

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

1-1-2010

Lost in Translation?: An Essay on Law and Neuroscience

Stephen J. Morse

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the Behavioral Neurobiology Commons, Behavior and Behavior Mechanisms Commons, Criminal Law Commons, Criminal Procedure Commons, Criminology Commons, Diagnosis Commons, Law and Psychology Commons, Mental Disorders Commons, and the Public Law and Legal Theory Commons

Repository Citation

Morse, Stephen J., "Lost in Translation?: An Essay on Law and Neuroscience" (2010). *Faculty Scholarship at Penn Law*. 368.

https://scholarship.law.upenn.edu/faculty_scholarship/368

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Law and Neuroscience

Current Legal Issues 2010

VOLUME 13

Edited by
MICHAEL FREEMAN F.B.A.
*Professor of English Law
University College London*

OXFORD
UNIVERSITY PRESS

28

Lost in Translation? An Essay on Law and Neuroscience

*Stephen J. Morse**

28.1 Introduction

The rapid expansion in neuroscientific research fuelled by the advent of functional magnetic resonance imaging [fMRI] has been accompanied by popular and scholarly commentary suggesting that neuroscience may substantially alter, and perhaps will even revolutionize, both law and morality. This essay will attempt to put such claims in perspective and to consider how properly to think about the relation between law and neuroscience. The overarching thesis is that neuroscience may indeed make some contributions to legal doctrine, practice, and theory, but such contributions will be few and modest for the foreseeable future.

The first part of this essay describes the law's implicit folk psychological view of human behaviour and why any other model is not possible at present. It then turns to dangerous distractions that have bedevilled clear thinking about the relation between scientific explanations of human behaviour and law. Next, the essay considers how to translate the mechanistic findings of neuroscience into the folk psychological concepts the law employs. Finally, illustrative case studies of the legal relevance of neuroscience studies are presented. The discussion and all the examples focus on criminal law and on competence for the sake of simplicity and coherence, but the arguments are almost all generalizable to other legal contexts.

28.2 The Criminal Law's Implicit Psychology and Legal Criteria

Lawyers take the criminal law's implicit psychology for granted because there is seldom any need to identify or to question it. The new neuroscience may call the

* Ferdinand Wakeman Hubbell Professor of Law & Professor of Psychology and Law in Psychiatry, University of Pennsylvania Law School and School of Medicine. The author thanks Ed Greenlee for his invaluable assistance. His personal attorney, Jean Avnet Morse, provided sound, sober counsel and moral support.

law's psychology into question, however, so it is crucial consciously to recognize it and to understand what would be entailed if it were undermined.

Criminal law presupposes the 'folk psychological' view of the person and behaviour. This psychological theory causally explains behaviour in part by mental states such as desires, beliefs, intentions, willings and plans. Biological, other psychological, and sociological variables also play a role, but folk psychology considers mental states fundamental to a full explanation of human action. Human behaviour cannot be adequately understood if mental state causation is completely excluded or eliminated. Lawyers, philosophers, and scientists do of course 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 disputants use to convince others presuppose the folk psychological view of the person. Brains do not convince each other; people do.

For example, the folk psychological explanation for why you are reading this paper is, roughly, that you desire to understand the relation of neuroscience to law to improve your work, you believe that reading the paper will help fulfil that desire, and thus you formed the intention to read it. This is a 'practical' explanation rather than a deductive syllogism.

Folk psychology does not presuppose the truth of free will, it is perfectly consistent with the truth of determinism, it does not hold that we have minds that are independent of our bodies (although it, and ordinary speech, sound that way), and it presupposes no particular moral or political view. It does not claim that all mental states are necessarily conscious or that people go through a conscious decision-making process each time that they act. It allows for 'thoughtless', automatic, and habitual actions and for non-conscious intentions. It does presuppose that human action will at least be rationalizable by mental state explanations or that it will be responsive to reasons, including incentives, under the right conditions. The definition of folk psychology being used does not depend on any particular bit of folk wisdom about how people are motivated, feel, or act. Any of these bits, such as that excited utterances are reliable enough to be an exception to the hearsay rule, may be wrong, as neuroscience might help disclose. Moreover, neuroscience might help the law define and identify legally relevant mental states with more precision. The definition insists only that human action is in part causally explained by mental states.

To understand the importance of mental states, consider the criteria for criminal responsibility: the elements of the *prima facie* case—primarily acts and mental states—and the absence of an affirmative defence. All are infused with mental states. All crimes include a 'voluntary' act requirement, which is defined, roughly, as an *intentional* bodily movement (or omission in cases in which the person has a duty to act) done while the agent is in a reasonably integrated state of consciousness. Although the meaning of an intentional bodily movement is seldom specified, the best definition is a bodily movement that can be in principle understood according to the person's mental states. One can almost always ask of any act, 'Why did you do that?', and expect some explicit or implicit mental explanation. If there is none even implicitly possible, it is probable that the agent's bodily movement was not an act at all.

Other than crimes of strict liability, all crimes also require a culpable further mental state, such as purpose, knowledge, or recklessness. Some crimes are also defined with the *mens rea* of negligence, which appears to be the absence of a mental state. This is a controversial issue among legal scholars, but the best explanation is that the failure to pay attention when the agent was creating the substantial and unjustifiable level of risk that supports criminal liability is itself a type of culpable omission. On the other hand, some scholars believe that negligence is indistinguishable from strict liability or, even if it is distinguishable, it is not really a mental state at all.

All affirmative defences of justification and excuse involve an inquiry into the person's mental state, such as belief that self-defensive force was necessary or the lack of knowledge of right from wrong. Of course the person's mental state is influenced by biological, psychological, and sociological variables, and knowledge of these variables may help determine what the person's mental state was, but the law is ultimately concerned with the mental state itself, rather than the causes of it.

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 such as desires, beliefs, willings, and plans. We are the sort of creatures that can act for and respond to reasons, including legal rules and standards that are expressed and understood linguistically. The law treats persons generally as intentional creatures and not as mechanical forces of nature.

Law is primarily action-guiding¹ and could not guide people *ex ante* and *ex post* unless people could use rules as premises in their reasoning about how they should behave. Otherwise, law as an action-guiding normative systems of rules and standards would be useless, and perhaps incoherent.² Law can directly and indirectly affect the world we inhabit only by its influence on human beings who can potentially use legal rules to guide conduct. Unless people were capable of understanding and then using legal rules to guide their conduct, law would be powerless to affect human behaviour.³ As John Searle wrote:

Once we have the possibility of explaining particular forms of human behavior as following rules, we have a very rich explanatory apparatus that differs dramatically from the explanatory apparatus of the natural sciences. When we say we are following rules, we are accepting

¹ See, George Sher, *In Praise of Blame* (New York: Oxford University Press, 2006) 123 (stating that although philosophers disagree about the requirements and justifications of what morality requires, there is widespread agreement that 'the primary task of morality is to guide action'); John R. Searle, 'End of the Revolution' (2002) 43 *N.Y. Rev. of Books* at 33, 35.

² Scott J. Shapiro, 'Law, Morality, and the Guidance of Conduct' (2000) 6 *Legal Theory* 127, 131–2.

³ *Ibid.*, at 131–2. This view assumes that law is sufficiently knowable to guide conduct, but a contrary assumption is largely incoherent. As Shapiro (*ibid.*) writes: 'Legal skepticism is an absurd doctrine. It is absurd because the law cannot be the sort of thing that is unknowable. If a system of norms were unknowable, then that system would not be a legal system. One important reason why the law must be knowable is that its function is to guide conduct.' I do not assume that legal rules are always clear and thus capable of precise action guidance. If most rules in a legal system were not sufficiently clear most of the time, however, the system could not function. Further, the principle of legality dictates that criminal law rules should be especially clear.

the notion of mental causation and the attendant notions of rationality and existence of norms...

