Introduction

Like other animals, we humans are distinctive. Physically, we have our own particular constitution. Behaviourally, we are special, acting, at least to an elevated degree, with purpose, sentience and self-consciousness. We are normative beings, having both an emotional and rational grasp of what we ought to do, framed by the existence of other conscious beings, by the environment we share, and by the institution of law. We experience great internal battles between our rational and appetitive powers and, depending on the outcome, may conform our behaviour, or not, to our normative commitments. We possess a range of interacting capacities – to move, to think, to comprehend, to want, to feel, to have reasons, and through these capacities, we are enabled to act. We are curious beings, pursuing knowledge not only for other ends but also for its own sake. And, although we are social and political, we are also intensely personal, existentially attuned to our condition in life. These depictions of human beings are depictions of a creature with a brain – with a human brain, of course. When it comes to understanding why we are the way we are perhaps there is nothing more to say: we are like *this* because our brains make us like this.

Each chapter in this multidisciplinary collection contributes, in some way, to understanding whether, how, and the extent to which this follows. More specifically, the chapters are concerned to examine the particularly prescient implications for legal responsibility of rapidly emerging neuroscientific understandings of the human brain. With unique authority, law regulates human action, a phenomenon about which neuroscience has much to say. By identifying neural correlates, it might tell us something that we did not previously understand about particular forms of culpable conduct; it can suggest interventions or neuro-enhancements to reduce an individual's propensity for criminal behaviour. It can illuminate the condition of the brain and speak of whether a defendant suffers a neurological 'deficit' or injury that might affect self-control, empathy, impulses, understanding or capacity for

foresight. Using fMRI and other techniques, it may help a court to understand whether a witness is telling lies or the truth, and by observing impulses in the brain, it might speak to whether a defendant intends rather than merely suspects, say, a prohibited outcome. Insofar as it casts relevant light on an accused's state of mind, or on a witness's testimony and depending on court rules, it might constitute evidence, and so the need arises to understand the complex evidence itself and its legal status/import.

Such is its significance that the interaction between neuroscience and law is now the subject of a dedicated field of study, 'neurolaw' (a term that this collection happily appropriates). By necessity the field integrates a number of perspectives. Here, the contributions come from cognitive neuroscience, law, psychology, criminology and philosophy. From the varying perspectives, three broad themes, with particular importance for law, emerge: (1) whether brains are the locus of responsibility, including legal responsibility, and what it might mean to say that they are; (2) what kind of information neuroscience really does provide (to law, in particular); and (3) the relevance in the courtroom of what neuroscience has to say. The themes correspond to the book's sections that can usefully be characterised as, respectively; conceptual, epistemic and legal.

The first section of this book is about the conceptual relationship between brains and human behaviour. According to the various proposed accounts of that relationship, the implications for responsibility are then explored. Naturally reductionism occupies a central place in this discussion: can human behaviour and states of mind be reduced to the brain, and if so, what kind of reduction might this be? Does human action even correlate to brain states? Are some or all of these reductions fatal (or not) to responsibility? On these and related questions, Dennis Patterson and Michael Moore seem to occupy different points on a continuum rather than different planes entirely. Patterson suggests that we can do better than refer to brain states in explicating human behaviour. In this regard, he rejects materialism, identifying laws, baseball, and poetry as three actions in the world that are not material. He proposes that for reductionism to offer anything close to an explanation of actions like these, it would need to justify avoiding their contextual and intentional elements, a task that seems, to him, impossible. Patterson's alternative approach, rooted in teleology, is to say that an understanding of action must appeal to reasons, hermeneutics and contexts rather than to efficient causal and deterministic explanations.

For Moore, reductionism, whatever form it might properly take, nonetheless has its limitations. He proposes that we cannot reduce who we are to the individual neurons in the brain, if it is to follow from such a reduction that the self disappears. But, he shows that many ordinary reductions are actually quite harmless in this respect: even if these reductions suggest that human beings are ultimately machines, the idea survives them that we are wonderful machines with a very special mechanism indeed. An interesting aspect of Moore's position is the idea that actually the most meaningful challenge to our mentalistic explanations might be the wholehearted eliminative materialism (EM) that seeks to replace mind with brain, rather than to so reduce it. If true, Moore agrees with Fodor that EM is, an 'intellectual catastrophe'. Fortunately, Moore finds no reason to assume any truth in it.

