on the basis that s9 was a 'filter' to remove cases where it was not necessary to consider fitness to stand trial because there was insufficient evidence that the accused was responsible for the offence at the outset. On this view the proper place for consideration of the mental element was later in the trial if the accused was found fit to stand trial. Furthermore, it was argued that, if proof of *mens rea* were to be required at the s9 stage, the court would have no choice but to discharge the accused even though he had committed the necessary act, something, it is claimed, Parliament did not intend. Alternatively, if the court were to find at the s9 hearing that the accused did have the necessary *mens rea*, that finding could be compromised if the accused were later found unfit to stand trial.[12]

8 See Warren Brookbanks, 'Special Hearings under CPMIPA' [2009] New Zealand Law Journal 30, 40.

9 [2009] NZCA 378 [48].

10 See discussion in Brookbanks (n 8), 33.

11 17/07/09, HC Christchurch, CRI 2001 009 0835552 (French J).

12 Ibid [60].

The court noted that support for the prosecution's approach was to be found in earlier decisions of the High Court[13] and in a Ministry of Justice publication.[14] In the light of these authorities and arguments by the Crown, French J concluded that s9 does *not* require proof of all the ingredients of the offence, an intention which the legislature could readily have indicated by substituting the words 'committed the offence' for those currently in contention. However, a more difficult issue for determination was whether Parliament intended to exclude *any* inquiry into the defendant's mental state at the time of the alleged offending.

In contrast to the Crown position, the defence in *Cumming* sought to persuade the court that s9 allows consideration of some aspects of *mens rea*, a position affirmed by overseas authority, in particular *R v Antoine*[15] and *R v Ardler*.[16] The court's treatment of these decisions in *Cumming* illustrates how they have come to impact the development of New Zealand law around s9 hearings.

Antoine was concerned with the interpretation of the phrase 'did the act or made the omission charged' in s 4A (2) Criminal Procedure (Insanity) Act 1964 (UK). The similarity of the phrase to s9's 'caused the act or omission that form the basis of the offence' is noted. In England, determination of the question is a jury matter and the standard of proof is beyond a reasonable doubt. Section 9, in contrast, only requires the court to be satisfied 'on a balance of probabilities'. The issue in *Antoine* was whether the offender, charged with murder, could rely on the defence of diminished responsibility in determining whether he 'did the act or made the omission charged'. The House of Lords held that diminished responsibility was not available and overruled earlier authority suggesting the statute required proof of all the ingredients of the offence. Lord Hutton noted that the purpose of the English provision[17] was to strike a 'fair balance' between protecting a person who has, in fact, done nothing wrong, but is unfit to plead, and protecting the public from a defendant who has actually done an injurious act which would have been a crime if committed with *mens rea*. His Honour went on to say that the section achieves the necessary balance by distinguishing between a person who has not carried out the *actus reus* of the crime charged against him and a person who has committed an act or omission which would be a crime if accompanied by *mens rea*.[18]

However, Lord Hutton also acknowledged that offences do not always divide neatly into separate *actus reus* and *mens rea* compartments, since some *actus reus* elements may also imply a mental element (for example, proof of the *actus reus* of possession of an offensive weapon is dependent on the defendant's intention in order to determine if the weapon is defensive).[19] Lord Hutton's approach was to say that the mandate to determine whether the accused did the 'act' did not require consideration of whether the defendant had the requisite *mens rea* for the offence. However, if the defendant had an arguable defence of accident, mistake or self-defence, supported by 'objective' evidence, which he would have raised if the trial had proceeded in the normal way, then the jury should not find the accused

13 See *R v Codd* [2006] 3 NZLR 562 at [38]; *T v Roberts* HC Auckland CRI 2005 092 014492, 22 November 2006 (Fogarty J) and *R v De Wes (Ruling (No 1))* HC Gisborne, CRI 2006 016 003323, 3 November 2008 (Keane J) [14].

14 See *Guide to the Criminal Procedure (Mentally Impaired Persons) Act 2003* (Ministry of Justice 2004).

15 [2001] 1 AC 340.

16 [2004] ACTCA 4.

17 See Criminal Procedure (Insanity) Act 1964, s 4A(2) (UK).

18 *R v Antoine* [2001] 1 AC 340, 375.

19 *R v Cumming* 17/07/09, HC Christchurch, CRI 2001 009 0835552 [73] (French J).

did the 'act', unless satisfied to the requisite standard of proof on all the evidence that the prosecution has negatived any defence(s) that may have been raised.[20]

On this basis, mistake, accident and self-defence may properly be characterised as *actus reus* defences since they are concerned not primarily with the accused's particular mental state at the time of the killing, but whether the accused's actions were otherwise lawful. Lord Hutton's point, while conceding that the defences named 'almost invariably' involve some consideration of the accused's mental state,[21] is that only independent evidence pointing to the lawfulness of the accused's actions, *as a matter of independent observation*, will be sufficient to trigger the nominated defence. Evidence coming from the accused himself or herself or expert evidence as to factors *internal* to the accused's psychological make-up will not suffice, hence the exclusion of mental state defences in determining responsibility for *act* or *omission*.

The case of *Ardler*[22] followed Antoine in finding that the inquiry at a 'special hearing' is directed not simply to *actus reus*, nor the full elements of *actus reus* and *mens rea*, but to an 'unlawful' act. In the context of a prosecution for rape, the court held that 'objective evidence' of raising issues, including mistake, accident, self-defence and lack of specific intent necessary to constitute an offence, would be available to negative the physical acts of an offence, but that the prosecution was not required to negative any lack of mental capacity to act intentionally or voluntarily. For this reason, the court held that pleas of mental impairment, provocation, or diminished responsibility were unavailable at a special hearing. An example given in the judgment of a specific intent 'of the particularity necessary to constitute the offence' was the crime of arson, where in the relevant jurisdiction, the specific intent of 'endanger the life of another' had to be established in order that the 'acts' proved constituted arson and not some lesser offence.[23]

In *Cumming* the prosecution contended that both *Ardler* and *Antoine* could be distinguished because of the different wording in the overseas statutes. However, French J observed that while the wording was not identical, it was very similar and it was difficult to see how wording disparity would lead to a different interpretation.[24] The principal distinction, the court found, was the fact that in New Zealand the s9 inquiry *preceded* the fitness to stand trial hearing, contrary to the practice in other jurisdictions where it *followed* an unfitness finding and a finding of mental impairment .

French J considered that the differences in the order of the 'facts' hearing and burden of proof requirements were independent of the reasoning of the courts in interpreting the words 'the act' and expressed the correct position (at least for the purposes of New Zealand law) as an adaptation and extension of the formula adopted by the English Court of Appeal in *R (on the application of Young) v Central Criminal Court*.[25] There Rose LJ approved the ruling of the trial judge as to the relevant principles emerging from Lord Hutton's speech in *Antoine*, namely:

(1) so far as possible, the inquiry should focus on an accused's actions as opposed to his state of mind;

(2) this distinction is dictated by the language [of s9] and the social purpose it serves;

20 *Antoine* (n 18).

21 Ibid (n 18) 376.

22 *Ardler* (n 16)

23 Ibid [77].

24 *R v Cumming* (n 19) [86].

25 [2002] 2 Cr App R 12.

(3) but the distinction cannot be rigidly adhered to in every case because of the diverse nature of criminal offences and criminal activity. In particular, it cannot be adhered to when *mens rea* is a composite element of the *actus reus*.

In addition, French J approved two further criteria arising from a consideration of the overseas case law:

(4) if there is objective evidence which raises the issues of mistake, self defence and accident, then the Court should not find the accused caused the act or omission unless satisfied on the balance of probabilities that the prosecution has negatived that defence;

(5) it is not open to an accused to argue absence of *mens rea* by reason of mental impairment. To the extent that passages in *Ardler* suggest otherwise, they are contrary to *Antoine* and the underlying legislative policy and should not be followed.

As the discussion so far has illustrated, the question of the availability of defences at a s9 hearing has become a matter of some significance. The essential argument is that if the purpose of such hearings is to determine responsibility for the physical elements of offences only, then there would seem to be little scope for evidence of defences, especially those that go to denial of the *mens rea*. Yet, as has already been established, there is dispute as to whether *any* defences wholly lack a *mens rea* component, even defences like mistake and accident which, it is claimed, are usually simple denials of *mens rea*, not the *actus reus*. Similarly, it might be argued that self-defence, which is a justification rendering what would otherwise be an unlawful assault a lawful act, is also dependent on *mens rea* notions, since the defendant's subjective belief in the circumstances prevailing is an essential element of the statutory defence.[26] It is hard to escape this analysis in respect of any defence we might choose to nominate, since most defences, as doctrines of the criminal law, concern unusual or abnormal states of mind.[27] Indeed the only 'defence' which it might be claimed is a 'pure' *actus reus* defence is the defence of involuntariness, since it amounts to a denial that the *actus reus* was produced with conscious volition – a notion independent of the requirements for *mens rea*.[28] Since on this view most defences contain a *mens rea* component in their conceptual structure, the question becomes: which defences are apt to establish *actus reus* elements when advanced via *objective evidence* given independently of the accused own subjective account? The examples given by Lord Hutton in *Antoine* are illustrative.

His Lordship gives as one example the case of a defendant who has struck another person with his fist, the blow causing death. In such a case, Lord Hutton suggests, it would be open to the jury at a special hearing (trial of the facts) to acquit the defendant charged with manslaughter *if a witness gave evidence* that the victim had attacked the defendant with a knife before the defendant struck him.

Lord Hutton's second example is where a woman has been charged with the theft of a handbag but a witness gives evidence that on sitting down at a table in a restaurant the defendant had placed her own handbag on the floor and, on getting up to leave, picked up the handbag placed beside her by a woman at the next table. In such a case it would be open to the jury to acquit.[29]

---

26 Crimes Act 1961, s 48 'in the circumstances *as he believes them to be*'.

27 Jerome Hall, *General Principles of Criminal Law* 2nd edn (Bobbs-Merrill 1960) 19.

28 *Kilbride v Lake* [1962] 590, 592 (Woodhouse J): 'This elementary principle obviously involves proof of something which goes behind any subsequent and additional inquiry that might become necessary as to whether *mens rea* must be proved as well'.

29 *Antoine* (n 18) 377.