... The content of the rule does not just describe what is happening, but plays a part in *making it happen*.⁴

Legal rules are not simply mechanistic causes that produce 'reflex' compliance, although they can certainly help to inculcate law-abiding 'habits'. They operate within the domain of folk psychology. Legal rules are thus action-guiding primarily because they provide an agent with good moral or prudential reasons for forbearance or action. For example, no instinct governs how fast a person drives on the open highway. Among the various explanatory variables, however, the posted speed limit and the belief in the probability of suffering the consequences for exceeding it surely play a large role in the driver's choice of speed. 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. The inculcation of law-abiding habits, for example, has an intentional component and we constantly act in the 'shadow of the law', especially when criminal conduct is at stake.

The legal view of the person does not hold that people must always reason or consistently behave rationally according to some preordained, 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 all actions for which agents deserve to be praised, blamed, rewarded, or punished are the product of mental causation⁵ and, in principle, responsive to reason. 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 violate expectations of what they owe each other and only people can do wrong.

Many scientists and some philosophers of mind and action consider folk psychology to be a primitive or pre-scientific view of human behaviour and the next section of this paper considers such views. For the foreseeable future, however, the law will be based on the folk psychological model of the person and behaviour described and this paper will proceed on that premise. 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 lawyers

⁴ Searle, note 1, above, at 35.

⁵ I do not mean to imply dualism here. I am simply accepting the folk-psychological view that mental states—which are fully produced by and realizable in the brain—play a genuinely causal role in explaining human behaviour.

and legal policy makers not lose sight of this model lest they fall into confusion when various claims based on neuroscience are made. Once again, any neuroscientific data or evidence must always be relevant to the law's folk psychological criteria. If neuroscience is to have any influence on current law and legal decision-making, it must be through this framework.

This author's view is that a primary task for neuroscience is to explain agency, not to explain it away reductively. That is, neuroscience should attempt to explain how a two and a half pound hunk of organic matter can produce consciousness, intentionality, reasons responsiveness, mental causation, and all the other mental apparatus that is so central to our lives.

28.3 Potential Distractions and Confusions

This section of the paper considers a number of underlying and related issues that are often thought to be relevant to criminal responsibility and competence but that are irrelevant or confusions and distractions: free will, causation as an excuse, causation as compulsion, prediction as an excuse, dualism, and the non-efficacy of mental states (alluded to just above).

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. 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, but it never addresses the presence or absence of free will.

When most people use the term free will or its lack in the context of legal responsibility, they are typically using this term 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, such as that the defendant was legally insane or acted under duress, but such usage of free will only perpetuates misunderstanding and confusion. Once the legal criteria for excuse have been met, for example—and none 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 about 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 think the problem is not resolvable. This is a real 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, at present, having or lacking libertarian freedom is not a criterion of any civil or criminal law doctrine.

Neuroscience is simply the most recent mechanistic causal science that appears deterministically to explain behaviour. It 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 it is better, more persuasive science than some of its predecessors. 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 insanity, for example, 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 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 paper 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. 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 history marked by poverty and abuse played a predisposing causal role in a defendant's criminal behaviour. Or suppose that an alleged new mental syndrome played a causal role in explaining criminal conduct. The claim is often made that such causes, which are not within the actor's capacity to control rationally, 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 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 an excusing condition. For example, imagine a person with paranoid suspiciousness who constantly and hypervigilantly scans his environment for cues of an impending threat. Suppose our person with paranoia now spots a genuine threat that no normal person would have recognized

and responds with proportionate defensive force. The paranoia played a causal role in explaining the behaviour, but no excusing condition obtained. If the paranoia produced a delusional belief that an attack was imminent, then a genuine excuse, legal insanity—an irrationality-based defence—might be appropriate.

In short, a neuroscientific causal explanation for criminal conduct, like any other type of causal explanation, does not *per se* mitigate or excuse. It provides only evidence that might help the law resolve whether a genuine excuse existed or data that might be a guide to prophylactic or rehabilitative measures.

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. If compulsion is literal, say, a person's arm moves because the person had a neuromuscular spasm or because a much stronger person pushed the arm, the person has not acted at all. Metaphorical compulsion is more difficult to understand, but it includes cases in which someone acts in response to a do-it-or-else threat (such as the excuse of duress) or acts in response to strong internal urges or desires (such as the control test for legal insanity). In all metaphorical compulsion cases the person acts, however, and deciding when to mitigate or excuse in such cases is a normative legal question.⁶

It is crucial to recognize 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 compelling. Suppose, for example, that a person with paedophilic urges has them weakly and is weakly sexed in general. If the person molested a child there would be no ground for a compulsion excuse. If causation were *per se* 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. But causation 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 *per se* an excusing or mitigating condition, even if the predictability of the behaviour is perfect. To understand this, just consider how many things each of us does that are perfectly predictable for which there is no plausible excusing or mitigating condition. Even if the explanatory variables that enhance prediction are abnormal, excuse or mitigation is warranted only if a genuine excusing or mitigating condition is present. For example, recent research demonstrates that a history of childhood abuse coupled with a specific genetically produced neurotransmitter abnormality vastly increases the risk that a person will behave antisocially as an adolescent or young adult.⁶ Again, such information may be of prophylactic or rehabilitative use for people affected, but no excuse or mitigation is applicable just because these

⁶ Avshalom Caspi et al., 'Role of Genotype in the Cycle of Violence in Maltreated Children' (2002). 297 *Science* 851.

variables make antisocial behaviour far more predictable. 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.

Most informed people are not 'dualists' about 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, therefore, as if law's emphasis on the importance of mental states as causing behaviour is based on a pre-scientific, 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. 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 play no independent and partial causal role.

Despite the lack of understanding of the mind-brain-behaviour relation, some scientists and philosophers question whether mental states have any causal effect, treating mental states as psychic appendixes that evolution has created but that have no genuine function. These claims are not strawpersons. They are seriously made by serious, thoughtful people. If accepted, they would create a complete and revolutionary paradigm shift in the law of criminal responsibility and competence (and more widely). They are treated in a later section.

In conclusion, legal actors must always keep the folk psychological view present to 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 truth of determinism, causation and predictability do not in themselves answer any doctrinal or policy issue.

28.4 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 of 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 the doctrines, practices or institutions are incoherent, illegitimate, or unjustified.

This section will consider the internal potential relevance of neuroscience. It begins with general considerations concerning the relevance of neuroscience to law, using criminal law and competence determinations as its primary examples. Then it turns to a taxonomy of types of internal contributions. Finally, it considers more radical, external challenges to law from neuroscience.

28.4.1 General considerations concerning translation

The law's criteria for responsibility and competence are essentially behavioural—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,⁷ 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 in approach to folk psychology. Neuroscience, in contrast, is purely mechanistic and eschews folk psychological concepts and discourse. Thus, the gap will be harder to bridge.

The brain enables the mind. If your brain is dead, you are dead, you have no mind, and you do not behave at all. Therefore, facts we learn about brains in general or about a specific brain in principle could provide useful information about mental states and human capacities in general and in specific cases. Some believe that this conclusion is a category error.⁸ 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 to writing this paper 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. Then, 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. At present, 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.

Some preliminary points of general applicability must be addressed first, however. The most important is simply to repeat the message of the prior section of this paper. 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. Further, neuroscience is insufficiently developed to detect specific, legally relevant mental content or to provide a sufficiently accurate diagnostic marker for even severe mental disorder. Nonetheless, certain aspects of neural

⁷ Alan A. Stone, *Law, Psychiatry, and Morality* (New York: Cambridge University Press, 1984) 95–6.

