

Neuroscience and Criminal Responsibility

David Hodgson

Neuroscience is telling us more and more about how our brains work, and about the effects that physical abnormalities and changes can have. As Professor Bennett's paper suggests, it is appropriate that this understanding be used in critically appraising and refining legal rules concerning criminal responsibility, including rules concerning the defence of insanity.

Professor Bennett concludes that it is 'up to the Cockburns and Tindals of the judiciary to determine how such understanding assists the "criminal law as it pertains to mentally incompetent defendants".' Perhaps it is the law reformers and the legislatures, rather than the judiciary, who will have the biggest say in this; but the important thing, I suggest, is that ultimately it has to be a combination of legal expertise and common sense (including what has been disparagingly called 'folk psychology') that determines how neuroscientific knowledge is to be used.

My reason for saying this is that the basic approach of the criminal law is to say that people should not be punished unless they *deserve* to be punished; and in general they don't deserve to be punished unless they have engaged in conduct in breach of the criminal law in circumstances where they are truly *responsible* for that conduct. I have argued in various places¹ that this approach should be maintained, along with the associated idea that criminal penalties are *just retribution* for criminal conduct. And while neuroscience can tell us much about what physical conditions of the brain have what effects on conduct, it cannot answer the question of whether someone is truly responsible for conduct or deserves to be punished for it.

I'm not sure whether Professor Bennett agrees with me on this: he seems to express disapproval of the view (expressed in the article in his last footnote) that in determining mental impairment 'the mere fact of organic damage alone is insufficient. It is the fact of disorder which is decisive, not the cause of it'. I say it is the existence and degree of *responsibility for conduct* that is decisive, and that physical conditions

¹ See for example my article 'Guilty mind or guilty brain: criminal responsibility in the age of neuroscience' (2000), *The Australian Law Journal* 74, 661-80

and their effects are important just to the extent that they bear on this question.

In this paper, I will outline why I contend that ideas of desert and retribution should be maintained, the difficulties of doing so in the light of philosophical and scientific considerations, and the role I see for neuroscience in the development and application of the criminal law.

1 Retributive and consequentialist purposes

The criminal justice system in Australia and other countries with similar legal systems serves two broad types of purposes, which are together considered as justifying the imposition of restraints or other detriments on offenders, and which inform decisions as to when to impose such restraints or detriments and what they should be:

- (a) Retributive, backward-looking purposes; and
- (b) Consequentialist, forward-looking purposes.

The former purposes are based on the idea, mentioned earlier, that a person who has acted criminally *deserves* to be punished for this conduct, and that it is *just* that appropriate detriment be inflicted on that person. And the idea that a person deserves punishment for criminal conduct presupposes that the person is truly *responsible* for it.

The latter purposes involve no such ideas. They simply look to the *good consequences* that imposition of detriment on offenders may be expected to have, notably:

- (i) Demonstration that certain types of conduct are unacceptable, and deterrence of the criminal and of others from engaging in that conduct;
- (ii) Restraint of the criminal from further crime during incarceration;
- (iii) Reform of the criminal;
- (iv) Placating victims and perhaps compensating them (although the latter may be considered a matter for civil rather than criminal law); and
- (v) Reassurance of the community that they are protected and that criminals will be punished (promoting confidence in security and in the rule of law, and discouraging self-help).

The interplay of these two sets of purposes is illustrated by:

- (a) The need for a *guilty mind* in criminal cases generally, and exceptions to this requirement; and
- (b) The principles applied in determining what *punishment* is appropriate.

As a general rule, criminal liability requires proof beyond reasonable doubt of a guilty mind, which in turn requires proof that the offender knowingly did something against the law in circumstances where the offender was responsible for this. Accordingly, if there is evidence raising questions about these matters, the prosecution must exclude such things as mistake of fact, self-defence and duress. There

may also be a question as to whether, by reason of some mental abnormality, responsibility is absent, giving rise to defences of insanity and sane automatism.

However, there are also a limited number of offences of strict or absolute liability, where the requirement of a guilty mind is reduced or absent, because consequentialist considerations are considered sufficient to justify placing an onus on citizens to make quite sure that some adverse event does not occur.

