Efficient causation and neuroscientific explanations of criminal acts

Nick J. Davis[1]

General introduction

In recent years, legal scholars have looked to the brain and behavioural sciences to help in delimiting criminal behaviour. But what do the lawyers want from the neuroscientists? And are the scientists in a position to deliver the goods? In this chapter I will explore what the neurosciences can offer legal analysis, and will discuss the practical and theoretical limits we will face in looking for the roots of criminal action. In particular I will look at the extent to which we can ascribe intent to a person's action, and whether neuroscience can offer anything legally relevant to this fundamental question.

How should we understand the intention behind a person's actions? It is a requirement of criminal law that a person should intend an act to occur, in order for the act to be malicious; this is the principle famously set down by Edward Coke: "*actus non facit reum nisi mens sit rea",* the act is not guilty unless the mind is also guilty). Where then do we look for the malicious intent? Recently legal scholars have looked to psychology and the neurosciences for guidance in interpreting the behaviour and motivation of a person who has committed an act. However we need to find the appropriate level at which to talk about psychological constructs that are of relevance to the law. Does it make sense to talk about the

[1] Senior Lecturer in Psychology at Manchester Metropolitan University, UK. I am grateful to Prof. Bebhinn Donnelly-Lazarov for helpful comments.

passage of individual action potentials along the axon of a neuron? Or should we talk more abstractly about higher-level goals? Dennis Patterson[2] has argued for the latter position, that we should talk in terms of *reasons* for action. I will examine this view, and present an argument based on what is currently known about the neuroscience of intention.

Mens rea and actus reus

The law has from its earliest days had a dual approach to action. A person must have committed an act, and must also have intended to commit that act. If a person is behind the wheel of a car that injures a pedestrian, she has surely committed an injurious act – an *actus reus*. But for the driver to be criminally liable she must have caused the act. For this latter condition the driver must have acted through intention, for example by choosing to drive after drinking alcohol or by choosing to drive with excessive speed. This choice, the *mens rea*, must be demonstrated if the driver is to be found liable for the injury to the pedestrian. It is possible to have each of these two components separately: the driver's car may have mounted the pavement after being struck by another car (*actus reus* but no *mens rea*), or the driver may have left the house that day with the intention of running over her ex-partner but decided against it at the last moment (*mens rea* but not followed through with *actus reus*). Similarly, the offensive act could be one of omission (failing to maintain the car's brakes) or one of possession (if one defines unlawful possession as an act). In any case there must be a deliberateness to the person's act for the person to hold liability. I will not deal further with acts of omission here, since the principles apply equally to deliberately acting as to deliberately non-acting. I will also not deal with offensive acts caused by recklessness, which also count as acts in *mens rea*.

[2] Dennis Patterson, 'Neuroscience and the explanation of human action', this volume.

For legal purposes, an 'act' has a fairly clear definition. A person's bodily movement, whether intended or involuntary, must lead to the harmful situation. This movement could be a gross motor act such as a kick, a punch or a tilt of a steering wheel, or could be a movement of the vocal apparatus or the typing fingers, in the case of offenses of language. In any case, an overt change of the person's bodily state must lead to the harm, even when the harm is separated in time or directness from the act[3]. This definition protects cognitive liberty, which is the freedom to think whatever one pleases – as long as no illegal act results from those thoughts[4]. Although there are some toothsome scholarly arguments around whether some acts truly count as acts[5], it is generally clear when a person's body has caused an event to occur.

However the body may not always be under the person's control, and involuntary movements such as tics or spasms, and acts committed while asleep or unconscious do not generally occasion legal liability, unless the person knowingly neglected to prevent those movements (such as if a person neglected to take medication that would suppress spasms). What is missing therefore is the sense of mental causation of the act, or the sense that the person intended the act to occur. This is the commonly agreed definition of mens rea: the intention or foreknowledge that a given act would be likely to cause a given harm.

---

[3] In *Thabo Meli and Others vs R* [1954] 1 WLR 228 (PC), the appellants were found to have killed a man, not through a blow to the head, which the victim survived, but from the attempt to dispose of (what they believed to be) the corpse by rolling it off a cliff. Rolling a dead body off a cliff does not constitute murder, but the chain of the day's events included acts intended to kill. So coincidence of cause and effect may be flexible.

[4] J-C Bublitz, 'My mind is mine?! Cognitive liberty as a legal concept'. In E. Hildt & A. Francke (Eds.), *Cognitive Enhancement* (Springer, 2013).