⁸ M. R. Bennett and P. M. S. Hacker, *Philosophical Foundations of Neuroscience* (Malden, MA: Blackwell Publishing, 2003); Michael Pardo and Dennis Patterson, 'Philosophical Foundations of Neuroscience' (2010) *U.Ill.L. Rev.* 1211.

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 now lack such science, but future research may remedy this.

Questions concerning competence or predictions of future behaviour address a subject's present condition. Thus, the retrospective problems besetting retrospective responsibility analysis do not apply to such questions. 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 being offered, at a level the law considers normatively acceptable to warrant respecting the subject's choice and autonomy.

At present, most neuroscience studies on human beings involve very small numbers of subjects. Further, most studies average the neurodata over the subjects and the average finding may not accurately describe the brain structure or function of an actual subject in the study. 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. Over time, however, these problems may ease as imaging and other techniques become less expensive and more accurate, as research designs become more sophisticated, and as the sophistication of the science increases generally. It is also an open question whether accurate inferences or predictions about individuals are possible using group data for a group that includes the individual. This is a very controversial topic, but even if it is difficult or impossible now, it may become easier in the future.

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. There is usually some bit of behaviour, such as addiction, schizophrenia, or impulsivity, that they would like to understand better by investigating its neural correlates. To do this properly presupposes that they have identified and validated the behaviour under neuroscientific investigation. On occasion, the neuroscience might suggest that the behaviour is not well-characterized or is neurally indistinguishable from other, seemingly different behaviour. In general, however, the existence of legally relevant behaviour will already be apparent. 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.

Whatever neuroevidence is adduced must be translated into the folk psychological criteria the law employs. That is, the expert must be able to explain precisely how, for example, the neuroevidence bears on whether the agent acted,

formed a required *mens rea*, or met the criteria for an excusing condition. 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. In addition, 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.

At present, we lack the ability neurally to identify the content of a person's legally relevant mental states, such as whether the defendant acted intentionally or knowingly, but we are increasingly learning about the relation between brain structure and function and behavioural capacities, such as executive functioning, that are apparently relevant to broader judgements about responsibility and competence. We are unlikely to make substantial progress with neural assessment of mental content, but we are likely to learn more about capacities that will bear on excuse or mitigation.

Finally, and most importantly, because the responsibility and competence criteria are behavioural, actions speak louder than images. This is a truism for all criminal responsibility 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. And 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 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 clearly 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, such evidence will be of limited help, however.

If actions speak louder than images, however, what room is there for using neuroevidence? 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 recognize this unfortunate condition. No further tests of any sort will be necessary to prove this. In such cases, neuroevidence will be at most convergent and increase our confidence in what we already had confidently concluded. Whether it is worth collecting the neuroevidence will depend on how cost-benefit justified obtaining convergent evidence will be.

The most striking example of just such a case was the US Supreme Court's decision, *Roper v Simmons*,⁹ which categorically excluded the death penalty for capital murderers who killed when they were sixteen or seventeen years old because such killers do not deserve the death penalty. The *amicus* briefs were replete with neuroscience data showing that the brains of late adolescents are not fully biologically mature, and advocates used such data to suggest that the adolescent killers could therefore not fairly be put to death. Now, we already knew from common-sense observation and 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.

Whether adolescents are sufficiently less rational on average than adults to exclude them categorically from the death penalty is of course 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, thus confusing the positive and 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 are 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 the neuroscience at present is not likely to be of much help. I term this the 'clear cut' problem. Recall that neuroscientific studies usually start with clear cases of well-characterized 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 between less clear behaviour and controls is greater. Thus the neural markers of clear cases will provide little guidance to resolve behaviourally ambiguous cases of legally relevant behaviour. For example, suppose 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. Now, small but statistically significant differences in neural structure or function between people who are clearly suffering from schizophrenia and those who are not have been discovered. 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.

With these general considerations in mind, let us now turn to the specific types of contributions neuroscience might make to the law as it now exists.

⁹ *Roper v Simmons* 543 US 551 (2005).

28.4.2 Potential internal contributions of neuroscience

28.4.2.1 *Help demonstrate that the particular bit of folk wisdom or apparent truth about the world underlying a particular doctrine or legal practice is wrong and thus suggest that the doctrine or practice should change or be abandoned*

The proponent must show how the evidence confirms or challenges the folk psychological suppositions that underlie the doctrine in question. Showing that brain activation, even abnormal activation, is present or absent and played a causal role is virtually never per se legally relevant. The brain is always actively playing a role. The proponent must show specifically how the neurodata confirm or challenge the underlying folk psychological assumptions.

For example, excited utterances are an exception to the hearsay doctrine because the law assumes that they are likely to be true. Neuroscience might demonstrate that when people are excited in the way the law requires for this exception, the parts of the brain activated or deactivated are consistent with decreased accuracy of a subject's recollections. (One would have to induce experimentally excitement in the subject and keep him or her still enough to be validly scanned during the experimental task.) But the gold standard demonstration would be behavioural because if it turns out that excited utterances are in fact more likely to be accurate, the exception makes sense even though accompanying brain states seem inconsistent. In any case, jettisoning such a doctrine would not cast doubt on the folk psychological model. It presupposes it.

For another example, Terry Maroney has demonstrated that Supreme Court justices use incorrect emotional common sense as a basis for deciding certain cases.¹⁰ This is clearly poor judicial reasoning and incorrect behavioural premises should be corrected by the science. Note that the science must begin with correctly identifying the statistically normative emotional response and neuroscience will do little more than confirm this.

28.4.2.2 *Suggest the need for new doctrine or practices*

One might claim that neuroscientific findings demonstrate that the law needs new doctrines and that injustice is created by their absence. For example, only a minority of states now have a control test for legal insanity. I believe that virtually all cases in which justice might demand a control test for criminal responsibility can be explained according to a theory of impaired rationality,¹¹ but one can certainly imagine neuroscience data that might be consistent with a control problem entirely independently of any rationality defect. How it would be consistent would have to be demonstrated. The psychological intermediary variables would have to be

¹⁰ See Terry A. Maroney, 'Adolescent Brain Science and Juvenile Justice', above, 255.

¹¹ Stephen J. Morse, 'Uncontrollable Urges and Irrational People' (2002) 88 *Va. L. Rev.* 1025.

identified. If all this were done, it might be strong evidence to adopt a control test if one believes that people with such problems deserve some type of excuse or that they cannot be deterred. Note that the capacity to control oneself is a folk psychological concept. Machines may have internal regulatory mechanisms built into their design, but only people control themselves by suppressing their desires and intentions to act in prohibited ways. Folk psychology is again presupposed.

Permit me a bit of special pleading for my next example. Anglo-American criminal law does not have a generic mitigating doctrine that is applicable at trial. Unless a defendant meets the criteria for a complete excuse such as legal insanity or duress, the law expects the defendant to control himself despite temptations and provocations. Mitigation is then left to the discretion of the sentencing judge. I have argued, in contrast, that criminal law should provide defendants with a partial mitigating doctrine applicable at trial that would in appropriate cases reflect a defendant's substantially diminished rational or control capacities.¹² Neuroscience studies might help confirm that some defendants have much greater difficulty obeying the law in cases of provocation and temptation than criminal law and common sense presuppose. If this were true, in the interest of fairness, of accurately responding to a defendant's culpability, perhaps the criminal law would be justified in adopting a generic partial mitigating doctrine.

28.4.2.3 *Help adjudicate an individual case by providing evidence of the folk psychological criteria*

Again, the proponent of using the evidence would have to show that the neurodata confirm or challenge the presence of folk psychological criterial mental states and actions.