As regards punishment, the governing principle stated in *Veen (No. 2)*,² to the effect that in no case should the punishment exceed what is proportionate to the criminality involved in the offence, is an important application of the retributive purpose of punishment. It is considered unjust that more should be imposed on an offender than the offending conduct deserves.

Retributive considerations may operate in addition to *reduce* the penalty from the *Veen* maximum, for example if there is evidence of mental abnormality that contributed to the offending conduct. However, if this abnormality means that the offender is more *dangerous* than persons without it, then consequentialist considerations may mean that no reduction in penalty should be given. This means the court does not look for a perfect match between desert and punishment: it is not possible to achieve perfection; and in any event, so long as an offender had a real choice as to whether or not to commit an offence, it is reasonable that protection of society be considered as justifying a penalty limited only by what is proportionate to the criminality of the offending conduct.

2 Arguments against retributivism

There is a school of thought that retributive purposes should be abandoned or at least de-emphasised, which has recently been given some impetus by developments in neuroscience: see article by Greene and Cohen³ and material on philosopher Thomas Clark's website.⁴

It is argued that crime is an illness to be treated rather than wrongdoing deserving punishment, and that:

- (a) Retribution is
 - (i) inhumane and based on primitive impulses for vengeance; and

² *Veen v The Queen (No 2)* (1988) 143 CLR 465.

³ 'For the law, neuroscience changes nothing and everything' (2004), *Phil. Trans. R. Soc. London B* 359, 1775-85. This can be found on the internet by googling the title.

⁴ <http://www.naturalism.org/>

- (ii) not justified, because the real responsibility for criminal conduct lies in the genes and circumstances of the criminal, and free will and responsibility are illusions; and
- (b) Everything that is reasonable about the criminal justice system can be supported by its consequentialist purposes. In particular, although it could be argued that general deterrence and reassurance of the community are achieved so long as the person punished *appears to be guilty*, there are powerful consequentialist reasons for having a system that so far as possible ensures that only those who are truly guilty are punished, proportionately to their guilt, namely:
 - (i) to promote confidence in the justice system and the assurance that law-abiding citizens are not punished; and
 - (ii) to limit punishment to those who need and can benefit from deterrence and reform.

However, on a purely consequentialist approach, if a mistake is made and an innocent person is punished, this can be considered a bad thing only if the overall consequences are worse – *injustice* as such does not count. Thus, if the accidental and undetected punishing of an innocent person happens to have better consequences than the exoneration of that person, because the harm to the individual is outweighed by the general deterrent effect and placating of the community, the purposes of the criminal law would be served.

3 Challenges to free will and responsibility

The denial of free will and responsibility referred to above is not based on new ideas, but rather on old ideas that have been given fresh impetus by recent science.

In the 18th century Pierre Laplace argued that everything must happen as determined by Newton's (wholly deterministic) laws of motion, leaving no room for any contribution from free choice. The Scottish philosopher David Hume argued that what we do is determined by the preponderance of our desires: we always do what we most want to do, and what we most want to do is in turn determined by our characters and our circumstances.

In the 19th century, Darwin's theory of evolution offered an explanation of how random mutations and natural selection could lead to the emergence of organisms that give the appearance of making decisions and pursuing goals, while in fact being controlled by the operations of physical brains operating wholly in accordance with laws of nature.

At the end of the 19th century and beginning of the 20th century, the work of Sigmund Freud drew attention to the extent to which our behaviour is affected by unconscious processes, thus further challenging the idea that what we do is a matter of conscious choice. At around the

same time, the deterministic physics of Newton was displaced by the indeterminism of quantum mechanics; but this has not generally been regarded as supporting free will and responsibility, because it did no more than introduce the possibility of randomness into physical processes, and randomness may be considered the antithesis of efficacious decision-making and responsibility for conduct.