[5] In a famous example, the aforementioned Edward Coke unsuccessfully defended the preacher Edmund Peacham, who was found in possession of a sermon which had never been, and apparently would never be, delivered: *The Case of Edmund Peacham* (1615) 2 How. St. Tr. 869.

So where does that mental intention come from? Here I will examine the arguments that link mental causation to the physical properties of the person and her environment, and will review the ways in which the new neurosciences help us to understand how to ground the metaphysical theories in the functions of the brain and body.

The philosophical status of neuroscience

Most modern neuroscientists would consider themselves heirs to the tradition of *materialism*. That is, all phenomena in existence, including mental processes, arise solely from the properties of matter. A materialist would hold that there is no need to invoke a higher power or spirit to explain our inner life, and that a full understanding of mental process will eventually come from understanding the properties of the matter that constitutes the person. This materialism stands in contrast to the *dualism* of philosophers such as Rene Descartes, who could not reconcile material and mental phenomena. A practical problem in dualism is in explaining how non-material phenomena are able to influence the physical apparatus of the brain and thereby the body. Descartes chose the pineal gland as the site in the body where non-material mental processes interact with the physical body. But locating a location in the brain is not the same as identifying a process, and there still do not appear to be any good theories in the dualist tradition of how material and non-material phenomena interact.

Modern materialism is more properly called *physicalism*, with the different name acknowledging that there is more to the physical realm than the matter itself. The interactions among material phenomena also contribute to the functions of the system. So a computer processor can be understood in terms of the lattice of gates that is etched into it, but the true nature of the processor is only revealed when the gates are configured to run a program. A complete physicalist theory of mental processes may possibly draw on the electromagnetic properties of the cells that constitute the body, on the dynamics of thermal or metabolic transport around those cells, or on the relationship between the person and the surrounding environment. Although physicalism is somewhat difficult to define[6], there is at least the general sense that the physical laws of the universe contain within them the basis for all physical and mental phenomena.

A conundrum in modern neuroscience is the status of *determinism*. Determinism is the philosophical position that the current state of any system (the brain, the plumbing, or the whole universe) is completely determined by its state immediately prior. That is, if we could know everything about the state of the brain in one instant, including the position and momentum of every particle, we could in principle know how the state of the brain will unfold over time, and could therefore predict how the person will act at all times in the future[7]. Whilst there are of course practical and physical barriers to complete knowledge of a system like the brain, such as the famous uncertainty principle of quantum mechanics, true determinism excludes the idea of uncaused events. However these uncaused events are precisely what free will requires, since otherwise the future choices of the brain would be completely constrained by the developing state.

[6] A. Ney, 'Defining physicalism' (2008) 3/5 *Philosophy Compass* 1033-1048

[7] This is a gross oversimplification, of course. Rummens and Cuypers pick this apart, establishing that there are different forms of predictability, and suggesting that human agency allows for a certain degree of control over some forms of determinism. See S. Rummens and S. Cuypers, 'Determinism and the paradox of predictability' (2010) 72 *Erkenn* 233-249

Physicalism and physical incompleteness

The physicalist approach to intention treats the physical realm as a *closed system*, in that all phenomena exist within the system and are sustained by the system. So for this to be true we must find an explanation for mental phenomena that arises within the system, and does not require an external intervention. An argument against physicalist approaches is that there does not seem to be a good way to start doing this within our current knowledge of the physical universe. David Chalmers famously drew the distinction between the 'easy problems' of the philosophy of mind, which are usually problems that can be put as questions of physiology or information processing, and the 'hard problems' of the sense of perception and cognition, for which we do not have a good way to frame a question in physical terms[8].

Why is a neuroscientific explanation attractive?

Many legal scholars are now turning to the neurosciences to reduce the uncertainties in legal processes. Despite the problems that neuroscience faces in defining the proper substrate for analysis, nevertheless there is a body of knowledge in the brain sciences, and a set of methods for developing ideas, that may add significantly to the legal analysis of specific cases. Neuroscience has value in understanding a person's motivations, and in determining a proper outcome for an individual.

[8] D. Chalmers, 'Facing up to the problem of consciousness' (1995) 2 *Journal of Consciousness Studies* 200-219

Mitigation and treatment

Neuroscience can help us to understand when we must act compassionately towards a person who needs help more than they need punishment. The nature and the severity of many crimes may indicate that a person should be committed for treatment for their own good and for the good of others. For example, the M'Naghten Rules of the 1840s allow courts to test the state of mind of a defendant at the time of the act, based on the following principle:

> [T]o establish a defence on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong.[9]

The field of biological psychology offers some insights into the range of factors that may bias an otherwise healthy person towards certain types of behaviour. It is known that the personality dimensions of Conscientiousness and Neuroticism are moderately heritable[10], suggesting a degree of predetermination in our actions: a person who regularly experiences worry and anger may have inherited a predisposition for these emotions from their parents. Intelligence and socio-economic status share both genetic and environmental determination, although these factors may be mitigated through education.