For example, suppose there was clear evidence that a person lacked the neural capacity that supports the folk psychological process of controlling one's intention to act in a prohibited way. This would be relevant to criminal responsibility for the insanity defence or sentencing purposes in some jurisdictions. How much lack of capacity would be required to excuse or mitigate could not be answered by neuroscience. It is a normative issue that folk psychological lawmakers and law appliers would have to address. Now, most cases will not be so clear and actions speak louder than images. Again, however, the use of neuroscience evidence to help adjudicate individual cases according to folk psychological legal criteria presupposes folk psychology.

28.4.2.4 *Help to implement a current policy more effectively*

The question in such cases is whether using neurodata would be cost-benefit justified. For example, criminal sentencing and many civil and quasi-criminal

¹² Stephen J. Morse, 'Diminished Rationality, Diminished Responsibility' (2003) 1 *Ohio St. J. Crim. L.* 289.

commitment laws contain a prediction of future behaviour criterion. This would provide the easiest and most straightforward case for relevance because, in such cases, mental state analysis is not crucial and neural markers might increase predictive accuracy. Notice that such doctrines address not the person being predicted, but the people who must decide how to respond to a particular prediction, a folk psychological decider performing the folk psychological process of deciding. Neuroscience might also enhance the accuracy of predicting who will benefit from diversion and from treatment programmes for those sentenced or civilly committed. Again, even if neurodata are useful in principle, is there sufficient value-added to justify the cost of collecting it?

28.4.3 External challenges from neuroscience

Recall that any external challenge to doctrines, institutions or practices suggests that they are incoherent and unjustifiable *ab initio*. This part addresses the two most radical external challenges from neuroscience: the challenge from neurodeterminism and the threat to the law's concept of the person, which grounds its concept of responsibility. The first challenge is familiar and has a good compatibilist response. The second is more radical and threatens the nature of law itself. The second fails on conceptual and empirical grounds, however. The coherence of responsibility and competence doctrines and practices are safe. At least for now.

28.4.3.1 *The challenge from neuroscientific determinism*

Many think that the proof of neuroscientific determinism undermines the legitimacy of all blaming practices, which are crucial to the criminal law. As the earlier section on distractions indicated, this is a mistake, but it is an external claim because it challenges the legitimacy of present criminal law as a whole and not just specific aspects of it. There is a profound metaphysical question about the truth of determinism, but this question is not itself resolvable by neuroscience. Further, as we have seen, the truth of determinism is fully consistent with the law's view of the person and all legal doctrines. The truth of determinism is not inconsistent with the view that mental states matter to the causation of behaviour and that human beings are capable of being guided by reason, including the law's commands. Let us therefore turn to a genuine challenge.

28.4.3.2 *The challenge to personhood*

The claim advanced by many that mental states are epiphenomenal and do no explanatory work presents a challenge to the coherence of all law, and not just to responsibility practices. If the concept of mental causation that underlies folk psychology and current conceptions of responsibility is false, our responsibility practices, and many others, would appear unjustifiable.

Such extreme claims are not strawpersons. Here is a lengthy quote from a widely quoted article by neuroscientists Joshua Greene and Jonathan Cohen that expresses the mechanistic conception:¹³

[A]s more and more scientific facts come in, providing increasingly vivid illustrations of what the human mind is really like, more and more people will develop moral intuitions that are at odds with our current social practices... Neuroscience has a special role to play in this process for the following reason. As long as the mind remains a black box, there will always be a donkey on which to pin dualist and libertarian positions... What neuroscience does, and will continue to do at an accelerated pace, is elucidate the 'when,' 'where' and 'how' of the mechanical processes that cause behaviour. It is one thing to deny that human behaviour is purely mechanical when your opponent offers only a general philosophical argument. It is quite another to hold your ground when your opponent can make detailed predictions about how these mechanical processes work, complete with images of the brain structures involved and equations that describe their function... At some further point... [p]eople may grow up completely used to the idea that every decision is a thoroughly mechanical process, the outcome of which is completely determined by the results of prior mechanical processes. What will such people think as they sit in their jury boxes? Will jurors of the future wonder whether the defendant... *could have done otherwise?* Whether he really *deserves* to be punished...? We submit that these questions, which seem so important today will lose their grip in an age when the mechanical nature of human decision-making is fully appreciated. The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless.

Alternatively, to use another of Greene and Cohen's arguments, suppose that 'neuroscience holds the promise of turning the black box of the mind into a *transparent bottleneck*.'¹⁴ They mean that the brain is the final mechanistic pathway through which all types of explanations of behaviour must ultimately operate and that neuroscience will be able to demonstrate that brain mechanisms, not mental states, are doing all the work.¹⁵ They speculate that we may some day possess 'extremely high-resolution scanners that can simultaneously track the neural activity and connectivity of every neuron in the human brain...' and that, with the help of computers and software, can help people see the neural events that are alone causally responsible for their behaviour.¹⁶ If such mechanistic understanding and knowledge were available and widespread, Greene and Cohen are probably correct that notions of responsibility would wither away because most would believe that

¹³ Joshua Greene and Jonathan Cohen, 'For the Law, Neuroscience Changes Nothing and Everything' in S. Zeki and O. Goodenough (eds.), *Law & the Brain* (New York: Oxford University Press, 2006) 207, 217–8.

¹⁴ *Ibid.* at 218.

¹⁵ *Ibid.*

¹⁶ I will assume that the scanning and computing abilities that the argument employs are possible, although the brain has 10^{11} cells and at least 10^{15} connections. Is it really likely, however, that the computer would predict what precise sentences we would speak? At present, of course, the speculation is pure science fiction and, in my opinion, is likely to remain so. The real problem with the argument is not that it assumes a (barely) plausible computational ability, but that it assumes that mental states can be reduced simply to brain states, an assumption that this subsection addresses.

it was the brain that 'did it', not the agent, and we do not hold brains morally responsible.

This picture of human activity exerts a strong pull on the popular, educated imagination as well as on the theorizing of scientists. Consider the following example. In an ingenious recent study,¹⁷ investigators were able to predict accurately based on which part of the brain was physiologically active whether a shopper-subject would or would not make a purchase. Activity in these regions predicted immediately subsequent purchases 'above and beyond self-report variables'.¹⁸ As we shall see in this subsection, this does not mean that the person's weighing of preferences and prices and the final decision played no role. Activity in the nucleus accumbens, the insula, and the mesial prefrontal cortex is not 'weighing' and 'deciding'. The latter are the activities of people, not brains.¹⁹ The findings interestingly, although unsurprisingly, suggest, however, that specific brain regions play a crucial role in particular types of psychological processes.

This study was reported in the 'Science Times' section of the *New York Times* by John Tierney.²⁰ Here is how the story was 'spun', beginning with its title: 'Findings: The Voices in My Head Say "Buy It!" Why Argue?'. The shopper is simply the hapless puppet of brain processes and plays no role as an agent in the purchase process. The decision is not up to the shopper; it is up to their brain. The conclusion considers how the study might help us deal with feckless consumerism:

You might remove the pleasure of shopping by somehow dulling the brain's dopamine receptors so that not even the new Apple iPhone would get a rise in the nucleus accumbens, but try getting anyone to stay on that medication. Better the occasional jolt of pain. Charge it to the insula.²¹

In addition to getting the study wrong—insula activation was associated with excessive prices and the decision not to purchase²²—it betrays once again the mechanistic view of human activity. What people do is simply a product of brain regions and neurotransmitters. The person disappears. There is no shopper. There is only a brain in a mall.

What if all these thinkers who claim that we are just victims of neuronal circumstances (VNCs) are correct? Suppose neuroscience convinces us that agency and folk psychology are an illusion, that intentional bodily movements and reflexes are morally indistinguishable because both are simply the outcomes of mechanistic

¹⁷ Brian Knutson et al., 'Neural Predictors of Purchases' (2007) 53 *Neuron* 147.