'The expression 'objective evidence' arises in this context in the judgment of Lord Hutton in *Antoine*.[30] The objective evidence Lord Hutton has in mind specifically is evidence of the nature discussed in the examples above. Clearly, his Honour has in mind independent eyewitness evidence able to offer testimony as to the presence, or absence, of an *actus reus* element. The organising principle in these examples would appear to be that, where *independent testimony* points to evidence that the defendant acted in a manner that was inconsistent with the essential external element of the offence charged (for example, an unlawful assault in the case of manslaughter or an *unlawful taking* in the case of theft), such evidence is available, together with any other evidence led at the hearing, from which a tribunal of fact may find that the *actus reus* of the offence has not been proven.

Occasionally, such independent evidence may be supplied through expert testimony pointing to the fact that the accused suffered from a condition which had impaired his/her capacity to act voluntarily. This is implicit in another example given by Lord Hutton. His Honour envisions a situation in which the defendant kicks out and strikes another person in the course of an uncontrollable fit brought about by a medical condition. In this case the defence counsel advances the defence that the defendant, in law, did not do the 'act' because his action was involuntary. But in that case there would have to be evidence that the defendant suffered from the condition.[31]

In each of these examples, what is determinative is the presence of independent testimony that tends to negate the accused's responsibility for an *external element* of the offence charged. The critical issue is not, it would seem, whether the particular defence is characterised as a *mens rea* or *actus reus* defence, but whether the particular defence advanced is capable of negating an *actus reus* element. Defences including self-defence, mistake, accident, act of a stranger, involuntariness,[32] automatism, impossibility of compliance, causal necessity, and alibi may serve this purpose. Others, including intoxication, provocation, diminished responsibility, insanity, compulsion, teleological necessity, generally serve to negate responsibility for the *mens rea* of crime and, as such, are irrelevant to the s 9 type of inquiry.

Cases where the *actus reus* includes a mental element which must be proved as part of the *actus reus* fall to be determined on a case-by-case basis and depend on the specificity of the elements in the offence charged. In such cases it will usually be a matter of statutory interpretation as to whether a particular mental element is a true *mens rea* element, or whether it is properly categorised as a material fact relevant to determination of the *actus reus*.[33]

However, in New Zealand, judicial opinion remains divided on whether evidence as to *mens rea* elements can *ever* be adduced at a s9 hearing; or whether s9 should be interpreted as meaning that only the commission of the physical act or acts need be proved.[34] The New Zealand Court of Appeal has observed that the 'objective evidence' model contended for in cases like *Antoine* and *Ardler* requires 'difficult distinctions' to be made.[35] The court found that such distinctions between defences supported by objective evidence and mental state defences that cannot be relied upon at a special hearing would be unnecessary if the s 9 hearing were limited to proof that the defendant committed physical acts that form the

30 *Antoine* (n 18) 377.

31 Ibid.

32 See *Police v Espanto* 1/05/09, Morris DCJ, DC North Shore, CRI 2008 044 009415 where the court held that while *mens rea* need not be proved in terms of intention, or knowledge of indecency, it must still be established that the accused's act was voluntary, or deliberate.

33 See *R (Young) v Central Criminal Court* [2002] 2 Cr App R 12.

34 As in *R v Lyttleton* HC Auckland CRI 2008–04466, 4 November 2009 (Wylie J).

35 See *R v Te Moni* [2009] NZCA 560 [79] (CA).

basis of the offence, as opposed to the *actus reus*. Yet the court concluded that such an approach did not appear to set 'a sufficiently high threshold to meet the objective of s9, which is to ensure that the court has made a *finding of criminal culpability* before the sanctions which can apply to a person who is unfit to stand trial can be imposed on that person'.[36] It is difficult to imagine how a finding of 'criminal culpability' might be made in the absence of some inquiry into *mens rea*. Thus the current law is in a state of some uncertainty. Nevertheless, there seems to be a general acceptance that the principal focus of a s9 inquiry should be on an accused's actions as opposed to state of mind, although such a distinction cannot be insisted upon where mental elements like consent (or its absence) are axiomatic to the definition of a particular offence.[37]

Two recent cases illustrate the way in which the provision has been applied in the case law with regard to the mental element in the definition of the offence. In *WH v Police*,[38] the accused faced a charge of possession of an offensive weapon under s 202(4)(b) of the Crimes Act 1961 (NZ). After a s9 hearing, the District Court found that the appellant had caused the act that formed the basis of the offence charged. The appeal against that finding was based on the ground that the evidence was insufficient to establish, on the balance of probabilities, that the appellant 'had in his possession an offensive weapon in circumstances that prima facie showed an intention to use it to commit an offence involving bodily injury or the threat or fear of violence'.

The appellant lived with his elderly parents. On the occasion in question, after an argument over money, he had taken a kitchen knife and, in anger, stuck it into the kitchen wall. Neither parent was in the room when this occurred. Neither experienced any fear as a result of his actions. When the appellant continued banging on the wall of the caravan in which he slept, and into which he had moved after his actions in the kitchen, making a very loud noise and showing no intention of stopping, his father called the police.

The offence of possession of an offensive weapon is defined as having possession 'of any offensive weapon . . . in circumstances that prima facie show an intention to use it to commit an offence involving bodily injury or the threat or fear of violence.'[39] Thus, the 'circumstance' element of the *actus reus* included a mental element that was indivisible from the *actus reus* itself and, therefore, part of the *actus reus* that had to be proved in order to establish evidential sufficiency.

The issue on appeal was whether the judge was entitled to find, on the balance of probabilities, that the evidence established objectively a *prima facie* intention on the part of the appellant to use the knife to commit an offence involving the threat or fear of violence. The District Court judge found that putting a knife in the wall was an act that could be indicative of a situation where there would be a fear or threat of violence and found the test in s9 was made out.

However, on appeal, Venning J found that the father's expressed concern was with the appellant's continual banging on the caravan, which led him to call the police. He held that, since the threat or fear of violence, which the prosecution relied on, must arise out of the act of sticking the knife into the wall, without it being done in the presence of the complainant meant that the appellant's conduct was as consistent with other actions involving property damage and banging on his caravan as showing an intent to commit an act involving the threat or fear of violence. Because, from the evidence, it appeared that the

36 *R v Te Moni* (n 35) [author's emphasis added].

37 See *R v Cummings* (n 19) (French J).

38 HC Auckland CRI-2012–404–000371, 28 November 2012 (Venning J).

39 Crimes Act 1961, s 202A(4)(b) (NZ).

appellant's father did not perceive the knife or the appellant's act of sticking it into the wall as a threat, nor was he frightened by the act, the evidence could not support a finding, even on the balance of probabilities, that the appellant had possession of the knife in circumstances that showed an intention to use it to commit an offence involving threat or fear of violence. The appeal was allowed and the finding of 'evidential sufficiency' was quashed and the appellant discharged in terms of s 17(1) of the CPMIP.

The decision was, in effect, a finding that the appellant lacked the mental element contained within the *actus reus* of the offence charged. It is clear that the court on appeal could not have determined the appellant's liability for 'caus[ing] the act . . . that forms the basis of the offence with which the defendant is charged' without an investigation into that mental element.

In *Police v MH*,[40] after the defendant was charged with 10 counts of sexual violation and other related charges, the question of his fitness to plead arose. The court observed that the 'mental element' of an offence does not have to be proved at a s9 hearing. However, proof of a mental element was not wholly excluded since the court found that for the sexual violation charges the prosecution had to prove on the balance of probabilities (a) penetration of the complainant's vagina or anus, and (b) that this occurred without her consent. Thus, the only mental element required to be proven for the purposes of s9 was that of consent by the complainant.

The complainant and the defendant had been in an 'on again/off again' relationship for about 10 years prior to the alleged events. At the time of the alleged offending they had been living in the defendant's mother's home. The prosecution evidence consisted mainly of the testimony of the complainant, principally from DVD recordings of interviews. The one defence witness was a consulting psychiatrist who was of the opinion that the complainant suffered from factitious disorder, a psychiatric disorder where people repeatedly present themselves to medical authorities with complaints or histories that cannot be reconciled with the evidence. They do this because of a deep-seated need to be cared for and so that health professionals will believe them and care for them. Importantly, for the purposes of this case, was the fact that false claims of rape are associated with factitious disorder.

The court concluded that there was strong evidence that the complainant suffered from factitious disorder and did so at the time of making the complaints. There were serious problems with the reliability of her claims regarding illness and a strong possibility this could have escalated to complaints of rape as another way of getting attention. In addition, there was no credible independent evidence supporting her claims of sexual violation. In the circumstances, the court found that the evidence against the defendant was not sufficient to establish, on the balance of probabilities, that he caused the act which formed the basis of the offence charged and that the finding applied to each of the 18 charges against him.

In concluding the court made an important observation about the burden of proof suggesting, in the judge's view, a need for reform. This was the fact that s 13(3) CPMIP states that a discharge does not amount to an acquittal. Noting that the logic of that proposition 'entirely escapes me', the judge asked: if the prosecution was unable to prove the matter on the balance of probabilities, what possible chance would it have of proving it beyond reasonable doubt, as would clearly have been required had the matter proceeded to a full criminal trial? His Honour said:

40 DC Auckland CRI-2012–092–007682, 17 May 2013 (McElrea DCJ).

> I would have thought that a person who has been discharged as a result of a s9 hearing under this Act is entitled to be treated as if he had been acquitted and I trust that the defendant in this case will be so treated.[41]

This issue has not yet been addressed as a matter of law reform. However, the view that the matter should be proved beyond a reasonable doubt is supported by strong authority.[42] It remains to be seen if the idea is adopted as part of an ongoing review of the CPMIP.