---

[9] *M'Naghten's case* (1843) 10 Cl & F 200.

[10] R. Riemann, A. Angerleitner and J. Strelau, 'Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales' (1997) 65 *Journal of Personality* 449-475

None of the factors listed here have probative value, however a compassionate judge may incorporate evidence from medical and scientific experts to understand how predisposing circumstances may mitigate to wrongness of an act. For example:

> When a neurological test can demonstrate that a defendant's thought process does not operate as one a judge would normally encounter - whether that is due to diminished culpability or a brain injury - a judge could consider that information, counteracting any implicit biases she may be holding about the defendant's obvious characteristics.[11]

I will return below to the issue of bias, but the hope here is that some kind of measurement of a person's brain state or mind state might help in arriving at an appropriate punishment or treatment for a person convicted of an offence.

Intention

FIGURE 1

The factors described above are in a sense *post hoc*, in that a person must have committed an act before we can worry about them. However modern neuroscience allows us to delve deeper into the intentional quality of action. Figure 1 shows a schematic version of a famous experiment in neuroscience, the so-called Libet Clock. Libet asked participants to press a key on a computer at a time of their choosing, and to report what time was on the

[11] B. Donald and E. Bakies, 'A glimpse inside the brain's black box: Understanding the role of neuroscience in criminal sentencing' (2016) 85 *Fordham Law Review* 481-502 at 500.

clock when the participant "felt the will to move" (called the W-judgement)[12]. The participants consistently reported that the time of the W-judgement occurred before the time of the actual key-press, which would of course be necessary as the intention must precede the act. However Libet's experiment revealed something more interesting: the activity of the brain showed a clear shift ahead of the W-judgement, measurable up to half a second before the key-press. Since this shift in activity preceded the participants' experience of deciding to press the key, many scientists and philosophers have taken this as evidence that our sense of intention is no more than the conscious brain developing a post hoc explanation for an action that the non-conscious brain has already decided. Variants of Libet's original experiment have confirmed the basic premise, that brain activity changes before the 'decision' to act, and sophisticated neuroimaging can detect activity up to ten seconds before a free choice that accurately predicts that choice[13].

Seductive allure

It must be admitted that there remains a certain attraction in pinning messy legal realities onto a seemingly clean scientific field such as neuroscience. In fact, the so-called "seductive allure" of the neurosciences may bias people into believing evidence without properly questioning it, which would undermine the use of such evidence in court[14].

---

[12] B. Libet, C. A. Gleason, E. W. Wright and D. K. Pearl, 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act' (1983) 106 *Brain* 623-642

[13] C. S. Soon, M. Brass, H. J. Heinze and J. D. Haynes, 'Unconscious determinants of free decisions in the human brain' (2008) 11 *Nature Neuroscience* 543-545

[14] Classically this was reported by D. Weisberg, F. Keil, J. Goodstein, E. Rawson and J. Gray, 'The seductive allure of neuroscience explanations' (2008) 20 *Journal of Cognitive Neuroscience* 470-477, but see also D. P. McCabe and A. D. Castel, 'Seeing is believing: the effect of brain images on judgments of scientific

Although there is some doubt about the solidity of this finding[15], nevertheless there is something evocative about brain images that lends credibility to the facts of a case. This is not unique to the brain sciences of course: many miscarriages of justice have occurred because courts have been poorly equipped to properly examine scientific evidence presented by expert witnesses.

However the allure of the brain sciences, and the seeming solidity of the evidence and its presentation, induces a false confidence in people who handle such evidence. For example, the following passage was written by an American judge and a legal clerk (emphasis mine):

> Without disregarding the criminal justice system's ability to hold those accountable for their actions, neuroscience can be utilized to demonstrate that certain actions may actually be the result of developmental problems associated with the brain, like the effects of complex trauma on children. *A judge may also use neuroscience to combat her implicit biases, which have ways of manifesting themselves in the courtroom and therefore need to be explicitly acknowledged.* Neuroscience can offer additional insight into a defendant's thought process and accordingly provide a means for the judge to address and correct those biases.[16]

I would argue that reliance on neuroscientific evidence is itself a source of bias. No scientific fact is beyond questioning, and we should be as prepared to acknowledge and to

reasoning' (2008) 107 *Cognition* 343-352, and D. Weisberg, J. Taylor and E. Hopkins, 'Deconstructing the seductive allure of neuroscience explanations' (2015) 10 *Decision Making* 429-44.