Arguments against free will and responsibility advanced by philosophers in the 20th century have included the argument that, for anything that happens, there must be sufficient prior causes, and there must in turn be sufficient prior causes for those causes, and so on; so that there is no room for any contribution from a human decision-maker that is not itself determined by prior causes outside the decision-maker's control. Thus the Australian philosopher J. J. C. Smart argued⁵ that every event is either causally determined or due to chance, and that there is no logical room between determinism and chance.

A basic dilemma about free will and responsibility has been well expressed by contemporary British philosopher Galen Strawson,⁶ in the following four propositions:

- (a) We do what we do because of the way we are, in terms of character and motivation.
- (b) So we cannot be responsible for what we do unless we are responsible for the way we are.
- (c) We cannot be responsible for the way we are when we first make decisions in life (that must be all down to genes and environment).
- (d) So we can never by earlier decisions become responsible for the way we are or for what we do.

These philosophical arguments have been reinforced in various ways by recent work in neuroscience. As more becomes known about how the brain works, in terms of processes that can be understood in terms of physics, chemistry and biology, the less room there may appear to be for free will. Neuroscience approaches the brain as a kind of machine, and seeks to explain its operations as being the operations of a machine.

One particular area of scientific research that is sometimes seen as excluding the possibility of free will is that undertaken by neuroscientist Benjamin Libet and his colleagues.⁷ These experiments suggest that

⁵ 'Free-will, praise and blame' (1961), *Mind* 70, 483-94.

⁶ See 'Luck swallows everything' (1998), *Times Literary Supplement*, 26 June, 8-10, and 'The bounds of freedom' in R. Kane (ed.) (2002), *Oxford Handbook of Free Will* (New York: Oxford University Press).

⁷ Libet, B., Gleason, C., Wright, W., and Pearl, D., 'Time of conscious

consciousness comes too late to initiate certain actions that the agent considers to be voluntary. In particular, where subjects were asked to make a movement at any time they decided to, neurological preparation for the movement was detected about half a second *before* the time identified by the subject as the time of the decision to move, suggesting that the subject's feeling that the decision was freely chosen must be an illusion.

Developments in neuroscience have led to suggestions from neuroscientist such as Colin Blakemore⁸ and Francis Crick⁹ that our feeling that we have free will is an illusion that arises because we are unaware of the physical brain processes that actually cause our actions; and psychologist Daniel Wegner has discussed experiments¹⁰ in which people claim to have made free choices in circumstances where they could not possibly have done so.

4 Virtues of retribution

As mentioned earlier, I contend there are good reasons for retaining retribution in the administration of criminal law, both as a general basis for determining when detriment is to be imposed on citizens and as an important factor in determining what detriment is to be imposed.

(a) If one has regard only to the consequences of what the State does to its citizens, they are being treated as *objects* to be manipulated for the general good. That is not appropriate in relation to people who are capable of acting rationally; and it does not encourage people to take responsibility for their conduct.

(b) Far from being inhumane, it is humane to refrain from imposing detriments on persons who have not been proved to have voluntarily breached a public law, and to impose on persons who have been proved to have breached a public law detriments that are no greater than are proportionate to the criminality of the offences committed. In Western societies, it is widely regarded as *legitimate* for governments to impose sanctions on people who have acted voluntarily in breach of public laws and, with limited exceptions, *not legitimate* for governments to impose sanctions or other loss of liberty on people who are innocent of doing

intention to act in relation to onset of cerebral activities (readiness potential): the unconscious initiation of a freely voluntary act' (1983), *Brain* 106, 623-42.

⁸ Blakemore, C. (1988), *The Mind Machine* (London: BBC) at 257, 269-271.

⁹ Crick, F. (1994), *The Astonishing Hypothesis* (London: Simon & Schuster) at 266.

¹⁰ Wegner, D. (2002), *The Illusion of Conscious Will* (Cambridge MA: MIT Press).

this: to impose sanctions or loss of liberty on people who are not at fault in this regard, without powerful justification, is considered a gross violation of human rights.

(c) Retribution that is proportionate to the criminality of the offence does not justify treatment that is harsh in a way that does not contribute to deterrence, and is consistent with pursuing rehabilitation . Where there is harsh treatment that does not contribute to deterrence and/ interferes with rehabilitation, this is due to defects in the system rather than correct application of retribution.