4 Reflections on the meaning of 'act' in s9

What seems reasonably clear is that, when the legislature formulated the terms of s9, it had in mind a simple procedure to determine whether an offender had committed acts or omissions that would have amounted to the *physical ingredients* of the alleged crimes. It was not considered appropriate for the courts to investigate *mens rea* issues, given that the accused's mental capacity was in contention.[43] As has already been observed, the intention seems to have been to provide a filter whereby a person who was presumptively innocent of the alleged offence(s) could be diverted from the criminal justice system before they were at risk of being found unfit to stand trial. Unfortunately, the words chosen were not apt to effect the intended purpose of the provision and have led to considerable litigation, centring on the phrase 'caused the *act* or *omission*'. The essential difficulty is that the word 'act' is not sufficiently austere as to exclude the possibility of the mind also intruding. The issue might have been resolved had the legislation simply referred to the 'physical ingredients' of the offence in place of the expression 'act or omission'.[44] For many ordinary *conduct* offences the *actus reus* may be fully expressed in a simple form of conduct. So it is possible to say, for example, that the 'act' of D striking V on the nose, constituted the *actus reus* of assault. Or, to take another simple example, the 'act' of D in lighting a fire under V's house damaging it constituted the *actus reus* of criminal damage. In such cases, there may be little difficulty in determining whether D did the act constituting the basis of the offence he or she is charged with. Certainly, no inquiry needs to be made about the accused's mental state to determine whether he or she did the act. The difficulties arise, however, firstly, where the offence charged contains an ulterior intent element, proof of which is necessary to determine the character of the alleged 'act',[45] or, secondly, where the alleged 'act' is not an act at all, because it was performed unconsciously.

A. Ulterior intent crimes

Dealing with the first area of difficulty take, for example, the crime of 'wounding *with intent to cause grievous bodily harm*.'[46] If we place the emphasis on the words in s9 'the *offence . . . charged*', the *actus reus* of the offence of wounding with intent is not proved simply upon evidence that D punched V on the nose. While that may well be enough for a simple assault, for the aggravated 'act' 'that forms the basis of the *offence . . . charged*' something more is required. To determine that, it is necessary to inquire as to what was in D's mind when D struck V, to determine whether the act forms the basis of the particular alleged offence. However, even if the emphasis is placed on the words 'the basis', then a simple assault may not be sufficient to establish the *basis* for a charge of wounding with intent, regardless of

41 *Police v MH* DC Auckland CRI-2012–092–007682, 17 May 2013 [35] (DC)

42 See R Chambers, 'Trial Rights for the Mentally Impaired' (2011) 24 New Zealand Universities Law Review 478, 483.

43 See Ministry of Justice, *Report to the Health Committee* (Ministry of Justice 2000).

44 See *R v Te Moni* (n 35) [79].

45 For an illuminating discussion on this issue, see Chambers (n 42) 483.

46 Crimes Act 1961, s 188 (NZ).

the differing *mens rea* elements. For a wounding, there must at least be a flow of blood following a break in the continuity of the skin.[47] Therefore, evidence of a mere assault would not do the work necessary to establish the 'basis' of the offence charged.

Thus, it would seem that many offences defined in criminal legislation that bear an ulterior intent element may not be susceptible to a simple analysis in terms of whether the accused performed an 'act' that constituted their 'basis'. The hoped-for simplicity of the evidential sufficiency hearing evaporates in the face of a concerted investigation into the *actus reus* elements of a crime and the evidence necessary to meet the statutory burden of proof. For this reason it is suggested that s9 may need to be redrafted in order to clarify what elements must be proved at the inquiry and the nature of evidence that will be sufficient for that purpose.

B. Unconscious conduct

A second area of difficulty concerning the meaning of 'act' is whether it accommodates unconscious conduct. At this stage, this is a purely theoretical question because there has not been a case in New Zealand where the alleged 'act or omission' was committed while the offender was in a state of impaired consciousness. 'Act' is not defined in the statute. According to one definition, an act is an 'intentional bodily movement performed by an agent whose consciousness is reasonably intact'.[48] By implication, a person whose consciousness is *not* 'reasonably intact' cannot, by this definition, commit an 'act'. If a person does something that looks criminal in a state of diabetic automatism or while suffering from a rapid eye movement sleep disorder, this might look like 'acting' but, in reality, their conduct is a mere event in which the law, properly, should have no interest in assigning criminal culpability. They have not 'acted'.[49] What other risks may be associated with their conduct may well involve other agencies, but should not be the province of the criminal law.

On this basis it could be argued that a person who lacks the present capacity to commit a criminal 'act' for reasons of lack of consciousness or overbearing physical compulsion, must necessarily be excluded from a finding that their 'act' formed the basis of the offence charged against him or her. Whatever they may be, they are not 'actors' in a legal sense and must be discharged at a s9 hearing. This would include all sleepwalkers, concussive automatons, those compelled by external forces and anyone else whose 'act' was not the product of a conscious, deliberate choice. The problem, however, is that, as the law presently stands, the courts have only allowed *objective evidence* of accident, mistake and self-defence to be advanced as defences at s9 hearings. This appears to exclude all mental state defences and involuntariness-based defences, including automatism.[50]

This is another area requiring further clarification. There would appear to be no reason in principle why the nominated defences should be sufficient to exclude liability at a s9 hearing, but not defences based on involuntariness or lack of conscious awareness. There is clear potential for injustice if an offender is deemed to have 'acted' because certain physical events are evidentially attributable to him or her, when he or she was either unconscious at the time of the alleged act or physically incapable of resisting the pressure to 'act'.

47 See Andrew Simester and Warren Brookbanks, *Principles of Criminal Law* 4th edn (Thomson Reuters 2012) para 17.3.2.

48 See Stephen Morse, 'Symposium: The Mind of a Child: The Relationship Between Brain Development, Cognitive Functioning, and Accountability under the Law: Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note' (2006) 3 Ohio State Journal of Criminal Law 397, 399.

49 See *R v Luedecke* [2008] ONCA 716 [53] (Dougherty J) (ONCA).

50 See discussion in *R v UP* YC Auckland, CRI2010–204–000314, 17 March 2011, Judge A J Fitzgerald [16] (DC).

C. Proof of other actus reus elements

The discussion so far has tended to focus on the scope of defences available to negate responsibility for the external elements of an offence charged. But this does not exhaust the possibilities for exculpation at a s9 hearing. Since the *actus reus* of an offence includes both circumstances and consequences, in addition to the conduct element, evidence that a relevant circumstance or consequence had not been proved to the requisite standard should, in principle, negate responsibility for the *actus reus*. Similarly, a failure to establish causation in an offence requiring proof of particular consequences would equally negate responsibility for the *actus reus* in an appropriate case. It should also be said that proof of causation is an explicit requirement of the statutory provision, in that the prosecution is required to establish through evidence that the defendant 'caused' the act or omission etc.[51] For example, in a prosecution for homicide, where death is a specified consequence in the *actus reus* of culpable homicide (whether murder or manslaughter), the prosecution must prove both that the consequence occurred and that the defendant's behaviour caused that consequence.[52] Such an inquiry may be conducted independently of any investigation as to *mens rea*, which normally would only arise once any outstanding issues of causation had been resolved. Absence of proof of causation would negate responsibility for the act or omission for the purposes of s9.[53]

5 Reform

It is clear from the foregoing analysis that the 'evidential sufficiency' hearing defined in s9 CPMIP has given rise to a number of unforeseen challenges to trial judges tasked with determining unfitness to stand trial. In September 2011, officials in the New Zealand Ministries of Justice and Health wrote to their respective ministers seeking approval to release an Issues Paper, entitled *Improving the Criminal Justice Process for Mentally Impaired Offenders*, and to invite consultation with stakeholders. The Issues Paper, which initiated a review of the CPMIP and the IDCCR, noted that the legislation had created a number of problems with the practical application of certain provisions, leading to procedural inefficiencies, conflicting case law and uncertainty about some of the processes. To date none of the options for reform presented in the Issues Paper have been adopted by government, and the review is ongoing, albeit stationary at the present time. Some of the reform suggestions identified in the Issues Paper have now been overtaken by case law developments. Others await further consideration.

One of the more important recommendations for reform concerned the order of proceedings for determining unfitness. As has been noted, the current s9 inquiry into a defendant's responsibility for the offence (the 'evidential sufficiency' or 'involvement' hearing) precedes the inquiry into the offender's fitness to stand trial. The apparent rationale for this was to ensure that the defendant was not subjected to a fitness to plead inquiry unless it was first determined that he or she had committed the alleged acts. In effect, the procedure acted as a filter to remove a presumptively innocent offender from the trial process before they were at risk of an indeterminate unfitness finding. This order is, as noted earlier, the opposite to what occurs in the UK and in other jurisdictions in respect of the equivalent 'trial of the facts', where unfitness is determined *before* there is any inquiry into the accused's responsibility for the offence. The obvious advantage of this model is

51 CPMIPA, s 9.

52 Andrew Simester and Bob Sullivan, *Criminal Law Theory and Doctrine* (Hart 2003) 86.

53 See e.g. *Police v Palu* 8/06/09, Priestley J, HC Auckland, CRI 2008 404 0083 [11], where Priestley J observed that the trial judge had to be satisfied 'that there was sufficient evidence that "the respondent had *caused* the assaults lying at the heart of the four charges . . ."'. See also *R v Akuhata* [2013] NZHC 2669.

that where a defendant is found fit, the matter may proceed to trial in the normal way. Where an unfitness determination is made, the involvement hearing (trial of the facts) becomes, in effect, an alternative to a trial.

New Zealand courts have on a number of occasions commented on the problems arising because of the order of the evidential sufficiency hearing. In particular, in the case of *Te Moni*,[54] an appeal against conviction on a charge of sexual violation by rape, the trial court failed to make a finding in terms of s9 as mandated by the statute, which prompted the Court of Appeal to question whether the s9 hearing ought to come after the fitness assessment has been made. In that event, the hearing would occur only where there is to be no trial. This would avoid the procedural inefficiency and resource implications of a court having to hear the same evidence and witness testimony twice. The concern expressed by the Court of Appeal was that the current practice requires an accused person whose fitness to stand trial is in doubt to undergo a form of trial as part of a process to determine whether he or she is fit to do so.[55] The court observed that, if the s9 hearing happened *after* the assessment of fitness to stand trial, the process could be tailored to deal with the reality that the accused person was unable to properly participate. In the context of *Te Moni*, that would have meant that the requirement for the complainant to give evidence twice would have been avoided. For these reasons the Issues Paper has recommended that the order of the procedure for determining fitness be reversed, so that the fitness hearing would precede the involvement hearing.