[15] Some methodological and other concerns were raised by M. Farah and C. Hook, 'The seductive allure of "seductive allure"' (2013) 8 *Perspectives on Psychological Science* 88-90

[16] Donald and Bakies, supra note 11, at 482.

challenge neuroscience evidence as we are to challenge evidence from bite marks or from DNA[17].

Given the newness of the relationship between the law and the neurosciences, there is a responsibility to get this right from the outset. Counsel must be given the confidence to present, and to challenge, evidence derived from the brain sciences, and expert witnesses must ensure that evidence presented in court is digestible without being overly simplified.

Practical problems in finding neuroscientific explanations

The previous section sounds rather optimistic. The neurosciences can tell us so much about predisposition and intention, so why do we not hand over judicial decisions to the scientists? There are several practical problems in reading brain states and in relating brain states to mind states.

Complexity and redundancy in the brain

FIGURE 2

There are currently significant gaps in our knowledge of how action plans arise in the brain and are thereby enacted. The neuroanatomy of the motor system is well known. The primary motor cortex, lying in a strip at the back end of the frontal cortex, projects to the spinal cord, and from there to the muscles. Although there are some controversies about

[17] For concerns around the use of bite marks, see G. V. Reesu and N. L. Brown, 'Inconsistency in opinions of forensic odontologists when considering bite mark evidence (2016) 266 *Forensic Science International* 263-270. There is a readable account of problems with DNA evidence in P. A. Smith, 'When DNA implicates the innocent' (2016) 314 *Scientific American* 11-12.

exactly how information is processed in the primary motor cortex[18], there are helpful regularities such as a somatotopic mapping from brain areas to muscles, meaning that changes in activity measured in these areas have a predictable and predictive relationship with changes measured in the muscles. 'Upstream' of the primary motor cortex it becomes increasingly difficult to separate clear motor information from cognitive function. The higher motor areas of the premotor cortex (PMC) and the supplementary motor area (SMA) deal with higher-order plans related to goals and motivations, rather than to specific actions. The motor areas of the brain are shown in Figure 2. The experiments of Libet and Haynes and others, mentioned above, took place under rigidly controlled conditions and with detailed and accurate knowledge of the state of the brain before and after any decision was undertaken. In the real (legal) world we have no such conditions, and no such certainty around the outcomes that we see.

The complexity of the brain adds another problem. Like many complex systems, including the Internet, termite colonies and financial economies, the brain exploits redundancy and plasticity to afford robustness against external attacks. We see this in people who have suffered a stroke. Even when a section of the cortex is damaged by the injury, the function that is subserved by the affected area may be preserved as the brain reorganises and relearns. This issue goes to the heart of what Patterson[19] describes as efficient causation in the neurolegal sense. It is simply not possible to identify a single causal chain of events in the brain that leads to a given outcome.

---

[18] N. J. Davis, S. P. Tomlinson and H. M. Morgan, 'The role of beta-frequency oscillations in motor control' (2012) 32 *Journal of Neuroscience* 403-404

[19] Patterson, supra note 2.

Philosophical problems in finding neuroscientific explanations

Traditions in neuroscience

So far I have been careful around making the obvious link between the mind and the brain. There is no question that the brain is the centre of all mental processes. Damage to the brain is clearly linked to altered mental processing, and even Descartes looked in the brain for the seat of his dualist interactions. However it would not be accurate to suggest that the brain is the closed system that we are looking for. The most obvious point is that the brain does not exist on its own, but is connected via a variety of tissues to the sensory organs, to the muscles, and to the internal organs of the body. In a trivial sense, the brain is crucially situated in the body, and is immersed in the sensory milieu. Notwithstanding philosophical conundrums such as the brain-in-a-vat thought experiment of Gilbert Harman[20], the brain is part of the body and is indivisible from it.

A branch of behavioural psychology suggests that not only should we consider the brain in the context of the whole body, but further that the functions of the brain and body only make sense in relation to the flow of information passing around and through them. This field of *ecological psychology* observes that all of the information needed to behave freely in the world is available directly from the sensory arrays, and that little, if any, information needs to be stored in a memory representation[21]. Extending this further, the programme of *embodied cognition* seeks to understand mental function in relation to the environment and to

[20] This was a neat thought experiment devised by Harman in G. Harman, *Thought* (Princeton, NJ: Princeton University Press, 1973), and was developed by Hilary Putnam in H. Putnam, *Reason, truth and history* (Cambridge, UK: Cambridge University Press, 1981).