¹⁸ *Ibid.* at 147.

¹⁹ M. R. Bennett and P. M. S. Hacker, 'Philosophical Foundations of Neuroscience: An Excerpt from Chapter 3', in Maxwell Bennett et al. (eds.), *Neuroscience & Philosophy: Brain, Mind & Language* (New York: Columbia University Press, 2007) 15, 18–23 (describing ascription of psychological attributes to the brain as 'senseless'). But see Daniel Dennett, 'Philosophy as Naive Anthropology: Comment on Bennett and Hacker' in *ibid.* at 73, 86–8 (claiming that it makes sense to attribute 'attenuated' sorts of psychological attributes to parts of the brain).

²⁰ John Tierney, 'Findings: The Voices in My Head Say "Buy It!" Why Argue?' (2007) *New York Times*, 16 January, F1.

²¹ *Ibid.* at F6. ²² *Ibid.*

biophysical processes? What if all the contending conceptions about responsibility depend on a mistake about human activity? What if, for example, mental states do not explain actions but are simply *post hoc* rationalizations the brain creates to 'make sense of' the bodily motions or non-motions that brains produce? We are just mechanisms, although the illusion of conscious will may play a positive role in our lives.²³ Some people, including many psychologists and neuroscientists, think that new discoveries about the causation of behaviour are leading inexorably to a purely mechanistic view of the link between the brain and behaviour, and thus to a purely mechanistic view of human behaviour. Will the agentic person disappear and be replaced by the biological victim of neuronal circumstances?

If Greene and Cohen are right, we are all allegedly 'merely victims of neuronal circumstances.' But are we? And will criminal justice system as we know it, which includes robust notions of personhood and desert, wither away as an outmoded relic of a prescientific and cruel age? And not only criminal law is 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 was simply the outcome of various 'neuronal circumstances'. Although I predict that we will see far more numerous attempts to introduce neuroevidence in the future, the dystopia that Greene and Cohen predict is not likely to come to pass.

It is important to note from the outset, however, that compatibilism or other responses to the determinist challenges to responsibility will not save the disappearing person. Determinism is consistent with either of two inconsistent views of human behaviour. The truth of determinism is consistent with the existence or non-existence of agency, with the causal role or non-causal role of mental states in explaining behaviour. Compatibilism presupposes that a folk psychological account of action is accurate and that distinctions based on it, such as the difference between actions and non-actions, should make a moral and legal difference. The new, VNC claims deny precisely this. The person and responsibility can only be saved if VNC is false or, if it is true, we learn to live with the illusion that it is false. Otherwise, all agency-based conceptions of responsibility must be abandoned.

At present, no such radical, external challenge from neuroscience even remotely approaches plausibility. It is true that the law's fundamental presuppositions about personhood and action are open to profound objection. Most fundamentally, action and consciousness are scientific and conceptual mysteries.²⁴ We do not

²³ This claim should not be confused with the apparently similar claim that 'personhood' is an illusion. See Martha J. Farah and Andrea S. Heberlein, 'Personhood and Neuroscience: Naturalizing or Nihilating?' (2007) 7 *Am. J. Bioethics* 37, 40 (claiming that our construct of 'personhood' is simply the illusory product of innate and automatic brain systems that is 'projected' onto the world). There are many problems with the logic of this claim, but even if it is correct, it does not deny that creatures like us have mental states, such as desires and beliefs, that can be causally explanatory. Most charitably interpreted, it simply denies the explanatory usefulness of the normative concept of a 'person'.

²⁴ See Robert Audi, *Action, Intention and Reason* (Ithaca: Cornell University Press, 1993) 1–4 (describing the 'basic philosophical divisions' in each of the four major problem areas in action theory); Colin McGinn, *The Mysterious Flame: Conscious Minds in a Material World* (New York: Basic Books, 1999) (describing the immense difficulty of explaining consciousness and doubting the ability of human beings to do so).

know how the brain enables the mind²⁵ and we do not know how action is possible. At most we have hypotheses or *a priori* arguments. Moreover, causation by mental states seems to depend on now largely discredited mind–brain dualism that treats minds and brains as separate entities that are somehow in communication with one another.²⁶ How can such tenuously understood concepts be justifiable premises for legal practices such as blaming and punishing? If our picture of ourselves is wrong, as many neuroscientists claim, then our responsibility practices are morally unjustified according to any moral theory we currently embrace. On the other hand, given how little we know about the brain–mind and brain–action connection, to claim based on neuroscience that we should radically change our picture of ourselves and our practices is a form of neuroarrogance.

To see in more specific detail why we need not abandon our robust conception of agency despite such claims, let us turn to the indirect and allegedly direct evidence for them. The real question is whether scientific and clinical investigations have shown that agency is rare or non-existent; that conscious will is largely or entirely an illusion. Four kinds of indirect evidence are often adduced: first, demonstrations that a very large part of our activity is undeniably caused by variables we are not in the slightest aware of; second, studies indicating that more activity than we think takes place when our consciousness is divided or diminished; third, laboratory studies that show that people can be experimentally misled about their causal contribution to their apparent behaviour; and, fourth, evidence that particular types of psychological processes seem to have associated neurophysiological activity in specific regions of the brain. None of these types of evidence offers logical support to VNC, however.

Just because a person may not be aware of all the causes for why he formed an intention does not mean that he did not form an intention, that he was not a fully conscious agent when he did so, and that his intention played no causal role in explaining the person's behaviour. Even if human beings were never aware of the causes of their intentions to act and of their actions, it would not necessarily follow that they were not acting consciously, intentionally and for reasons that make eminent sense to anyone under the circumstances.

Human consciousness can undeniably be divided or diminished by a wide variety of normal and abnormal causes.²⁷ We have known this long before contemporary scientific discoveries of what causes such states and how they correlate with brain structure and processes. Law and morality agree that if an agent's

²⁵ Paul R. McHugh and Philip R. Slavney, *The Perspectives of Psychiatry*, 2nd edn (Baltimore: Johns Hopkins University Press, 1998) 11–12.

²⁶ It is almost impossible not to talk 'dualistically' in ordinary speech and writing. Every time a monist neuroscientist uses a personal pronoun in speaking or writing, for example, he seems to imply that there is a genuine 'him' or 'her' that is somehow distinguishable from his brain activity. This does not mean, however, that the neuroscientist (or anyone else) is really a crypto-dualist. It is simply an inevitable feature of current language, and perhaps it always will be.

²⁷ See Jeffrey L. Cummings and Michael S. Mega, *Neuropsychiatry and Behavioral Neuroscience* (New York, Oxford University Press, 2003) 333–43 (description of dissociative and related states and their causes and treatments); D. Vaitl, N. Birbaumer et al., 'Psychobiology of Altered States of Consciousness' (2005) 131 *Psychol. Bull.* 98.

capacity for consciousness is non-culpably diminished, responsibility is likewise diminished. Some suggest that it is diminished because bodily movements in the absence of fully integrated consciousness are not 'actions'.²⁸ Others believe that apparently goal-directed behaviour that is responsive to the environment, such as sleepwalking, is action, but that it should be excused because diminished consciousness reduces the capacity for rationality.²⁹ Let us assume that the former view is correct, because it offers more direct support to VNC and therefore the greatest challenge to traditional notions of individual responsibility. Let us also assume that divided or diminished consciousness is more common than it appears to be. Nevertheless, neither of these assumptions supports the more radical, general VNC thesis and the arguments for automatistic imperialism have been termed the 'automaticity juggernaut'.³⁰

Demonstrating that divided or partial consciousness is more common than it appears certainly extends the range of cases in which people are not responsible or have diminished responsibility. Such studies do not demonstrate, however, that most human bodily movements that appear intentional and rational (apparently rational actions) occur when the person has altered consciousness. One cannot generalize to all human behaviour from genuinely deviant cases or from cases in which a known abnormality is present. A model of action (or, we should say, non-action) built on sleepwalking, for example, is hardly a threat to orthodox notions of individual responsibility.