Another issue, already noted in this article and addressed in the Issues Paper, concerns the burden of proof in evidential sufficiency hearings. Under s9, the standard of proof for determining evidential sufficiency is the balance of probabilities. In every other jurisdiction with an equivalent procedure, the standard is beyond reasonable doubt. The rationale for the lower standard of proof was the view that criminal liability is not determined at an involvement hearing and therefore the standard should be the civil standard. Professor Mackay and I addressed this issue in our article 'Protecting the Unfit to Plead: A Comparative Analysis of the "Trial of the Facts"'.[56] We argued that because the s9 procedure aims to determine elements of criminal responsibility and not simply the issue of trial capacity, the full burden of proof seemed inescapable. We argued that the presumption of innocence established in s 2(c) of the New Zealand Bill of Rights Act 1990 was, in principle, as applicable here as in any other context where the issue of criminal responsibility was to be determined.

As the Issues Paper has noted, raising the standard of proof would provide greater protection for an accused person, although it would make the prosecution's task more difficult. At the time of writing there is no evidence of any official proposal to alter the burden of proof as suggested. Nevertheless, I endorse the view of the late Justice Robert Chambers that the reasons for using a high standard of proof in the ordinary criminal context, namely, reducing the margin of error and properly reflecting the high value placed on individual liberty, apply equally in this context.[57] I would recommend that the change to the burden of proof be effected at the earliest possible date to reflect this reality.

54 CA 695/2008 [2009] NZCA 560.

55 *Te Moni* (n 35) [96].

56 (2005) Juridical Review 173.

57 Chambers (n 42) 484.

6 Conclusion

Unfitness to stand trial is now a commonly litigated trial procedure in many common law-based jurisdictions. In most such jurisdictions, including New Zealand and England, the rules are statutory, albeit supplemented by decisions of the higher courts. In New Zealand, while the procedures governing the substantive question of whether an offender is fit to stand trial are now reasonably well settled, the same cannot be said for those procedures which are an adjunct to the fitness determination, in particular the so-called 'trial of the facts' or, in New Zealand, the 'evidential sufficiency' hearing.

This novel procedure, designed to achieve procedural efficiency and to eliminate innocent offenders from the consequences of an unfitness finding, has proven difficult to interpret and complex to administer. While these problems are largely matters for the courts to resolve through the developing jurisprudence in this area of the law, some matters are better left for the legislature to address as a matter of law reform. Firstly, the current location of the hearing *before* a determination of unfitness to stand trial seems misconceived and has created a raft of unnecessary difficulties. These could be largely resolved by placing the hearing *after* the determination of unfitness, as occurs in other jurisdictions. Secondly, the normative requirements of criminal justice would seem to dictate that placing the burden of proof on the prosecution to prove 'evidential sufficiency' beyond a reasonable doubt is unavoidable.

These changes would bring the procedure in New Zealand into line with the approach adopted in other jurisdictions and offer the prospect of rendering the s9 hearing a more rational and comprehensible procedure within the broader framework of determining unfitness to stand trial.

NILQ 65(2): 229–42

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* xii–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 *Brawner*, 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 *appreciate* 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 Acquittees' (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 Callaham 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 Pohlot* 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 Correra* 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); *Zamora 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 Cir1983) (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. *Hendershott 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 *Brawner* (n 15).

56 See ibid 1002.

57 See, for example, *Regina v Wallett* [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); *Brawner* (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.

NILQ 65(2): 243–58

Criminal law and neuroscience: present and future

STEPHEN J MORSE JD, PHD

*Ferdinand Wakeman Hubbell Professor of Law and Professor of Psychology and Law in Psychiatry, and Associate Director, Center for Neuroscience and Society, University of Pennsylvania**

1 Introduction

In a 2002 editorial published in *The Economist*, the following warning was given: 'Genetics may yet threaten privacy, kill autonomy, make society homogeneous and gut the concept of human nature. But neuroscience could do all of these things first.'[1] But neither genetics nor any other science that was predicted to revolutionise the law, including behavioural psychology, sociology and psychodynamic psychology, to name but a few, has had this effect. This will also be true of neuroscience, which is simply the newest science on the block. Neuroscience is not going to do the terrible things *The Economist* fears, at least not in the foreseeable future. Neuroscience has many things to say, but not nearly as much as people would hope, especially in relation to criminal law. At most, in the near to intermediate term, neuroscience may make modest contributions to legal policy and case adjudication. Nonetheless, there has been irrational exuberance about the potential contribution of neuroscience, an issue I have addressed previously and referred to as 'Brain Overclaim Syndrome'.[2]

I first address the law's motivation and the motivation of some advocates to turn to science to solve the very hard normative problems that law addresses. The next part discusses the law's psychology and its concepts of the person and responsibility. Then I consider the general relation of neuroscience to law, which I characterise as the issue of 'translation'. The following part canvasses various distractions that have bedevilled clear thinking about the relation of scientific, causal accounts of behaviour to responsibility. Next, I examine the limits of neurolaw and consider why neurolaw does not pose a genuinely radical challenge to the law's concepts of the person and responsibility. The penultimate part

* Member, MacArthur Foundation Research Network on Law and Neuroscience; Diplomate, American Board of Professional Psychology (Forensic). I thank Ed Greenlee for his invaluable help. Jakob Elster and Michael Moore provided invaluable insights. An earlier version of this article was presented at a symposium, 'The Brain Sciences in the Courtroom' 22 October 2010, Mercer University, Walter F George School of Law. It was originally published as, 'Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty' (2011) 62 Mercer Law Review 837. It is published here in substantially altered form.

1 'The Ethics of Brain Sciences: Open Your Mind' *The Economist*, 23 May 2002, 77, <www.economist.com/node/1143317/print> accessed 5 October 2011.

2 Stephen J Morse, 'Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note' (2006) 3 Ohio State Journal of Criminal Law 397; 'Brain Overclaim Redux' (2013) XXXI Law and Inequality 509.

makes a case for cautious optimism about the contribution that neuroscience may make to law in the near and intermediate term. A brief conclusion follows. Throughout, I use the criminal law for my examples because the relation of neuroscience to criminal law, especially criminal responsibility, has most captured the public and legal imagination.

2 The source of neuro-exuberance

Everyone understands that legal issues are normative, addressing how we should regulate our lives in a complex society. How do we live together? What are the duties we owe each other? For violations of those duties, when is the state justified in imposing the most afflictive – but sometimes justified – exercises of state power, criminal blame and punishment?[3] When should we do this, to whom, and how much?

Virtually every legal issue is contested – consider criminal responsibility, for example – and there is always room for debate about policy, doctrine and adjudication. In a recent book, Professor Robin Feldman has argued that law lacks the courage forthrightly to address the difficult normative issues that it faces.[4] The law therefore adopts what Feldman terms an 'internalising' and an 'externalising' strategy for using science to try to avoid the difficulties. In the internalising strategy, the law adopts scientific criteria as legal criteria. A futuristic example might be using neural criteria for criminal responsibility. In the externalising strategy, the law turns to scientific or clinical experts to make the decision. An example would be using forensic clinicians to decide whether a criminal defendant is competent to stand trial and then simply rubberstamping the clinician's opinion. Neither strategy is successful because each avoids facing the hard questions and impedes legal evolution and progress. Professor Feldman concludes, and I agree, that the law does not err by using science too little, as is commonly claimed. Rather, it errs by using it too much because the law is too insecure about its resources and capacities to do justice.

A fascinating question is why so many enthusiasts seem to have extravagant expectations about the contribution of neuroscience to law, especially criminal law. Here is my speculation about the source. Many people intensely dislike the concept and practice of retributive justice, thinking that they are prescientific and harsh. Their hope is that the new neuroscience will convince the law at last that determinism is true, no offender is genuinely responsible, and the only logical conclusion is that the law should adopt a consequentially-based prediction/prevention system of social control guided by the knowledge of the neuroscientist-kings who will finally have supplanted the platonic philosopher-kings.[5] On a more modest level, many advocates think that neuroscience may not revolutionise criminal justice, but neuroscience will demonstrate that many more offenders should be excused and do not deserve the harsh punishments imposed by the United States criminal justice system. Four decades ago, our criminal justice system would have been using psychodynamic psychology for the same purpose. More recently, genetics has been employed in a similar manner. The impulse, however, is clear: jettison desert, or at least mitigate judgments of desert. As will be shown below, however, these advocates often adopt an untenable theory of mitigation or an excuse that quickly collapses into the nihilistic conclusion that no one is really criminally responsible.

3 See e.g. *In re Winship* 397 US 358, 364 (1970) (holding that due process requires that every conviction be supported by proof beyond reasonable doubt as to every element of the crime).

4 Robin Feldman, *The Role of Science in Law* (Oxford University Press 2009).

5 Joshua Greene and Jonathan Cohen, 'For the Law, Neuroscience Changes Nothing and Everything' in Samir Zeki and Oliver Goodenough (eds), *Law and the Brain* (Oxford University Press 2006) 217–18, 224.

3 The law's psychology, concept of the person and responsibility

Criminal law presupposes a 'folk psychological' view of the person and behaviour. This psychological theory explains behaviour in part by mental states such as desires, beliefs, intentions, willings, and plans. Biological and other psychological and sociological variables also play a causal role, but folk psychology considers mental states fundamental to a full causal explanation and understanding of human action. Lawyers, philosophers and scientists argue about the definitions of mental states and theories of action, but that does not undermine the general claim that mental states are fundamental. Indeed, the arguments and evidence that disputants use to convince others presuppose the folk-psychological view of the person. Brains do not convince each other, people do. Folk psychology presupposes only that human action will at least be rationalisable by mental state explanations or will be responsive to reasons – including incentives – under the right conditions.

For example, the folk-psychological explanation for why you are reading this article is, roughly, that you desire to understand the relation of neuroscience to criminal responsibility or to law generally. You believe that reading the article will help fulfil that desire, so you formed the intention to read it. This is a practical, rather than a deductive, syllogism. Brief reflection should indicate that the law's psychology must be a folk-psychological theory, a view of the person as a conscious – and potentially self-conscious – creature who forms and acts on intentions that are the product of the person's other mental states. We are the sort of creatures who can act for and respond to reasons. The law treats persons generally as intentional creatures and not simply as mechanistic forces of nature.