[21] J. J. Gibson, *The ecological approach to visual perception* (Boston, MA: Houghton Mifflin, 1979)

available behaviours[22]. For example, mentally rotating an image of a hand is faster than mentally rotating an image of blocks, which is consistent with our familiarity with our own hands[23].

The difference between many fields of cognitive science comes down to what quality of the universe is being minimised or maximised. Many researchers who study the motor system believe that movement errors should be minimised[24], while those who study the sensory systems believe that some measure of sensory fidelity should be maximised[25]. The ecological branch focuses on the amount of information processing that is needed, believing that it is wasteful for the brain to simulate or calculate information that is cheaply available in the environment[26].

The disunity of neuroscience

All this is to say that the proper substance of neuroscientific analysis has not been sufficiently established. Researchers in different fields, or in different traditions, look at the cognitive architecture of the brain and see different things. Many other researchers are less concerned with the information processing within the brain, as they are with the tools they have for extracting that information. People who study language functions almost invariably use electroencephalography (EEG), while those who study brain-wide cognitive of memory

---

[22] E. Thompson and F. Varela, 'Radical embodiment: neural dynamics and consciousness' (2001) 5 *Trends in Cognitive Sciences* 418-425

[23] S. Kosslyn, G. DiGirolamo, W. Thompson and N. Alpert, 'Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography' (1998) 35 *Psychophysiology* 151-161

[24] C. Harris and D. Wolpert, 'Signal-dependent noise determines motor planning' (1998) 394 *Nature* 780-784

[25] D. C. Knill and A. Pouget, 'The Bayesian brain: the role of uncertainty in neural coding and computation' (2004) 27 *Trends in Neurosciences* 712-719

[26] D. N. Lee and P. Reddish, 'Plummeting gannets: a paradigm of ecological optics' (1981) 293 *Nature* 293-294

processes may reach for functional magnetic resonance imaging (fMRI). The wholeness of our understanding of the brain and its activity is threatened by the disparity between the methods and the theories used to understand it.

Can there be unity in the fields of behaviour and neuroscience? I believe there can. The philosophical bedrock comes from the notion that we live in a *computable* universe where systems progress according to lawful rules[27]. Those rules come from the laws of physics, and particularly the laws that apply at the scale of our body size, and to the substances that compose our bodies. For example, Fick's laws describe the diffusion of chemicals in a fluid, Maxwell's equations relate different forms of electromagnetic energy, and Boltzmann (following Newton) developed a basis for understanding entropy. With these laws in place, and a sense of their dynamics, we have the basis for understanding brain processes in relation to each other, which is the beginnings of a theory. However there is some considerable work to be done to get to that point.

Possible solutions

The need for better neuroscience

Where does this epistemic obstacle course leave us? Many of the issues raised are practical ones, and the solution may lie in better technology and better theories. Pushing the envelope of brain imaging allows us to see intentions develop, several seconds before action[28]. Another solution may be to attack the problem from another angle. The same advances in neuroimaging that showed us the anatomy of intention may also be used to probe

---

[27] D. Deutsch, *The Fabric of Reality* (London: Penguin Books, 1997).

[28] Soon et al., supra note 13

the content of memory. The field of lie-detection is relatively well-established, but is held back by the legal and moral questions around extracting the true state of another person's mind. The right to cognitive integrity is thought by many to be just as strong a human right as the right to bodily integrity, the violation of which is treated with the utmost seriousness in most jurisdictions.

Abstraction and emergence

We may also learn from the fields of science that deal daily with the most complex systems. The science of the climate is just such a field. The large-scale weather systems that gyre across the oceans are the result of uncountably many uncountably tiny interactions among molecules of water and gas and the rays of the sun. Yet we can ask simple, direct questions of them, such as: "should I wear a coat on Sunday", or: "will there be snow at Christmas". The weather scientists add all available information into their computers, and deliver their verdict: "unlikely to be necessary", or: "80% probable". People who ask and answer questions at the level of clothing decisions do not need to know the physical properties of water, or the likelihood of every possible interaction between every molecule of water and every possible solar ray that passes through the atmosphere. Instead we are comfortable with the statistical nature of the physical processes, and with the probabilistic nature of the predictions that arise from them. Similarly, we do not need to know the exact pathway of every neuron in a person's brain, nor the distribution of neurotransmitters and peptides, to know whether our friend is sad to have ended her relationship with her girlfriend. Again we are comfortable with approximations and probabilities. In both of these cases we develop an appropriate level of abstraction that suits the form of the question we want to ask. This does not deny the 'truth' of the most detailed level of analysis, but serves to remind us that most human cognition operates at a level of abstraction that might best be termed "beyond reasonable doubt".