(d) This retributive approach has further advantages, in particular:

(i) if punishment is not generally limited to those who deserve it, no one could feel secure that compliance with the law would generally ensure they were not subjected to loss of liberty or confiscation of property by the State; and

(ii) if punishment is not made dependent on wrongdoing and proportional to the seriousness of the wrongdoing, the law would be less respected and resort to self-help and vengeance may not be kept in check.

(e) The need to prove voluntary conduct in breach of a public law before imposing detriment on persons is a necessary restraint on the conduct of the State and its officials. If the imposition of detriment is considered as justified whenever doing so has good consequences, whether or not the person concerned has done something to deserve it, there is no clear basis in principle for inhibiting the State and its officials from arbitrary arrest and detention.

It may be argued that these are consequentialist considerations that can be taken into account in a broader consequentialist theory, and that they do not in any event bear on the question whether or not there is any *truth* in the idea that people are really responsible for their conduct.

However, I contend that, unless one regards people as truly responsible for their conduct, there is no reason in principle for giving different treatment to a person who *has had the bad luck* to be regarded by a State official to be a danger to the State, from that given to a person who *has had the bad luck* to be caused by genes and environment to breach a public law. Both would be equally undeserving of what happens to them. The principle of human rights referred to earlier would thus be in jeopardy if ideas of responsibility are abandoned or weakened.

And while it is true that these considerations do not bear on the *truth* of free will and responsibility, they give us good reason to consider very critically arguments that may be advanced against retribution and against free will and responsibility.

5 Three ways to maintain retribution

There are three broad ways in which various writers have sought to maintain retribution despite challenges to free will and responsibility. (a) One approach is to say that ideas of responsibility and retribution should be maintained even though there is no such thing as free will. Some argue, on the basis of considerations of the type mentioned in the previous section, that we should maintain the illusion that there is free will even though there is not.¹¹ One prominent legal theorist who might possibly be included in this category, or possibly in the next, is Michael Moore,¹² who argues that the persistence and ubiquity of retributive impulses are signs of the moral reality that retribution is a *good thing* and a primary aim of the law, even though there is no free will. However, he supports the view that it is sufficient for culpability and responsibility that a person has the capacity to act rationally and to respond to reasons; and that a person is excused if sufficiently compromised in his or her rational capacity or coerced to act against his or her wishes. In this respect his views are similar to those in the next category.

(b) The second approach, adopted by philosophers such as Daniel Dennett,¹³ is to argue that, although the world is deterministic for all practical purposes, nevertheless free will and responsibility are *compatible* with this determinism. Human beings have free will and responsibility just because they are free to act in accordance with their own choices and to do whatever it is they most want to do; and it does not matter that their choices and their wants may themselves be determined by prior circumstances and impersonal laws. A prominent legal theorist Stephen J. Morse¹⁴ argues that responsibility is explained by our capacity to grasp and be guided by good reasons, and that this is so despite the truth of determinism.

(c) The third approach is to say that we have free will and responsibility in a sense that is incompatible with determinism. This view is called *libertarianism*, and it is a minority view in philosophy and science.

My take on these three approaches is as follows.

¹¹ For example, Smilansky, S. (2002), *Free Will and Illusion* (New York: Oxford University Press).

¹² Moore, Michael (1997), *Placing Blame: a General Theory of Criminal Law* (Oxford: Oxford University Press).

¹³ (2003), *Freedom Evolves* (London: Allen Lane).

¹⁴ Stephen J. Morse, 'Rationality and responsibility' (2000), *Southern California Law Review* 74, 251.

As for (a), I don't think that retribution could be justified if there is no free will or responsibility for conduct, and I don't think it's a good idea to act on the basis of a pretence.