Law is primarily action-guiding and is not able to guide people directly and indirectly unless people are capable of using rules as premises in their reasoning about how they should behave. Unless people could be guided by law, it would be useless (and perhaps incoherent) as an action-guiding system of rules. Legal rules are action-guiding primarily because these rules provide an agent with good moral or prudential reasons for forbearance or action. Human behaviour can be modified by means other than influencing deliberation, and human beings do not always deliberate before they act. Nonetheless, the law presupposes folk psychology even when we most habitually follow the legal rules. Unless people are capable of understanding and then using legal rules to guide their conduct, the law is powerless to affect human behaviour.

The legal view of the person does not hold that people must always reason or consistently behave rationally according to some pre-ordained, normative notion of rationality. Rather, the law's view is that people are capable of acting for reasons and are capable of minimal rationality according to predominantly conventional, socially constructed standards. The type of rationality the law requires is the ordinary person's common-sense view of rationality, not the technical notion that might be acceptable within the disciplines of economics, philosophy, psychology, computer science and the like.

Virtually everything for which agents deserve to be praised, blamed, rewarded, or punished is the product of mental causation and, in principle, is responsive to reasons, including incentives. Machines may cause harm, but they cannot do wrong, and they cannot violate expectations about how people ought to live together. Machines do not deserve praise, blame, reward, punishment, concern, or respect because they exist or because of the results they cause. Only people, intentional agents with the potential to act, can do wrong and violate expectations of what they owe each other.

Many scientists and some philosophers of mind and action might consider folk psychology to be a primitive or prescientific view of human behaviour. For the foreseeable future, however, the law will be based on the folk-psychological model of the person and

behaviour described. Until and unless scientific discoveries convince us that our view of ourselves is radically wrong, the basic explanatory apparatus of folk psychology will remain central. It is vital that we do not lose sight of this model lest we fall into confusion when various claims based on neuroscience are made. If any science is to have appropriate influence on current criminal law and legal decision-making, the science must be relevant to and translated into the law's folk-psychological framework.

All of the law's doctrinal criteria for criminal responsibility are folk-psychological. Begin with the definitional criteria, the 'elements' of crime. The 'voluntary' act requirement is defined, roughly, as an *intentional* bodily movement – or omission in cases in which the person has a duty to act – done in a reasonably integrated state of consciousness. Other than crimes of strict liability, all crimes also require a culpable mental state, such as purpose, knowledge or recklessness. All affirmative defences of justification and excuse involve an inquiry into the person's mental state, such as the belief that self-defensive force was necessary or the lack of knowledge of right from wrong.

Our folk-psychological concepts of criminal responsibility follow logically from the action-guiding nature of law itself, from its folk-psychological concept of the person and action, and from the aim of achieving retributive justice, which holds that no one should be punished unless they deserve it and no more than they deserve. The general capacity for rationality is the primary condition for responsibility, and the lack of that capacity is the primary condition for excusing a person. If human beings were not rational creatures who could understand the good reasons for action and were not capable of conforming to legal requirements through intentional action or forbearance, the law could not adequately guide action and would not be just. Legally responsible agents are therefore people who have the general capacity to grasp and be guided by good reason in particular legal contexts.[6]

In cases of excuse, the agent who has done something wrong acts for a reason but is either incapable of rationality generally or incapable on the specific occasion in question. This explains, for example, why young children and some people with mental disorders are not held responsible. The amount of lack of capacity for rationality that is necessary to find the agent not responsible is a moral, social, political and, ultimately, legal issue. It is not a scientific, medical, psychological, or psychiatric issue.

Compulsion or coercion is also an excusing condition. Literal compulsion exists when the person's bodily movement is a pure mechanism that is not rationalisable by reference to the agent's mental states. These cases defeat the requirement of a 'voluntary act'. For example, a tremor or spasm produced by a neurological disorder is not an action because it is not intentional and, therefore, defeats the ascription of a voluntary act. Metaphorical compulsion exists when an agent acts intentionally but in response to some hard choice imposed on the agent through no fault of his or her own. For example, if a miscreant holds a gun to an agent's head and threatens to kill her unless she kills another innocent person, it would be wrong to kill under these circumstances. Nevertheless, the law may decide as a normative matter to excuse the act of intentional killing because the agent was motivated by a threat so great that it would be supremely difficult for most citizens to resist. Cases involving internal compulsive states are more difficult to conceptualise because it is difficult to define and assess 'loss of control'.[7] The cases that most fit this category are 'disorders of desire', such as addictions and sexual disorders. The question is why these acting agents lack control, but other people with strong desires do not. If people frequently yield to their apparently very

6 I adapt the felicitous phrase 'to grasp and be guided by good reason' from Jay Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press 1994) 86.

7 Stephen J Morse, 'Uncontrollable Urges and Irrational People' (2002) 88 Virginia Law Review 1025, 1035.

strong desires at great social, medical, occupational, financial, and legal cost to themselves, agents will often say they could not help themselves, they were not in control, and an excuse or mitigation is therefore warranted. But why mitigation or excuse should obtain is difficult to understand.

4 Lost in translation? Legal relevance and the need for translation

What in principle is the possible relation of neuroscience to law? We must begin with a distinction between internal relevance and external relevance. An internal contribution or critique accepts the general coherence and legitimacy of a set of legal doctrines, practices or institutions and attempts to explain or alter them. For example, an internal contribution to criminal responsibility may suggest the need for doctrinal reform of, say, the insanity defence, but it would not suggest that the notion of criminal responsibility is itself incoherent or illegitimate. By contrast, an externally relevant critique suggests that the doctrines, practices or institutions are incoherent, illegitimate or unjustified. Because a radical, external critique has little possibility of success at present (as is explained below), I make the simplifying assumption that the contributions of neuroscience will be internal and thus will need to be translated into the law's folk-psychological concepts.

The law's criteria for responsibility and competence are essentially behavioral – acts and mental states. The criteria of neuroscience are mechanistic – neural structure and function. Is the apparent chasm between those two types of discourse bridgeable? This is a familiar question in the field of mental health law,[8] but there is even greater dissonance in neurolaw. Psychiatry and psychology sometimes treat behaviour mechanistically, sometimes treat it folk-psychologically, and sometimes blend the two. In many cases, the psychological sciences are quite close to folk psychology in approach. Neuroscience, in contrast, is purely mechanistic and eschews folk-psychological concepts and discourse. Neurons and neural networks do not act intentionally for reasons. They have no sense of past, present and future, and no aspirations. They do not recognise that they will die. Thus, the gap will be harder to bridge.

The brain does enable the mind (even if we do not know how this occurs). Therefore, facts we learn about brains in general or about a specific brain could in principle provide useful information about mental states and about human capacities in general and in specific cases. Some believe that this conclusion is a category error.[9] This is a plausible view, and perhaps it is correct. If it is, then the whole subject of neurolaw is empty, and there was no point writing this article in the first place. Let us therefore bracket this pessimistic view and determine what follows from the more optimistic position that what we learn about the brain and nervous system can be potentially helpful to resolving questions of criminal responsibility if the findings are properly translated into the law's psychological framework.

The question is whether the new neuroscience is legally relevant because it makes a proposition about responsibility or competence more or less likely to be true. Any legal criterion must be established independently, and biological evidence must be translated into the criminal law's folk-psychological criteria. That is, the expert must be able to explain precisely how the neuroevidence bears on whether the agent acted, formed the required *mens rea*, or met the criteria for an excusing condition. In the context of competence evaluations, the expert must explain precisely how the neuroevidence bears on whether the subject was

8 See e.g. Alan A Stone, *Law, Psychiatry, and Morality* (American Psychiatric Press 1984) 95–96.

9 See e.g. Max R Bennett and Peter M S Hacker, *Philosophical Foundations of Neuroscience* (Blackwell 2003) 112, 270, 360, 381; Michael S Pardo and Dennis Patterson, *Minds, Brains and Law: The Conceptual Foundations of Law and Neuroscience* (Oxford University Press 2013)' and 'Philosophical Foundations of Law and Neuroscience' (2010) University of Illinois Law Review 1211.

capable of meeting the law's functional criteria. If the evidence is not directly relevant, the expert should be able to explain the chain of inference from the indirect evidence to the law's criteria. At present, as I explain below, few such data exist, but neuroscience is advancing so rapidly that such data may exist in the near or medium term. Moreover, the argument is conceptual and does not depend on any particular neuroscience findings.

5 Dangerous distractions concerning neuroscience and criminal responsibility and competence

This part considers a number of related issues that are often thought to be relevant to criminal responsibility and competence but that are in fact irrelevant, confusing and distracting: free will, causation as an excuse, causation as compulsion, prediction as an excuse, dualism, and the non-efficacy of mental states. It is important to correct these errors because much of the unjustified legal exuberance about the contributions of neurolaw flow from them. The legal exuberance also flows, however, from unrealistic expectations about the scientific accomplishments of neuroscience. A later part of this article addresses the scientific exuberance.

Contrary to what many people believe and what judges and others sometimes say, free will is not a legal criterion that is part of any doctrine, and it is not even foundational for criminal responsibility.[10] Criminal law doctrines are fully consistent with the truth of determinism or universal causation that allegedly undermines the foundations of responsibility. Even if determinism is true, some people act and some people do not. Some people form prohibited mental states and some do not. Some people are legally insane or act under duress when they commit crimes, but most defendants are not legally insane or acting under duress. Moreover, these distinctions matter to moral and legal theories of responsibility and fairness that we have reason to endorse. Thus, law addresses problems genuinely related to responsibility, including consciousness, the formation of mental states such as intention and knowledge, the capacity for rationality, and compulsion. The law, however, never addresses the presence or absence of free will.

When most people use the term 'free will' in the context of legal responsibility, they are typically using it loosely as a synonym for the conclusion that the defendant was or was not criminally responsible. They typically have reached this conclusion for reasons that do not involve free will – for example, that the defendant was legally insane or acted under duress – but such use of the term, free will, only perpetuates misunderstanding and confusion. Once the legal criteria for excuse have been met – and no excuse includes lack of free will as a criterion – the defendant will be excused without any reference whatsoever to free will as an independent ground for excuse.

There is a genuine metaphysical problem regarding free will, which is whether human beings have the capacity to act uncaused by anything other than themselves and whether this capacity is a necessary foundation for holding anyone legally or morally accountable for criminal conduct. Philosophers and others have debated these issues in various forms for millennia and there is no resolution in sight. Indeed, some people might think that the problem is insoluble. This is a philosophical issue, but it is not a problem for the law, and neuroscience raises no new challenge to this conclusion. Solving the free will problem would have profound implications for responsibility doctrines and practices, such as blame and punishment, but having or lacking libertarian freedom is not a criterion of any civil or criminal law doctrine.