The above touches on what has become known as *emergence*, or the notion that the proper level of analysis is not the tiniest level of physical detail, but at a higher level, where matter and forces interact lawfully to produce larger-scale effects. As John Blodwell[29] puts it:

> There is an underlying, unspoken assumption about what has come to count for us as existential bedrock, namely that at bottom 'reality' is stuff in space, evolving and interacting in accordance with the laws of physics. […] Ontological status then favours objects rather than operations.

However Blodwell argues that mental phenomena are not properties of the constituents of the brain (neurons, glia, cerebrospinal fluid, etc.) but of the interactions among those parts. These phenomena emerge when the parts communicate with each other.

Crucial to the discussion of emergence is the sense that the emergent phenomena are stable in some way. We are happy to believe that the 'person' (collection of cells and fluids collected within a matrix of skin and clothes) who stands in court bears some relation to the 'person' (collection of cells, etc.) that existed at the time of a supposed crime. And further that the same collection of cells will emerge from a period of corrective detention, chastened but improved. This temporal coherence gives meaning to dynamic systems such as personhood or cognitive processing.

[29] J. Blodwell, 'Physicalism and emergence' (2011) 18 *Journal of Consciousness Studies* 6-31 at 7

Reasons

Patterson has argued that the above examples focus on an inappropriate level of detail. Instead, the philosophical constructs of *reasons* are the best level on which to consider issues of motivation. What are reasons (in this context)? Robert Audi[30] defines a range of different types of reasons, which cover the different senses in which people mean the term. However the classes of reasons which concern us here are *normative reasons* and *explanatory reasons*. Normative, or justifying, reasons are reasons that make it sensible or likely that a person will act in a certain way. So the reason why I check if my dinner guests may be vegetarian is normative: I would not want to offend someone by presenting her with a steak if she does not eat meat. Actions result from a weighing up of the positive and negative normative reasons for those actions. By contrast, explanatory reasons are the reasons that led the person to commit the act. These are the causes (the Aristotelean *efficient causes* as Patterson notes) of the action by that specific person on that specific day.

Audi later[31] develops and elaborates the notion of reasons, relating them to different conceptions of moral virtue, or understanding the chains of reasoning (or *schemata*) that derive from basic reasons. However a limitation of the use of reasons in considering human behaviour is that they are opaque to external observers. The person who possesses the reason may have rational and well-justified reasons for her behaviour, but we still need that person to tell us what they are. So normative reasons may be guessed at from the time or the culture in which a person lives, but the explanatory reasons for a specific act may not be obvious unless a person chooses to reveal them.

[30] R. Audi, 'Acting for reasons' (1986) 95 *The Philosophical Review*, 511-546

[31] R. Audi, *Reasons, Rights, and Values* (Cambridge: Cambridge University Press, 2015)

Reasons are not, therefore, a directly helpful way to judge *mens rea*. Could they have any other uses? Examination of a person's reasons for action may reveal that a person believed they had acted on behalf of God, or in the cause of preserving the racial 'purity' of the nation. Neither of these is a good reason (or rather, an exculpatory reason), but the court may take these beliefs (or delusions) into account when sentencing for a crime.

Conclusions

Dennis Patterson[32] argues that the appropriate level of analysis for criminal neuroscience is that of reasons. I agree with Patterson that current neuroscientific explanations of action are unsatisfying, but I would further suggest that we keep trying. Steering away from physicalist explanations does not solve the problem the legal system has wrestled with for centuries: whether a person possessed a guilty intention (*mens rea*) at the time when they committed a guilty act (*actus reus*).