Sifferd's chapter, like Moore's, offers us a forensic analysis of reductionist enterprises. Her aim is to assess the compatibility of these various reductionist accounts with the mental causation that criminal law depends upon. Which explanation of the relationship between brain and mind best supports law's account of criminal responsibility and the folk concepts it entails? Sifferd defends the view that noneliminative reductionism fits the bill. She proposes that a noneliminative reductivist may readily embrace the causal power of mental states. Specifically, the account can admit the possibility that folk concepts pick out a disjunction of local physical states and do so reliably enough for us to consistently use such concepts to predict and understand human behaviour. In any particular case, a mental state has causal properties as a particular token instantiation of this local disjunction. In this way, Sifferd's account offers support for law's mens rea/actus reus paradigm, and like Moore, she affirms that 'we' remain responsible for what we do notwithstanding the fact that our doings have neural correlates. Indeed, for her, noneliminative reductionism supports rather than challenges or defeats this insight.

Donnelly-Lazarov considers the concept of intention, making the case that human intentions cannot be reduced, in any way, to brain states. Taking up a recent trend in the philosophy of action, her suggestion is that we know our actions non-observationally and that that this, alone, is what it means to intend. Moreover, by her account, there is no sensation of intending that accompanies a sensation of acting; the one is the other. For Donnelly-Lazarov, it follows that intention is not the kind of thing that human beings can be said to have at all, in the mind or the brain or

anywhere else. She suggests then that neuroscience will have to overcome a conceptual minefield if it is to account for intention in physical terms.

Nick Davis considers Patterson's view that reasons provide a better explanation for behaviour than physicalist explanations. He partly agrees but only for the time being. For Davis, law and neuroscience are natural friends. Legal scholars, he claims, will find it profitable to understand the basis of voluntary action and its underpinnings in the brain. However, he cautions that 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 and also in terms of the questions we are able to ask of the brain. Although, to this extent, he agrees with Patterson that current neuroscientific explanations of action are unsatisfying, he urges the neuroscientists to keep trying.

The second theme that the book explores is an epistemic one; it is about the nature and proper use of the information that our brains can provide. To begin, Mike Pardo examines the claims of lie-detection experiments. He notes that although the empirical problems associated with these experiments have received much attention, there are conceptual issues that warrant further scrutiny. Of particular concern to Pardo are the (false) presuppositions, embedded in these experiments, about what exactly is being measured. Pardo emphasizes the point that in order to confirm that brain activity correlates with any variable *X*, correct conceptual accounts of what exactly constitutes *X* are required. In particular, in the context of fMRI lie detection, scientists need to know what 'lying' is before they can identify its neural correlates. Pardo proceeds to analyse the (poorly understood) distinction between lying and deception and to scrutinise the concept of lying employed in the studies.

John Danaher also looks at the specific area of lie detection. He is less sceptical than some about its merits. In particular, Danaher suggests that EEG-based lie-detection tests, used appropriately, can assist the courts. Perhaps more important, these techniques, he claims, withstand philosophical opposition. Danaher includes, among this apparently ineffective opposition, the neurolaw mereological fallacy, identified by Patterson and Pardo (wherein states of mind or actions are wrongly attributed to the brain rather than to the whole person).

Forming the backdrop to both Danaher's and Pardo's accounts is some scepticism about the idea that neuroscience allows us to read the brain (or at least about whether it yet so allows). Does it? Searle has noted that computers do not really defeat humans at playing chess because the

machines are not really *playing* chess at all.[1] It might similarly be suggested that *reading* the brain is not really what neuroscience does. This is precisely Haselager and Mecacci's charge. The authors ask of fMRI techniques, for example, whether they 'understand' content meaning and suggest that contentwise such techniques are opaque. Haselager and Mecacci use the analogy of a cash point machine 'reading' a pin number. What is the pin machine doing with the code? They claim that a communication with the pin machine is not one of content; the machine receives an opaque instruction that allows it to identify the correct vehicle but not to *understand* the inputted content. This is the same limitation of brain reading as Haselager and Mecacci see it: The current stage of neuroscience does not enable it to read content. Rather, it allows it to categorise patterns of vehicle similarity. In essence, brain reading gives no systematic access to propositional attitudes. Why might this matter to law? The authors touch on a number of legal implications but focus on the Fifth Amendment. Their conclusion is not without irony that current brain-reading technology may fall outside the scope of the privilege that the Fifth Amendment affords defendants precisely because of its current shortcomings, and only as long as these last: There can be no self-incrimination, the idea goes, so long as the technology has no access to the content of our brains.