As for (b), compatibilism is a defensible view (it is probably the majority view among philosophers). Certainly, even if the world is deterministic for all practical purposes, there still would be an important distinction between those persons who are rational in their behaviour, and those who by reason of mental abnormality are not rational or not fully rational; and this distinction supports retributive ideas, in that those who are not rational are less likely to be deterred by the threat of punishment, and are more appropriately dealt with by treatment, and if necessary confinement, than by punishment. However, I think compatibilism is vulnerable to Strawson's argument set out above: if we and all our actions are ultimately and completely the products of our genes and environment, how can anything we do be other than the inevitable result of things outside our control?

As for (c), this is the view I favour, and I think there are good arguments for it that are not widely appreciated.¹⁵ The (bare) gist of my position is that, given our circumstances, the way we are plus laws of nature provide available alternatives, inconclusive reasons (and how they appeal), and unconscious tendencies, and also the capacity to decide between the alternatives on the basis of the reasons; and what we do is what we decide in exercise of that capacity. Thus the constraining effect of the way we are is limited to determining alternatives, reasons and unconscious tendencies. Our decisions are not otherwise constrained by any distinguishing features of the way we are (we are all alike in respect of our capacity to decide), and to this extent we are truly responsible for them. I support this by arguing that conscious experiences can make a positive contribution to decision-making; that this contribution is not one wholly determined by rule-determined processes; that we can respond to whole feature-rich gestalt experiences that cannot engage as wholes with rules or laws of any kind; and that the role of consciousness (and its advantage) is to enable us to contribute this response to decision-making. One of my articles arguing for this view is attached to this paper. (Reading it is not compulsory!)

6 The role of neuroscience

¹⁵ See my articles 'A role for consciousness' (2008), *Philosophy Now* 65, 22-24; 'Why I (still) believe in free will and responsibility', edited version published under the title 'Partly free' in (2007), *The Times Literary Supplement* on 6 July; and other articles at <http://users.tpg.com.au/raeda>.

There is no doubt that neuroscience will continue to develop and tell us more and more about how brains operate. Some see this as meaning that belief in free will and responsibility will be progressively eroded, because it will become increasingly obvious that there is just no room for free will. For reasons set out in the articles referred to in footnote 15, I don't think this view is justified; and I think claims to this effect should be carefully and critically examined.

There is however every reason to expect that neuroscience will play an increasing role in the criminal law; and so long as its claims are critically appraised, this should be welcomed. Neuroscience may contribute in at least the following ways:

- (a) Evidence of neuroscientists will increasingly be used in determining questions about criminal responsibility, in accordance with the categories the law prescribes.
- (b) Neuroscience may be expected to influence the development of the law concerning the attribution of criminal responsibility, particularly in the case of those affected by mental abnormalities.
- (c) Neuroscience is likely to assist in identifying brain conditions that involve particular risks of criminal behaviour, and in devising methods to minimise these risks.
- (d) More generally, increasing knowledge of how the brain works may be expected to guide programs for addressing environmental factors contributing to criminal conduct and for rehabilitation of offenders, and to contribute to the development and implementation of educational strategies to discourage criminal conduct.

In relation to the first two areas I have mentioned, there is an underlying difficulty. The categories used by the criminal law are not those of neuroscience, but rather are pre-scientific commonsense categories, which have their source in folk psychology. Categories such as willed or voluntary action, belief, and intention (of alleged offenders), and consent (of alleged victims), are often central to determining criminal liability; and they are not matters that are susceptible to scientific proof or even description. Even when what is in issue is a question of mental abnormality, the categories used by the law are non-scientific categories such as disease of the mind, knowing what one is doing, knowing that what one is doing is wrong, and substantially diminished responsibility.

These categories used by the law pre-suppose an active conscious agent who is generally responsible for conduct; whereas neuroscience focuses on the physical causes and effects of the functioning or malfunctioning of the brain. And there is not at present a generally accepted theoretical framework that links and reconciles these two approaches.

Further, if I am right in my contention that it is desirable that notions of personal responsibility and retribution be maintained by the criminal law, then it is to be hoped that the criminal law will continue to use categories that give effect to these notions; that is, categories which, like those presently used, pre-suppose active conscious agents generally responsible for their actions.