There is substantial empirical evidence to suggest that laboratory manipulations of unsuspecting subjects can cause the subjects to believe that their intentions were producing action when this was not the case.³¹ That subjects can be cleverly misled by experimental manipulations hardly indicates that intentions generally play no role in explaining our behaviour. Self-deception under laboratory conditions of deceit does not entail that intentions generally do not causally explain action. Universal deception about personal causal efficacy hardly seems to be a plausible evolutionary outcome.

Finally, there is accumulating evidence that various psychological processes have associated neural activity in localized regions of the brain. We have long known that many behavioural activities are biologically based in highly specific regions. For example, there is substantial evidence that ability to recognize faces is localized in a region of the temporal lobe of the right hemisphere referred to as the 'fusiform face area'. Should this area become lesioned, the subject loses the ability to recognize

²⁸ See, e.g., Michael S. Moore, *Act and Crime* (New York, Oxford University Press, 1993) 49–52, 135–55, 257–8 (arguing that cases of compromised consciousness should be treated as non-action); see also Michael S. Moore, 'More on Act and Crime' (1994) 142 *U. Pa. L. Rev.* 1749, 1804–20.

²⁹ Stephen J. Morse, 'Culpability and Control' (1994) 142 *U. Pa. L. Rev.* 1587, 1641–52 (arguing that clouded consciousness should be treated as an affirmative defence); see also Bernard Williams, 'The Actus Reus of Dr. Caligari' (1994) 142 *U. Pa. L. Rev.* 1661 (arguing that human activity with clouded consciousness is action).

³⁰ John F. Kihlstrom, 'The Automaticity Juggernaut—or, Are We Automatons After All?' in John Baer, James C. Kaufman, and Roy F. Baumeister (eds.), *Are We Free? Psychology and Free Will* (New York: Oxford University Press, 2008) 155–73.

³¹ See John A. Bargh, 'Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior' in Ran R. Hassin, James S. Uleman, and John A. Bargh (eds.), *The New Unconscious* (New York: Oxford University Press, 2005) at 37, 51–4 (reviewing the evidence and concluding that the 'will' is not primarily responsible for action).

faces, a condition called prosopagnosia.³² Now, however, functional neuroimaging techniques permit the exploration of brain activity during more complicated psychological processes and can identify associated neurophysiological activation for the processes. I have already discussed the example of brain regions associated with decisions to purchase an object.³³ For another example, a recent study demonstrated that investigators could determine from the region of brain activity which mental process—adding or subtracting—a subject had covertly intended to, but had not yet, performed.³⁴

The localization evidence is immensely interesting and suggestive, but it does not indicate that mental states play no role in causally explaining behaviour. There must be a biological substrate in the brain for all human behaviour. If your brain is dead, you are dead and not behaving at all. Nor is it surprising that particular regions of the brain are associated with particular psychological processes. For example, a leading, albeit controversial, theory of how the mind works suggests that it is composed of different systems that perform different functions.³⁵ Although we do not know how the brain enables the mind, it makes sense to assume that specific psychological processes would have brain substrates specific to each individual process. Based on what we already know about localization and on the reasonable assumption that it would be inefficient if all regions of the brain needed equal activation to support all psychological processes, localization is most likely to be true. Even if all this is correct, however, it does not follow that mental states do no causal explanatory work. It demonstrates at most that the neural network substrates for specific mental functions may be located in specific regions of the brain.

What is needed to support VNC is a general, direct demonstration that causal intentionality is an illusion *tout court*, but no such general demonstration has yet been produced by scientific study. The most interesting evidence has arisen from studies done by neuroscientist, Benjamin Libet,³⁶ which have generated an

³² Elinore McKone, Kate Croke, and Nancy Kanwisher, 'The Cognitive and Neural Development of Face Recognition in Humans' in Michael S. Gazzaniga (ed.), *The Cognitive Neurosciences*, 4th edn (Cambridge, MA: MIT Press, 2009) 467, 468–9; Daniel Trench, 'Functional Neuroanatomy: Neuropsychosis Correlates of Cortical and Subcortical Damage' in Stuart C. Yudovsky and Robert Hales (eds.), *Neuropsychiatry and Clinical Neuroscience*, 4th edn (Washington, DC: American Psychiatric Publishing, 2002) 71, 80–2.

³³ See notes 17–19, above, and accompanying text.

³⁴ John-Dylan Haynes et al., 'Reading Hidden Intentions in the Human Brain' (2007) 17 *Current Biology* 323, 323–8. It is important to recognize that the brain activity accurately predicted only which *type* of process the subject had covertly formed the intention to perform. It did not identify the specific content of the intention, such as which two numbers the subject intended to add or subtract. Despite the enormous advances in cognitive neuroscience, we do not know how to read minds using neuroimaging or any other technique. Cf. Martha J. Farah, 'Bioethical Issues in the Cognitive Neurosciences' in *The Cognitive Neurosciences* note 32, above, at III, 1309, 1309–10 (referring to the ability to identify traits and states as 'a crude form of mindreading').

³⁵ See, e.g., Jerry A. Fodor, *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983) (providing a modular theory).

³⁶ Benjamin Libet, 'Do We Have Free Will' in Benjamin Libet et al. (eds.), *The Volitional Brain: Towards a Neuroscience of Free Will* (Exeter: Imprint Academic, 1999) 47 (summarizing the findings and speculating about their implications). For a more recent, powerful demonstration of a similar finding, see C. S. Soon et al., 'Unconscious Determinants of Free Decisions in the Human Brain' (2008) 11 *Nature Neuroscience* 543.

immense amount of comment.³⁷ Indeed, many claim that Libet's work and later, similar studies are the first direct neurophysiological evidence of VNC.³⁸ Libet's exceptionally creative and careful studies demonstrate that measurable electrical brain activity associated with intentional actions occurs in the relevant motor area of the brain about 550 milliseconds before the subject actually acts and for about 350–400 milliseconds before the subject is consciously aware of the intention to act.

Let us assume, with cautious reservations,³⁹ the basic scientific methodological validity of these studies.⁴⁰ The crucial question then becomes whether the interpretation of these findings as supporting VNC is valid. Michael Moore has usefully shown that the Libetian conception of the role of brain events in causing behaviour is confused.⁴¹ Indeed, it is not clear precisely what the claim is, but the most profound challenge would be that mental states are epiphenomenal. If this is true, the folk psychological basis for all law is incoherent. Alfred Mele has shown that Libet's work does not establish VNC, has exposed numerous confusions, and has usefully described the type of experiment that might achieve this result.⁴² Rather than repeat their analyses, which bear close reading, this section will instead offer a more empirical and common-sense critique.

It does not follow from Libet's discovery of the temporal ordering that conscious intentionality does no causal work. It simply demonstrates that non-conscious brain events precede conscious experience. Once again, we have no idea how the brain enables the mind, but this seems precisely what one would expect of the mind-brain. Electrical impulses move quickly among neurons, but some lag between brain activity and conscious experience seems unsurprising. Once again, if the brain is dead, the person is dead. Prior electrical activity does not mean that intentionality played no causal role. Electrical activity in the brain is precisely that: electrical activity in the brain and not a mental state such as a decision or an intention. A readiness potential is not a decision.⁴³ A perfectly plausible reading

³⁷ Daniel Wegner, *The Illusion of Conscious Will* (Cambridge, MA: The MIT Press, 2002) 54–5 (characterizing the recounting of Libet's results as a 'cottage industry' and noting the large and contentious body of commentary).