The final theme appears to be a very workaday, legal matter. It is about whether and how neuroscience can be of assistance to our systems of criminal justice. But, in this domain, difficult scientific, psychological and philosophical questions persist alongside the legal ones. Let's assume that there is some controversy, in neuroscience, about whether a victim's brain injury is typically caused by an assault or not. How can judges and juries possibly understand the complex neuroscientific evidence? What is the current legal status of neuroscientific evidence: is it admissible or not according to the usual standards of admissibility? How useful is the evidence in fact? Joanna Glynn illuminates the legal context in England and Wales, in detail, analysing the extent to which neuroscientific evidence might be ruled inadmissible by virtue of concerns over reliability. The interesting possibility is explored that, in this regard, neuro-evidence might properly be treated more cautiously than other forms of evidence, where usually concerns about reliability go to weight

[1] John Searle, 'What your computer can't know', *The New York Review of Books* (2014). Along similar lines see: Searle, 'The Chinese room' in R. A. Wilson and F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences* (Cambridge, MA: MIT Press, 1999).

rather than to admissibility. Glynn explores some of the reasons for this caution, noting, for example, the suggestion that adversarial systems encourage medical experts to express their views in terms of certainties, even where the grounding science is much less certain.

Criminal courts need evidence; they need it, inter alia, to assess responsibility. What is the relationship between an offender's brain and that person's responsibility for a criminal action? Are some defendants less responsible, or more, than others who perform the same crime, and if so does it matter whether the relevant mitigating or aggravating factors are connected to brain states and structures or to societal or familial influences, for instance? The question matters a great deal for those whose behaviour departs from expected norms. Raynor is in no doubt that neuroscience leaves our common-sense and ordinary-language use of responsibility and culpability largely intact. Raynor asks for greater conceptual clarity about the meaning of the term *responsibility*, particularly in the context of criminology, penology and rehabilitation theory, noting that such clarity will enable a better understanding of how psychological reports might be relevant to sentencing; of how cognitive behavioural therapy affects offenders' choices; and about how social and individual factors can lead to abstention from offending. He makes the case that some circumstances that commonly occur in criminal justice call for a realistic view of responsibility based on how far people could reasonably have been expected to exercise control over their behaviour.

Neuroscience might tell us that some defendants are behaviourally 'normal' but, on a particular occasion, fail to act in accordance with the norm. Others are behaviourally 'abnormal', with a reduced capacity to exercise control and experience fear. One such 'abnormality' often manifest in the context of criminal offending is psychopathy. These kinds of questions emerge: In what way is a psychopath mentally different from a nonpsychopath? Are these differences attributable to structural brain 'abnormalities'? What does it mean to lack the capacity to conform one's behaviour to social, legal or moral norms, and what does it mean for the brain to evidence such a lack of capacity? Is it morally more blameworthy or less to offend in virtue of a lack of capacity than to offend while having a capacity and failing to exercise it? Marion Godman addresses these and related questions. She examines the growing challenge from scientists and philosophers to view deficits in empathy and moral cognition as reasons to excuse rather than to blame psychopaths. In responding to the challenge she notes that offending by psychopaths often takes the form of instrumental, rather than the more commonplace reactive, aggression.

Godman explains that the nature of their aggression indicates both that psychopaths exercise choices about their behaviour and that early intervention, to prevent offending, is justified. Her contribution is original in that Godman considers the features of psychopathy against the notion of fairness. Without committing to a particular view, Godman considers the interesting possibility that whether psychopathic tendencies are inculpatory or not, there might be a duty, falling on society, somehow to compensate the psychopath for a condition, and its associated failings, that, after all, he has not chosen to have.

In the final chapter, Elizabeth Shaw examines the potential use in criminal justice systems of direct biomedical interventions designed to alter the brain states or behaviour of psychopaths. Shaw examines two objections to the interventions: (a) the idea that for reasons of principle it would be futile to attempt to alter psychopathic traits biomedically and (b) the claim that there is something about the nature of psychopathy, which means that it could *never* be ethically permissible to offer such offenders treatments, that carry significant risks. Shaw rejects both objections and proposes a list of preconditions to be satisfied before neurointerventions could permissibly be given.

Many of the chapters in this book began life at a neurolaw conference hosted by Swansea University in December 2014. The debates were lively, and we learned much from each other's disciplinary approaches. This is a forum for integrating the various perspectives, for exploring what neuroscience can offer to law and for considering how the institution should respond to the undoubted enlightenment neuroscience provides, the challenges it presents, and the limitations it has yet to overcome.