10 Stephen J Morse, 'The Non-Problem of Free Will in Forensic Psychiatry and Psychology' (2007) 25 Behavioral Sciences and the Law 203, 204.

Neuroscience is simply the most recent, mechanistic causal science that appears deterministically to explain behaviour. Neuroscience thus joins social structural variables, behaviourism, genetics and other scientific explanations that have also been deterministic explanations for behaviour. In principle, however, neuroscience adds nothing new, even if neuroscience is a better, more persuasive science than some of its predecessors. No science, including neuroscience, can demonstrate that libertarian free will does or does not exist. As long as free will in the strong sense is not foundational for just blame and punishment and is not a criterion at the doctrinal level – which it is not – the truth of determinism or universal causation poses no threat to legal responsibility. Neuroscience may help shed light on folk-psychological excusing conditions, such as automatism or legal insanity, but the truth of determinism is not an excusing condition. The law will be fundamentally challenged only if neuroscience or any other science can conclusively demonstrate that the law's psychology is wrong, and that we are not the type of creatures for whom mental states are causally effective. This is a different question from whether determinism undermines responsibility, however, and this article returns to it below.

A related confusion is that behaviour is excused if it is caused, but causation *per se* is not a legal or moral mitigating or excusing condition. I termed this confusion the 'fundamental psycholegal error'.[11] At most, causal explanations can only provide evidence concerning whether a genuine excusing condition, such as lack of rational capacity, was present. For example, suppose a life marked by poverty and abuse played a predisposing causal role in a defendant's criminal behaviour or that an alleged new mental syndrome played a causal role in explaining criminal conduct. The claim is often made that such causes – for which the agent is not responsible – should be an excusing or mitigating position *per se*, but this claim is false.

All behaviour is the product of the necessary and sufficient causal conditions without which the behaviour would not have occurred, including brain causation, which is always part of the causal explanation for any behaviour. If causation were an excusing condition *per se*, then no one would be responsible for any behaviour. Some people might welcome such a conclusion and believe that responsibility is impossible, but this is not the legal and moral world we inhabit. The law holds most adults responsible for most of their conduct, and genuine excusing conditions are limited. Thus, unless the person's history or mental condition, for example, provides evidence of an existing excusing or mitigating condition, such as lack of rational capacity, there is no reason for excuse or mitigation.

Even a genuinely abnormal cause is not *per se* an excusing condition. For example, imagine an armed robber who suffers from intermittent hypomania and who only robs when he is clinically hypomanic because only then does he feel sufficiently energetic and confident. In other words, the hypomania is a 'but for' cause of his robberies. Nevertheless, he would not be excused for an armed robbery because hypomania seldom compromises rational capacity sufficiently to warrant an excuse. If he committed an armed robbery under the influence of a delusional belief his mania produced, then he might be excused by reason of legal insanity. In that case, the excusing condition would be compromised rationality and not the mania *per se*. In short, a neuroscientific causal explanation for criminal conduct, like any other type of causal explanation, does not *per se* mitigate or excuse. It only provides evidence that might help the law resolve whether a genuine excuse existed, or it may in the future provide data that might be a guide to prophylactic or rehabilitative measures.

11 Stephen J Morse, 'Culpability and Control' (1994) 142 University of Pennsylvania Law Review 1587, 1592–94.

Compulsion is a genuine mitigating or excusing condition, but causation – including brain causation – is not the equivalent of compulsion. Compulsion may be either literal or metaphorical and normative. It is crucial to recognise that most human action is not plausibly the result of either type of compulsion, but all human behaviour is caused by its necessary and sufficient causes – including brain causation. Even abnormal causes are not necessarily compelling. To illustrate, suppose that a person has weak paedophilic urges and weak sexual urges in general. If this person molested a child there would be no ground for a compulsion excuse. If causation was the equivalent of compulsion, all behaviour would be compelled and no one would be responsible. Once again, this is not a plausible account of the law's responsibility conditions. Causal information from neuroscience might help us resolve questions concerning whether legal compulsion existed, or it might be a guide to prophylactic or rehabilitative measures when dealing with plausible legal compulsion. Causation, however, is not *per se* compulsion.

Causal knowledge, whether from neuroscience or any other science, can enhance the accuracy of behavioural predictions, but predictability is also not a *per se* excusing or mitigating condition – even if the predictability of the behaviour is perfect. To understand this, consider how many things we do that are perfectly predictable but for which there is no plausible excusing or mitigating condition. If the variables that enhance prediction also produce a genuine excusing or mitigating condition, then excuse or mitigation is justified for the latter reason and independent of the prediction.

For example, recent research demonstrates that a history of childhood abuse coupled with a specific, genetically produced enzyme abnormality that produces a neurotransmitter deficit vastly increases the risk that a person will behave antisocially as an adolescent or young adult.[12] Does this mean that an offender with this gene by environment interaction is not responsible or less responsible? No. The offender may not be fully responsible or responsible at all, but not because there is a causal explanation. What is the intermediary excusing or mitigating principle? Are these people, for instance, more impulsive? Are they lacking rationality? What is the actual excusing or mitigating condition?

Again, causation is not compulsion, and predictability is not an excuse. Just because an offender is caused to do something or is predictable does not mean that the offender was compelled to do the crime charged or is otherwise not responsible. Brain causation – or any other kind of causation – does not mean that we are automatons, not really acting agents at all, or otherwise excused.

Most informed people are not 'dualists' concerning the relation between the mind and the brain. That is, they no longer think that our minds – or souls – are independent of our brains and bodies more generally and can somehow exert a causal influence over our bodies. It may seem as if law's emphasis on the importance of mental states as causing behaviour is based on a prescientific, outmoded form of dualism, but this is not the case. Although the brain enables the mind, we have no idea how this occurs and have no idea how action is possible.[13] It is clear that, at the least, mental states are dependent upon or supervene on brain states, but neither neuroscience nor any other science has demonstrated that mental states do not play an independent and partial causal role.

Despite our lack of understanding of the mind–brain–action relation, some scientists and philosophers question whether mental states have any causal effect, thus treating mental

12 See e.g. Avshalom Caspi et al, 'Role of Genotype in the Cycle of Violence in Maltreated Children' (2002) 297 Science 851. Indeed, the risk is nine times higher.

13 Paul R McHugh and Phillip R Slavney, *The Perspectives of Psychiatry* 2nd edn (Johns Hopkins University Press 1998) 11–12.

states as psychic appendixes that evolution has created but that have no genuine function. These claims are not strawpersons. They are made by serious, thoughtful people.[14] As discussed below, if accepted, they would create a complete and revolutionary paradigm shift in the law of criminal responsibility and competence (and more widely). Thus, this claim is an external critique and must be understood as such. Moreover, given our current state of knowledge, there is little scientific or conceptual reason to accept it.[15]

In conclusion, legal actors concerned with criminal law policy, doctrine and adjudication must always keep the folk-psychological view present in their minds when considering claims or evidence from neuroscience, and must always question how the science is legally relevant to the law's action and mental states criteria. The truths of determinism, causation and predictability do not in themselves answer any doctrinal or policy issue.

6 The limits of neurolaw: the present limits of neuroscience

Most generally, the relation of brain, mind and action is one of the hardest problems in all science. Again, we have no idea how the brain enables the mind or how action is possible.[16] The brain–mind–action relation is a mystery. For example, we would like to know the difference between a neuromuscular spasm and intentionally moving one's arm in exactly the same way. The former is a purely mechanical motion, whereas the latter is an action, but we cannot explain the difference between the two. We know that a functioning brain is a necessary condition for having mental states and for acting. After all, if your brain is dead, you have no mental states, are not acting, and indeed are not doing much of anything at all. Still, we do not know how mental states and action are caused.

Despite the astonishing advances in neuroimaging and other neuroscientific methods, we still do not have sophisticated causal knowledge of how the brain works generally and we have little information that is legally relevant. This is unsurprising. The scientific problems are fearsomely difficult. Only in the last decade have researchers begun to accumulate much data from functional magnetic resonance imaging (fMRI), which is the technology that has generated most of the legal interest. Moreover, virtually no studies have been performed to address specifically legal questions.

Before turning to the specific reasons for neuromodesty, a few preliminary points of general applicability must be addressed. The first and most important is contained in the message of the prior part. Causation by biological variables, including abnormal biological variables, does not *per se* create an excusing or mitigating condition. Any excusing condition must be established independently. The goal is always to translate the biological evidence into the criminal law's folk-psychological criteria.

Assessing criminal responsibility involves a retrospective evaluation of the defendant's mental states at the time of the crime. No criminal wears a portable scanner or other neurodetection device that provides a measurement at the time of the crime, at least not yet. Further, neuroscience is insufficiently developed to detect specific, legally relevant mental content or to provide a sufficiently accurate diagnostic marker for even a severe mental

14 See e.g. Greene and Cohen (n 5) 217, 218.

15 Stephen J Morse, 'Lost in Translation? An Essay on Law and Neuroscience' in Michael Freeman (ed), *Law and Neuroscience* (Oxford University Press 2011) 543–54.

16 McHugh and Slavney (n 13).

disorder.[17] Nonetheless, certain aspects of neural structure and function that bear on legally relevant capacities, such as the capacity for rationality and control, may be temporally stable in general or in individual cases. If they are, neuroevidence may permit a reasonably valid retrospective inference about the defendant's rational and control capacities and their impact on criminal behaviour. This will of course depend on the existence of adequate science to do this. We currently lack such science, but future research may provide the necessary data.

Questions concerning competence or predictions of future behaviour are based on a subject's present condition. Thus, the problems besetting the retrospective responsibility analysis do not apply to such issues. The criteria for competence are functional. They ask whether the subject can perform some task – such as understanding the nature of a criminal proceeding or understanding a treatment option that is offered – at a level the law considers normatively acceptable to warrant respecting the subject's choice and autonomy.