I expect that the contribution of neuroscience in these first two areas will for some time be limited to giving more detailed and accurate accounts of relevant aspects of brain functioning, which can then be used in commonsense reasoning to arrive at conclusions that engage with existing legal categories, or else guide future developments of the law. In particular, I would hope and expect that the law will retain the presumption that persons of sufficient maturity are generally responsible for their conduct; and I would also expect that, if this presumption is to be rebutted by neuroscientific evidence in particular cases, then this evidence will need to identify some brain abnormality and the effects of that abnormality, on the basis of which a commonsense conclusion can be reached as to whether and if so to what extent responsibility is excluded.

However, the combination of neuroscience and commonsense may enable the law to provide a more systematic approach to questions of this kind.

The present category of insanity is presumably intended to capture those cases where responsibility for conduct is effectively absent due to a brain abnormality that is of indefinite duration and/or difficult to treat; so that, if that brain abnormality is such as to make the person dangerous, there will be the option available of indefinite detention in order to protect the public. Neuroscience may well assist in achieving a better definition of these cases than is provided by the *M'Naghten* rules.

The present category of sane automatism is presumably intended to capture those cases where responsibility for conduct is absent either without there being any brain abnormality, or else because of a brain abnormality that is temporary and/or easy to treat and/or not such as to make the person a danger; so that it is reasonable to excuse the person altogether without there being any need for extensive future restraint or even monitoring. Again, neuroscience may assist in better defining these cases.

Where the law recognises that murder should be reduced to manslaughter in cases where there is substantially diminished responsibility, neuroscience may contribute to a better definition of when such diminished responsibility is to be recognised. This approach may also have application to those cases where addiction contributes to criminal conduct.

As regards the third area mentioned above, it is highly likely that neuroscience will be increasingly be capable of identifying brain conditions that involve particular risks of criminal behaviour, and of devising methods to minimise these risks. The question is what should be done with this capability.

I would myself strongly oppose its being used to justify detention and/or compulsory treatment of persons who have not yet committed any criminal offence. That course might be advocated by those who wish to abandon retribution and be guided purely by the consequentialist purposes of criminal law; but I say it would be contrary to the principle of human rights referred to earlier, which proscribes adverse treatment by the State of persons who have not committed any criminal offence, unless there is powerful justification. Mere identification of risk factors would not provide that justification; although I accept justification for restraint could be provided where what is identified is a substantial mental illness that makes the person a clear danger to himself or herself or to others, which could then support detention of the person for no longer than is necessary for medical treatment of the illness so as to minimise the danger.

Neuroscientific identification of risk factors could properly be taken into account in determining what sentence to pass (up to the limit considered proportionate to the criminality of the offence) on a person who has committed a criminal offence, and in guiding the treatment of that person during the period of that sentence. And it could properly provide a basis for offering voluntary treatment to persons who have not committed an offence.

One other area where identification of risk factors is now being used in some Australian jurisdictions is to justify extended detention of persons convicted of serious sex offences, who have served their sentences for their crimes but, having not successfully completed sex offender programs during their sentences, are considered at high risk of re-offending. I would argue that this procedure should be kept within strict limits, being applied only to serious sex offenders, and where the risk of re-offending is clearly proved to be high. Even then, it could be argued to be a denial of human rights, amounting to punishment for an offence going beyond that considered proportionate to the criminality of that offence; although against this, it could be argued that extended detention is not unfair in the case of someone who has already committed a serious crime and has not taken the opportunity, provided by the sentence for that crime, to address the risk factors that contributed to its commission.

Turning to the fourth area, it is to be expected that neuroscience will make substantial contributions towards the addressing of

environmental factors that contribute to criminal conduct and towards educational strategies to discourage criminal conduct, as well as towards rehabilitation of offenders. However, I would argue that central to both education and rehabilitation are belief in and acceptance of personal responsibility for conduct, and recognition and acceptance of appropriate moral standards of right and wrong. If neuroscience is permitted to undermine belief in and acceptance of personal responsibility for conduct and/or belief in the importance of moral considerations, then this would be counterproductive.