³⁸ William P. Banks and Susan Pockett, 'Benjamin Libet's Work on the Neuroscience of Free Will' in Max Velmans and Susan Schneider (eds.), *The Blackwell Companion to Consciousness* (Malden, MA: Blackwell, Publishing, 2007) 657, at 658.

³⁹ See, e.g., Henrik Walter, *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy* (Cynthia Klor trans.) (Cambridge, MA: MIT Press, 2001) 250–2; Jing Zhu, 'Reclaiming Volition: An Alternative Interpretation of Libet's Experiment' (Nov. 2003) 61 *J. Consciousness Stud.* 1–77.

⁴⁰ Banks and Pockett, note 38, above, at 659–62 (concluding after a careful review of possible artefacts that 'readiness potentials do start before the subject consciously 'decides' to move').

⁴¹ Michael S. Moore, 'Libet's Challenges to Responsible Human Agency' in Walter Sinnott-Armstrong and Lynn Nadel (eds.), *Conscious Will And Responsibility: A Tribute to Benjamin Libet* (New York: Oxford University Press, 2010), ch. 18.

⁴² Alfred R. Mele, *Effective Intentions: The Power of the Conscious Will* (New York: Oxford University Press, 2009). See also, M. R. Bennett and P. M. S. Hacker, *Philosophical Foundations of Neuroscience* (Malden, MA: Blackwell Publishing, 2003) 228–31 (criticizing Libet's account of action).

⁴³ Moreover, Libet does not carefully distinguish between urges or wants on the one hand and decisions and intentions on the other. Indeed, Alfred Mele argues that the experimental evidence is

of Libet's work is that various non-conscious causal variables, including non-conscious urges, precede action—who would have thought otherwise?—but intentionality is nonetheless necessary for action.

Libet also suggests that people can 'veto' the act during the delay between becoming aware of the intention and performing the intended action, which he surprisingly conceives of as an undetermined act. Other researchers appear to have localized the part of the brain that is associated with the activity of vetoing.⁴⁴ But, in addition to the implausibility of the veto being undetermined,⁴⁵ the conceptual foundations of the interpretation that the subjects were exercising a genuine veto are shaky at best.⁴⁶ This suggestion undermines the claim that the brain is doing all the work because it is an agent's mental state, a newly formed intention to veto, that causes the agent not to perform the act. In short, Libet's work presupposes agency at every step in the process.

Libet's task involved 'random' finger movements that involved no deliberation whatsoever and no rational motivation for the specific movements involved.⁴⁷ This is a far cry from the behavioural concerns of the criminal law or morality, which address intentional conduct when there is always good reason to refrain from harming another or to act beneficently. In fact, it is at present an open question whether Libet's paradigm is representative of intentional actions in general because Libet used such trivial behaviour.⁴⁸

In addition to direct problems with the alleged implications of Libet's work, there are also good reasons to reject it. Answers to the possibility of VNC are rooted in common sense, a plausible theory of mind, our evolutionary history, and practical necessity. Virtually every neurologically intact person consistently has the experience of first person agency, the experience that one's intentions flow from one's desires and beliefs and result in action. Indeed, this folk-psychological experience is so central to human life and so apparently explanatory that it is difficult to imagine giving it up or a good reason to do so, even if it were possible to give it up. As the eminent philosopher of mind, Jerry Fodor, has written:

[I]f commonsense intentional psychology were really to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're that wrong about the mind, then that's the wrongest we've ever been about anything. The collapse of the supernatural, for example, didn't compare... Nothing except, perhaps, our

much more consistent with the RP being associated with an urge rather than an intention or a decision: Alfred R. Mele, *Free Will and Luck* (New York: Oxford University Press, 2006) 33,40. I am not convinced that this problem is major, but associating the RP with 'desire' rather than intention perhaps weakens the case that Libet's work establishes NAT.

⁴⁴ See, Maurice Brass and Patrick Haggard, 'To Do or Not to Do: The Neural Signature of Self-control' (2007) 27 *J. Neurosci.* 9141 (identifying the part of the brain that is activated when the 'veto' is exercised).

⁴⁵ Banks and Pockett, note 38 above, at 667.

⁴⁶ Mele, note 42 above, at 34–5.

⁴⁷ Participating in the study and cooperating with the investigator can be rationally motivated, of course. But the experimental task was to move one's finger randomly, for no good reason.

⁴⁸ Banks and Pockett, note 38 above, at 662–3.

commonsense physics... comes as near our cognitive core as intentional explanation does. We'll be in deep, deep trouble if we have to give it up...

... But be of good cheer; everything is going to be all right.⁴⁹

Folk psychology has much explanatory power and is capable of scientific investigation.⁵⁰ There is compelling psychological evidence that intentions play a causal role in explaining behaviour.⁵¹ Finally, despite Mele's attempt, it is hard to imagine the nature of a scientific study that would prove conclusively that mental states do no work to creatures who have created and will assess that study with mental states.

The plausible theory of mind that might support mental state explanations is thoroughly material, but non-reductive and non-dualist. It hypothesizes that all mental and behavioural activity is the causal product of lawful physical events in the brain, that mental states are real, that they are caused by lower level biological processes in the brain, that they are realized in the brain—the mind-brain—but not at the level of neurons, and that mental states can be causally efficacious.⁵² It accepts that a fully causal story about behaviour will be multifield and multilevel.⁵³

There is a perfectly plausible evolutionary story about why folk psychology is causally explanatory and why human beings need rules such as those provided by law. We have evolved to be self-conscious creatures that act for reasons and are reasons responsive. Acting for reasons is inescapable for creatures like ourselves who inevitably care about the ends they pursue and about what reason they have to act in one way rather than another.⁵⁴ Because we are social creatures whose interactions are not governed primarily by innate repertoires, it is inevitable that rules will be necessary to help order our interactions in any minimally complex social group.⁵⁵ As a profoundly social species, it seems apparent that our ancestors would have been much less successful, and therefore much less likely to be our ancestors, if they were unable to understand the intentions of others, not sure they could convert their intentions into action, and were not also equipped with powerful assumptions that

⁴⁹ Jerry A. Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: MIT Press, 1987) xii.

⁵⁰ See, e.g., Bertram F. Malle, *How The Mind Explains Behavior: Folk Explanations, Meaning and Social Interaction* (Cambridge, MA: MIT Press, 2004) (providing a full theoretical account and empirical support). There is also growing recognition within psychology that 'mental-state inference is one of the most fundamental tools of social cognition'. Bertram F. Malle, 'Folk Theory of Mind: Conceptual Foundations of Human Social Cognition' in Ran R. Hassin, James S. Uleman, and John A. Bargh (eds.), *The New Unconscious* (New York: Oxford University Press, 2005) 225, 229.

⁵¹ See, Richard Holton, *Willing, Wanting, Waiting* (New York: Oxford University Press, 2009) 5–9 (reviewing psychologist Peter Gollwitzer's work and explaining how it supports the role of a distinct psychological kind, intention, as playing a causal role in behaviour); Mele, note 42 above, at 134–6.

⁵² See, e.g., John R. Searle, *Mind: A Brief Introduction* (New York: Oxford University Press, 2004) at 113–4 (terming his position 'biological naturalism' about consciousness).

⁵³ See, Carl Craver, *Explaining the Brain: Mechanisms and the Mosaic of Neuroscience* (New York: Oxford University Press, 2007) (providing an account of how neuroscience fits into a multifield, multilevel framework for explaining behaviour).

⁵⁴ Hilary Bok, *Freedom and Responsibility* (Princeton: Princeton University Press, 1998) at 75–91, 129–31, 146–51.