Now, let us consider the specific grounds for neuromodesty in cognitive, affective and social neuroscience, the sub-disciplines most relevant to law. At present, most neuroscience studies on human beings involve very small numbers of subjects, although this phenomenon is starting to change. Most of the studies have been done on college and university students, who are hardly a random sample of the population generally and of criminal offenders specifically. There is also a serious question of whether findings based on subjects' behaviour and brain activity in a scanner would apply to real world situations. Further, most studies average the neurodata over the subjects, and the average finding may not accurately describe the brain structure or function of any actual subject in the study. Replications are few, which is especially important for law. Policy and adjudication should not be influenced by findings that are insufficiently established, and replications of findings are crucial to our confidence in a result. Finally, the neuroscience of cognition and interpersonal behaviour is largely in its infancy and what is known is quite coarse-grained and correlational, rather than fine-grained and causal.[18] What is being investigated is an association between a condition or a task in the scanner and brain activity. These studies do not demonstrate that the brain activity is a sensitive diagnostic marker for the condition or either a necessary, sufficient or predisposing causal condition for the behavioural task that is being done in the scanner. Any language that suggests otherwise – such as claiming that some brain region is the neural substrate for the behaviour – is simply not justifiable based on the methodology of most studies. Moreover, activity in the same region may be associated with diametrically opposite behavioural phenomena – for example, love and hate.

There are also technical and research design difficulties. It takes many mathematical transformations to get from the raw fMRI data to the images of the brain that are increasingly familiar. Explaining these transformations is beyond me, but I do understand that the likelihood that an investigator will find a statistically significant result depends on how the researcher sets the threshold for significance. There is dispute about this, and the threshold levels are conventional. If the threshold changes, so does the outcome. I have been convinced by neuroscience colleagues that many such technical difficulties have largely been

17 Allen Frances, 'Whither DSM-V?' (2009) 195 British Journal of Psychiatry 391. Many studies do find differences between patients with mental disorders and controls, but the differences are too small to be used diagnostically. But see, generally, John P A Ioannidis, 'Excess Significance Bias in the Literature on Brain Volume Abnormalities' (2011) 68 Archives of General Psychiatry 773 (claiming, based on a meta-analysis of studies of brain volume abnormalities in patients with mental disorders, that many more studies than should be expected found statistically significant results and that this can be best explained by bias in the reporting of the data).

18 See e,g. Gregory A Miller, 'Mistreating Psychology in the Decades of the Brain' (2010) 5 Perspectives on Psychological Science 716 (providing a cautious, thorough overview of the scientific and practical problems facing cognitive and social neuroscience).

solved, but research design and potentially unjustified inferences from the studies are still an acute problem. It is extraordinarily difficult to control for all conceivable artefacts. Consequently, there are often problems of over-inference. Finally, it is also an open question whether accurate inferences or predictions about individuals are possible using group data when that group includes the individual.[19] This is a very controversial topic, but even if it is difficult or impossible now, it may become easier in the future. Over time, however, all these problems may ease as imaging and other techniques become less expensive and more accurate, research designs become more sophisticated, and the sophistication of the science increases generally.

Virtually all neuroscience studies of potential interest to the law involve some behaviour that has already been identified as of interest, and the point of the study is to identify that behaviour's neural correlates. Neuroscientists do not go on general 'fishing' expeditions.[20] There is usually some bit of behaviour – such as addiction, schizophrenia or impulsivity – that investigators would like to understand better by investigating its neural correlates. To do this properly presupposes that the researchers have already identified and validated the behaviour under neuroscientific investigation. Thus, neurodata can be no more valid than the behaviour with which it is correlated.

On occasion, the neuroscience might suggest that the behaviour is not well-characterised or is neurally indistinguishable from other, seemingly different behaviour. In general, however, the existence of legally relevant behaviour will already be apparent before the neuroscientific investigation is begun. For example, some people are grossly out of touch with reality. If, as a result, they do not understand right from wrong, we excuse them because they lack such knowledge. We might learn a great deal about the neural correlates of such psychological abnormalities, but we already knew without neuroscientific data that these abnormalities existed, and we had a firm view of their normative significance. In the future, however, we may learn more about the causal link between the brain and behaviour, and studies may be devised that are more directly legally relevant. I suspect that we are unlikely to make substantial progress with neural assessment of legally relevant mental content, but we are likely to learn more about capacities that will bear on excuse or mitigation.

The criteria for both responsibility and competence are behavioral; therefore, actions speak louder than images. This is a truism for all criminal responsibility and competence assessments. If the finding of any test or measurement of behaviour is contradicted by actual behavioural evidence, then we must believe the behavioural evidence because it is more direct and probative of the law's behavioural criteria. For example, if the person behaves rationally in a wide variety of circumstances, the agent is rational even if the brain appears structurally or functionally abnormal. We confidently knew that some people were behaviourally abnormal – such as being psychotic – long before there were any psychological or neurological tests for such abnormalities.

An analogy from physical medicine may be instructive. Suppose someone complains about back pain, a subjective symptom, and the question is whether the subject actually does have back pain. We know that many people with abnormal spines do not experience back

19 David L Faigman, John Monahan and Christopher Slobogin, 'Group to Individual Inference (G2i) in Scientific Expert Testimony' (2014) 81 U Chicago Law Review (forthcoming).

20 For an amusing exception, see Craig M Bennett, Abigail A Baird, Michael B Miller, and George L Wolford, 'Neural Correlates of Interspecies Perspective Taking in the Post-mortem Atlantic Salmon: An Argument for Multiple Comparisons Correction' (2009) 1 Journal of Serendipitous and Unexpected Results 1. Available at, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161.8384&rep=rep1&type=pdf>. The study scanned a dead Atlantic salmon to demonstrate that significant results can be obtained from the most unpromising investigation unless the research design properly controls for chance findings (false positives).

pain, and many people who complain of back pain have normal spines. If the person is claiming a disability and the spine looks dreadful, evidence that the person regularly exercises on a trampoline without difficulty indicates that there is no disability caused by back pain. If there is reason to suspect malingering, however, and there is not clear behavioural evidence of lack of pain, then a completely normal spine might be of use in deciding whether the claimant is malingering. Unless the correlation between the image and the legally relevant behaviour is very powerful, however, such evidence will be of limited help.

If actions speak louder than images, however, what room is there for introducing neuroevidence in legal cases? Let us begin with cases in which the behavioural evidence is clear and permits an equally clear inference about the defendant's mental state. For example, lay people may not know the technical term to apply to people who are manifestly out of touch with reality, but they will readily recognise this unfortunate condition. No further tests of any sort will be necessary to prove that the subject suffers from seriously impaired rationality. In such cases, neuroevidence will be at most convergent and increase our confidence in what we already had confidently concluded. Determining if it is worth collecting the neuroevidence will depend on whether the cost–benefit analysis justifies obtaining convergent evidence.

Roper v Simmons is the most striking example of a case in which the behavioural evidence was clear.[21] In *Roper* the US Supreme Court categorically excluded the death penalty for capital murderers who killed when they were 16 or 17 years old on the grounds that adolescents do not deserve the death penalty.[22] The amicus briefs were replete with neuroscience data showing that the brains of late adolescents are not fully biologically mature, and advocates used this data to suggest that adolescent killers could not be fairly put to death.[23] Now, we already knew from common-sense observation and from rigorous behavioural studies that juveniles are on average less rational than adults. What did the neuroscientific evidence about the juvenile brain add? It was consistent with the undeniable behavioural data and perhaps provided a partial causal explanation of the behavioural differences. The neuroscience data was therefore merely additive and only indirectly relevant, and the Supreme Court did not cite it, except perhaps by implication when it referred vaguely to 'other' scientific evidence.[24]

Whether adolescents are sufficiently less rational on average than adults to exclude them categorically from the death penalty is a normative legal question and not a scientific or psychological question. Advocates claimed, however, that the neuroscience confirmed that adolescents are insufficiently responsible to be executed,[25] thus confusing the positive and

21 *Roper v Simmons* 543 US 551 (2005).

22 Ibid 578–79.

23 Ibid 569.

24 Ibid 569, 573. The Supreme Court referred generally to other science, but it was not clear whether neuroscience played a specific role. The Supreme Court did cite neuroscientific findings in *Graham v Florida* 130 S Ct 2011 (2010), which categorically excluded juveniles from life without the possibility of parole in non-homicide cases (at 2034) and in *Miller v Alabama* 132 S Ct 2455 (2013), which held that the sentence of life without possibility of parole was constitutional for juveniles who committed homicide crimes but that it was unconstitutional to impose this penalty mandatorily (at 2460). In both cases, the citation was conclusory and generally non-specific, and I believe it was dictum. The Supreme Court was responding in *Graham* to an argument that no party had seriously made, which was that the science of adolescent development had changed significantly since *Roper* was decided. Also in *Miller*, the court drew a distinction between social science and 'science' (at 2464, n. 5). Social science, like neuroscience, is science (and arguably more directly relevant to legal criteria for the reasons this article has discussed). The important distinctions are between good and bad science and legally relevant and legally irrelevant science.

25 *Roper v Simmons* (n 21) 569.

the normative. The neuroscience evidence in no way independently confirms that adolescents are less responsible. If the behavioural differences between adolescents and adults were slight, it would not matter if their brains were quite different. Similarly, if the behavioural differences were sufficient for moral and constitutional differential treatment, then it would not matter if the brains were essentially indistinguishable.

If the behavioural data are not clear, then the potential contribution of neuroscience is large. Unfortunately, it is in just such cases that neuroscience at present is not likely to be of much help. I term the reason for this the 'clear cut' problem.[26] Recall that neuroscientific studies usually start with clear cases of well-characterised behaviour. In such cases, the neural markers might be quite sensitive to the already clearly identified behaviours precisely because the behaviour is so clear. Less clear behaviour is simply not studied, or the overlap in data about less clear behaviour is greater between experimental and control subjects. Thus, the neural markers of clear cases will provide little guidance to resolve behaviourally ambiguous cases of legally relevant behaviour and they are unnecessary if the behaviour is sufficiently clear.

For example, suppose that in an insanity defence case the question is whether the defendant suffers from a major mental disorder, such as schizophrenia. In extreme cases, the behaviour will be clear, and no neurodata will be necessary. Investigators have discovered various small but statistically significant differences in neural structure or function between people who are clearly suffering from schizophrenia and those who are not.[27] Nonetheless, in a behaviourally unclear case, the overlap between data on the brains of people with schizophrenia and people without the disorder is so great that a scan is insufficiently sensitive to be used for diagnostic purposes. In short, at present, in those cases in which the neuroscience would be most helpful, it has little to contribute. Again, this situation may change if neural markers become more diagnostically sensitive for legally relevant criteria.