Law and neuroscience are natural friends. Legal scholars will find it profitable to understand the basis of voluntary action, and its underpinnings in the brain. However the lawyers must also understand the limits of current understanding about the functions of the brain, both in terms of the technological limits of the machines we use to probe the brain, but also in terms of the questions we are able to ask of the brain. Conversely the scientists will benefit greatly from contact with a field in which their knowledge can be properly applied, since new relationships breed new questions, and constant questioning is what drives science to improve. For their part, scientists must remain open to these enquiries, and must provide

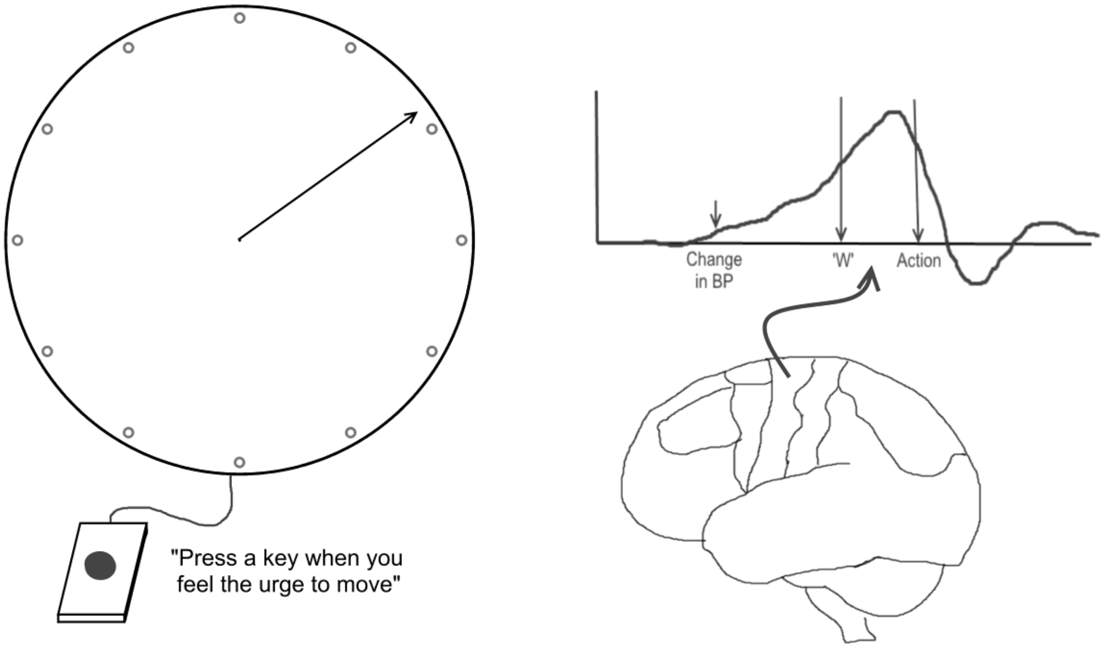
[32] Patterson, supra note 2

answers that engage with the question without skipping detail and without implying certainty or consensus. The so-called “seductive allure” of brain-based explanations may introduce bias to legal proceedings without adding to the scientific value, and to do so devalues the law and the neurosciences.

At the start of this chapter I asked what it was that lawyers want from the neurosciences, and whether the scientists were able to provide what is asked of them. Criminal law has one major question to ask of neuroscience: did this person intend to commit this act? Several additional questions may result from this one, such as: what was the person’s state of mind during the act? Or: has this person truly conquered her addiction, and so is less likely to offend again? However the question of pure intention is the big one. Neuroscience is not equipped to answer this question yet. The roots of intentional behaviour undoubtedly lie in the brain, but extracting that signal from the vast architecture of the brain is certainly beyond current technology, and may possibly be the wrong question to ask. Asking people to ‘intend’ in a lab, as in the paradigms of Benjamin Libet, or of Soon et al., reveal a network of brain areas that work together at different stages of the process. However none of these answer the more legally-relevant *post hoc* question: for the act that was committed by this person, did the person intend or not intend the outcome? Now the question becomes one of truth-detection, and again the neurosciences do not have a good, reliable means of separating truthful from deceitful responses. So the scientists cannot do what the lawyers seem to want of them: to answer the question of *mens rea*.