Some people think that executive capacity – the congeries of cognitive and emotional capacities that help to plan and regulate human behaviour – is going to be the Holy Grail to help the law determine an offender's true culpability. After all, there is an attractive moral case that people with a substantial lack of these capacities are less culpable, even if their conduct satisfied the *prima facie* case for the crime charged. Perhaps neuroscience can provide specific data previously unavailable to identify executive capacity differences more precisely.

There are two problems, however. First, significant problems with executive capacity are readily apparent without testing, and criminal law simply will not adopt fine-grained culpability criteria. Second, the correlation between neuropsychological tests of executive capacity and actual real world behaviour is not terribly strong.[28] Only a small fraction of the variance is accounted for, and the scanning studies will use the types of tasks the tests use. Consequently, we are far from able to use neuroscience accurately to assess non-obvious executive capacity differences that are valid in real world contexts.

7 The radical neurochallenge: are we victims of neuronal circumstances?

This part addresses the claim and hope alluded to earlier that neuroscience will cause a paradigm shift in criminal responsibility by demonstrating that we are 'merely victims of neuronal circumstances' (or some similar claim that denies human agency). This claim holds that we are not the kinds of intentional creatures we think we are. If our mental states play

26 Morse (n 15) 540.

27 On the other hand, there may be reason to be cautious about such findings. See generally Ioannidis (n 17).

28 See e.g. Russell A Barkley and Kevin R Murphy, 'Impairment in Occupational Functioning and Adult ADHD: The Predictive Utility of Executive Function (EF) Ratings versus EF Tests' (2010) 25 Archives of Clinical Neuropsychology 157.

no role in our behaviour and are simply epiphenomenal, then traditional notions of responsibility based on mental states and on actions guided by mental states would be imperilled. But is the rich explanatory apparatus of intentionality simply a post hoc rationalisation that the brains of hapless homo sapiens construct to explain what their brains have already done? Will the criminal justice system as we know it wither away as an outmoded relic of a prescientific and cruel age? If so, criminal law is not the only area of law in peril. What will be the fate of contracts, for example, when a biological machine that was formerly called a person claims that it should not be bound because it did not make a contract? The contract is also simply the outcome of various 'neuronal circumstances'.

Given how little we know about the brain–mind and brain–action connections, to claim that we should radically change our conceptions of ourselves and our legal doctrines and practices based on neuroscience is a form of neuroarrogance. Although I predict that we will see far more numerous attempts to introduce neuroevidence in the future, I have elsewhere argued that, for conceptual and scientific reasons, there is no reason at present to believe that we are not agents.[29] It is possible that we are not agents, but the current science does not remotely demonstrate that this is true. The burden of persuasion is firmly on the proponents of the radical view.

What is more, the radical view entails no positive agenda. Suppose we are convinced by the mechanistic view that we are not intentional, rational agents after all.[30] What should we do now? We know that it is an illusion to think that our deliberations and intentions have any causal efficacy in the world. We also know, however, that we experience sensations – such as pleasure and pain – and care about what happens to us and to the world. We cannot just sit quietly and wait for our brains to activate, for determinism to happen. We must and will deliberate and act.

Even if we thought that the radical view was correct and standard notions of genuine moral responsibility and desert were therefore impossible, we might still believe that the law would not necessarily have to give up the concept of incentives. Indeed, Greene and Cohen concede that we would have to keep punishing people for practical purposes.[31] Such an account would be consistent with 'black box' accounts of economic incentives that simply depend on the relation between inputs and outputs without considering the mind as a mediator between the two. For those who believe that a thoroughly naturalised account of human behaviour entails complete consequentialism, this conclusion might be welcomed.

On the other hand, this view seems to entail the same internal contradiction just explored. What is the nature of the agent that is discovering the laws governing how incentives shape behaviour? Could understanding and providing incentives via social norms and legal rules simply be epiphenomenal interpretations of what the brain has already done? How do we decide which behaviours to reward or punish? What role does reason – a property of thoughts and agents, not a property of brains – play in this decision?

If the truth of pure mechanism is a premise in deciding what to do, no particular moral, legal or political conclusions follow from it.[32] The radical view provides no guide as to how one should live or how one should respond to the truth of reductive mechanism.

29 Morse, 'Lost in Translation?' (n 15) 543–54; Stephen J Morse, 'Determinism and the Death of Folk Psychology' (2008) 9 Minnesota Journal of Law, Science and Technology 1.

30 Of course, the notion of being 'convinced' would be an illusion too. Being convinced means that we are persuaded by evidence or argument, but a mechanism is not persuaded by anything. A mechanism is simply neurophysically transformed.

31 Greene and Cohen (n 5) 218.

32 This line of thought was first suggested by Professor Mitchell Berman in the context of a discussion of determinism and normativity. Mitchell Berman, 'Punishment and Justification' (2008) 118 Ethics 258, 271.

Normativity depends on reason and, thus, the radical view is normatively inert. If reasons do not matter, then we have no reason to adopt any particular morals, politics or legal rules, or to do anything at all.

Given what we know and have reason to do, the allegedly disappearing person remains fully visible and necessarily continues to act for good reasons, including the reasons currently to reject the radical view. We are not Pinocchios, and our brains are not Geppettos pulling the strings.

8 The case for cautious neurolaw optimism

Despite having claimed that we should be exceptionally cautious about the current contributions that neuroscience can make to criminal law policy, doctrine and adjudication, I am modestly optimistic about the near and intermediate term contributions neuroscience can potentially make to our ordinary, traditional, folk-psychological legal system. In other words, neuroscience may make a positive contribution even though there has been no paradigm shift in thinking about the nature of the person and the criteria for criminal responsibility. The legal regime to which neuroscience will contribute will continue to take people seriously as people – as autonomous agents who may fairly be blamed and punished based on their mental states and actions.

In general, my hope is that over time there will be feedback between the folk-psychological criteria and the neuroscientific data. Each might inform the other. Conceptual work on mental states might suggest new neuroscientific studies, for example, and the neuroscientific studies might help refine the folk-psychological categories. The ultimate goal would be a reflective, conceptual–empirical equilibrium.

More specifically, there are four types of situations in which neuroscience may be of assistance: (1) data indicating that the folk-psychological assumption underlying a legal rule is incorrect, (2) data suggesting the need for new or reformed legal doctrine, (3) evidence that helps adjudicate an individual case, and (4) data that help efficient adjudication or administration of criminal justice.

Many criminal law doctrines are based on folk-psychological assumptions about behaviour that may prove to be incorrect. If so, the doctrine should change. For example, it is commonly assumed that agents intend the natural and probable consequences of their actions. In many or most cases it seems that they do, but neuroscience may help in the future to demonstrate that this assumption is true far less frequently than we think. In that case, the rebuttable presumption used to help the prosecution prove intent should be softened or used with more caution.

Second, neuroscientific data may suggest the need for new or reformed legal doctrine. For example, control tests for legal insanity have been disfavoured for some decades because they are ill understood and hard to assess. It is at present impossible to distinguish 'cannot' from 'will not'. Perhaps neuroscientific information will help to demonstrate and to prove the existence of control difficulties that are independent of cognitive incapacities. If so, then perhaps independent control tests are justified and can be rationally assessed after all. More generally, perhaps a larger percentage of offenders than we currently believe have such grave control difficulties that they deserve a generic mitigation claim that is not available in criminal law today. Neuroscience might help us discover that fact. If that were true, justice would be served by adopting a generic mitigating doctrine. On the other hand, if it turns out that such difficulties are not so common, we could be more confident of the justice of current doctrine.

Third, neuroscience might provide data to help adjudicate individual cases. Consider the insanity defence again. As in *US v Hinckley*,[33] there is often dispute about whether a defendant claiming legal insanity suffered from a mental disorder, which disorder the defendant suffered from, and how severe the disorder was.[34] At present, these questions must be resolved entirely behaviourally, and there is often room for considerable disagreement about inferences drawn from the defendant's actions, including utterances. In the future, neuroscience might help resolve such questions if the clear-cut problem difficulty can be solved. As mentioned previously, however, in the foreseeable future, I doubt that neuroscience will be able to help identify the presence or absence of specific *mens reas*.

Finally, neuroscience might help us to implement current policy more efficiently. For example, the criminal justice system makes predictions about future dangerous behaviour for purposes of bail, sentencing, including capital sentencing, and parole. If we have already decided that it is justified to use dangerousness predictions to make such decisions, it is hard to imagine a rational argument for doing it less accurately if we are in fact able to do it more accurately. Behavioural prediction techniques already exist. The question is whether neuroscientific variables can add value by increasing the accuracy of such predictions considering the cost of gathering such data. Very recently, two studies have been published showing the potential usefulness of neural markers for enhancing the accuracy of predictions of antisocial conduct.[35] Although these must be considered preliminary, 'proof of concept' studies,[36] it is perfectly plausible that in the future, genuinely valid, cost–benefit justified neural markers will be identified, and thus, prediction decisions will be more accurate and just.

9 Conclusion

At present, neuroscience has little to contribute to more just and accurate criminal law decision-making concerning policy, doctrine and individual case adjudication. This was the conclusion reached when I tentatively identified Brain Overclaim Syndrome eight years ago, and it remains true today. In the future, however, as the philosophies of mind and action, and neuroscience mutually mature and inform one another, neuroscience will help us understand criminal behaviour. Although no radical transformation of criminal justice is likely to occur, neuroscience can inform criminal justice as long as it is relevant to law and translated into the law's folk-psychological framework and criteria.

33 525 F Supp 1342 (DDC 1981).

34 Ibid 1346.

35 Eyal Aharoni et al, 'Neuroprediction of future arrest' 110 (2013) PNAS 6223; Dustin A Pardini et al, 'Lower Amygdala Volume in Men is Associated with Childhood Aggression, Early Psychopathic Traits, and Future Violence' (2014) 75 Biological Psychiatry 73.

36 For example, a re-analysis of the Aharoni et al study (n 35) by Russell Poldrack, a noted 'neuromethodologist' demonstrated that the effect size was tiny <www.russpoldrack.org/search?q=aharoni>. Also, the study used good but not the best behavioural predictive methods for comparison.