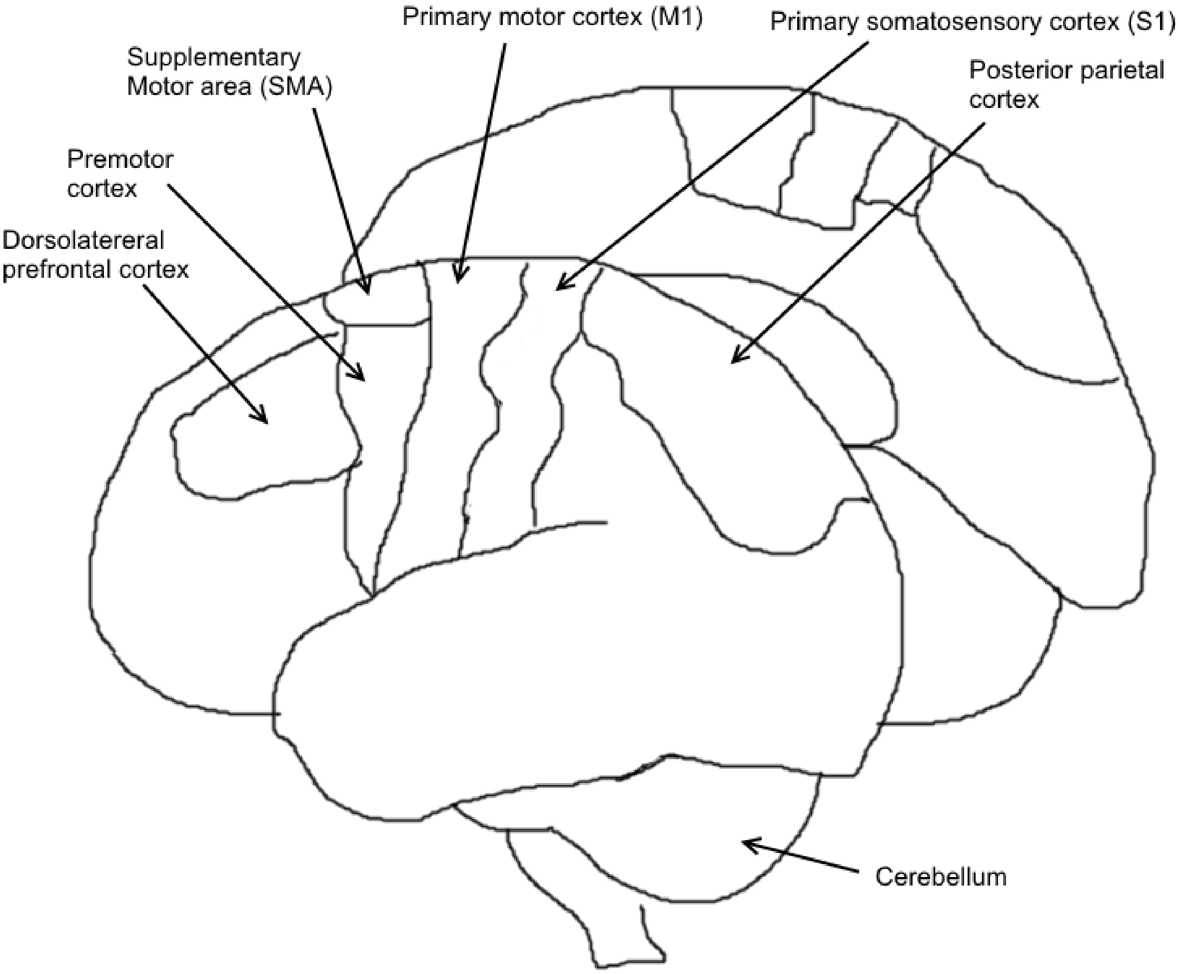
Neuroscience undoubtedly has much to offer the legal system. But in cases of criminal law, science can offer very little to aid the court. There seems little alternative but to continue to rely on the time-honoured system of judging each case on the best available evidence, however imperfect that system may seem.

FIGURE 1



Legend: In the classic experiments of Benjamin Libet (e.g. Libet et al., 1983, at note 12), a participant is asked to press a key whenever he or she "feels the urge (or the will) to move". The participant is asked to note the time shown on the clock face at the time of the 'will' – this is known as the W-judgement. The Bereitschaftspotential (BP), which is a measure of the electrical activity of the brain around the time of movement, shows a deflection leading up to the time of an action (the button-press), at which point the potential returns to baseline. The W-judgement precedes the movement, as we would expect. However the key finding of Libet's work is that the BP begins its deflection some hundreds of milliseconds prior to the time when the participant thought they had decided to move. This is taken by some as evidence that 'free will' is no more than post hoc rationalising by ne are of the brain for a decision already taken by another area.

FIGURE 2



LEGEND: The key motor-related areas of the brain. The primary motor cortex (M1) is the principal outflow of information to the muscles. However other areas are involved in planning actions (dorsolateral prefrontal areas and the posterior parietal cortex), and still others are involved in translating these plans into muscle dynamics (supplementary motor area, premotor cortex). The cerebellum and the primary somatosensory cortex (S1) monitor and guide ongoing action. The large amount of real estate devoted to action planning and coordination points to the complexity of information processing relating to action in the brain.