⁵⁵ Larry Alexander et al., *The Rule of Rules: Morality, Rules and the Dilemmas of Law* (Durham, NC: Duke University Press, 2001) 11–25 (explaining why rules are necessary in a complex society and contrasting their account with H. L. A. Hart's theory).

that stranger coming over the hill is equipped with the same capacity for harmful intentions as they are.⁵⁶

One of the qualities that makes us most human is the ability to infer that others have independent mental states and then to use that information to understand and predict the behaviour of others. Psychologists call this having a 'theory of mind'.⁵⁷ Human beings who do not develop an adequate 'theory of mind', such as those with autism, experience profound difficulties in their interpersonal lives. The lengthy conservation of mental states and their ubiquitousness and centrality suggest that they do play an important causal role and that they are very evolutionarily expensive if they do not. This is of course not an incontrovertible analytic argument against VNC, but surely the burden of persuasion is on those who argue to the contrary. At the very least, we remain entitled to presume that conscious intentions are causal until the burden is met.

Libet's work is fascinating, but it does not prove that humans are generally not conscious, intentional agents or capable of employing their conscious intentionality when they have good reason to do so.⁵⁸ Even if the work is methodologically valid, various conceptual and interpretive arguments undermine the claim that Libet has demonstrated that VNC is true.

In short, despite the often astonishing findings and impressive advances in neuroscience and allied disciplines, there is no compelling evidence yet that VNC is generally true. Do any of the foregoing arguments entail that mental states are not epiphenomenal? Of course not. But we are nowhere close to demonstrating epiphenomenalism experimentally (indeed, it would be hard to imagine what that study would look like), the best theories about reduction suggest that it is probably false, and evolution and common sense suggest that it is false. Future discoveries may undermine this conclusion, however, so I now turn to the implications of VNC.

VNC alas can provide no guidance about what people should do next and, in any event, degenerates into self-referential incoherence. Suppose that you were convinced by the mechanistic view that you were not an intentional, rational agent after all. (Of course, the notion of being 'convinced' would be an illusion, too.⁵⁹ Being convinced means that you were persuaded by evidence or argument, but a

⁵⁶ See Justin N. Wood et al., 'The Perception of Rational, Goal-Directed Action in Nonhuman Primates' (2007) 317 *Science* 1402, 1405 (demonstrating that the ability to understand the intentions of other creatures evolved in primates 40 million years ago); see also Esther Herrmann et al., 'Humans Have Developed Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis' (2007) 317 *Science* 1360 (comparing chimpanzees and orang-utans to two-and-a-half-year-old humans and discovering that they have approximately equal cognitive skills concerning the physical world, but that humans have superior cognitive skills for understanding social interaction).

⁵⁷ Geraldine Dawon and Karen Toth, 'Autism Spectrum Disorders' in Dante Cicchetti and Donald J. Cohen (eds.), *Developmental Psychopathology*, 2nd edn, *Risk, Disorder and Adaptation* (Hoboken, NJ: John Wiley & Sons, 2006) vol. 3, 327.

⁵⁸ See Jerry Fodor, 'Making the Connection' (2002) *Times Literary Supplement*, 17 May, at 4 (arguing that the new neuroscience rarely has much to contribute when the phenomenon in question is complex social behaviour).

⁵⁹ See Daniel C. Dennett, 'Calling in the Cartesian Loans' (2004) 27 *Behav. Brain Scis.* 661, 661 (wondering, in response to Professor Wegner, who is this 'we' that inhabits the brain).

mechanism is not persuaded by anything. It is simply neurophysically transformed.) What should you do now? You know that it is an illusion to think that your deliberations and intentions have any causal efficacy in the world. (Again, what does it mean according to the purely mechanistic view to 'know' something? But enough.) You also know, however, that you experience sensations such as pleasure and pain and that you care about what happens to you and to the world. You cannot just sit quietly and wait for your neurotransmitters to fire. You cannot wait for determinism to happen. You must, and will of course, deliberate and act.

If one still thought that VNC was correct and that standard notions of genuine moral responsibility and desert are therefore impossible, one might nevertheless continue to 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.⁶⁰ Through poorly understood automatic processes, it is possible that various potential rewards and punishments would shape behaviour even if they did not do so as premises in practical reasoning. Such an account would be consistent with 'black box' accounts of economic incentives. For those who believe that a thoroughly naturalized account of human behaviour entails complete consequentialism, such a conclusion might not be unwelcome.

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 thought and agents, not a property of brains—play in this 'decision'? Once again, the VNC account seems to swallow itself. Moreover, VNC proponents of consequentialism could hardly complain about those who refuse to 'accept' what the proponents think rationality requires. The allegedly misguided people who resist are simply the victims of their automatic brain states. They cannot be expected intentionally to use their capacity for reason to accept what the consequentialists believe reason demands. Indeed, the consequentialist's belief is also an illusory mental state or it exists but plays no role in explaining behaviour.

Even if our mental states play no genuinely causal role (about which we will never be certain until we solve the mind-body problem) human beings will find it almost impossible not to treat themselves as rational, intentional agents unless there are major changes in the way our brains work. Moreover, if one uses the truth of pure mechanism as a premise in deciding what to do, this premise yields no particular moral, legal, or political conclusions. It will provide no guide to how one should live or how one should respond to the truth of VNC. If reasons, which are mental states, are epiphenomenal and normativity depends on reason, VNC is normatively inert.⁶¹

⁶⁰ Greene and Cohen, note 13, above, at 218.

⁶¹ I was first prompted to this line of thought by Mitchell Berman's discussion of determinism and normativity. See, Mitchell Berman, 'Punishment and Justification' (2008). 118 *Ethics* 258, 271 n. 34.

28.5 Case Studies

This section considers case studies that will illuminate the conceptual and abstract theses of the preceding sections. It begins with examination of the legal implications of two recent, widely-noticed and excellent neuroscience studies that appear to have legal relevance. Then it considers a body of neuroscience research that many people think is useful to the law. Finally, it addresses the criminal responsibility of an individual paedophile for whom there was compelling evidence that a brain tumour was the source of the defendant's paedophilic urges and behaviour.

28.5.1 The neural correlates of third-party punishment⁶²

In scenarios involving third-party criminal punishment, the right dorsolateral prefrontal cortex (rDLPFC) was activated when the sixteen subjects decided that the harmdoer should be punished because there was culpability—a result consistent with previous findings that rDLPFC is activated when people make punishment decisions in two-party games. It also found that the amygdala, a region in an older portion of the brain that is associated with the expression of emotion and affective processing, was activated when the subjects decided how much to punish.

The study is compelling, but one can fairly raise objections to various aspects of the language of the study that are common but which overclaim or mislead. For example, it claims that it 'elucidated the neural dynamics that underlie human altruistic punishment'; that 'prefrontal and parietal activity is modulated by a punishment-related decisional process'; that 'the two fundamental components of third-party legal decision making... are not supported by a single neural system'; and that 'these findings seem to highlight an important conceptual overlap between moral reasoning and legal reasoning...'. The study does show which brain regions are apparently activated when subjects perform specific tasks, but this is different from 'elucidating neural dynamics'. If this means that the study demonstrates the causally explanatory interactions between neurons or neural networks, it is false. If it means simply that there is a correlation between activation in some regions and specific tasks—as the title of the study properly indicates—it is true but the language appears to make a more expansive claim. 'Modulated' suggests causal understanding that we do not possess. 'Supported by a single neural system' again goes beyond our understanding. It may well be true, but the language suggests an understanding of causation, of necessary and perhaps sufficient neural conditions for the behaviour, that we in fact lack. Finally, the suggestion that these neural correlates 'highlight an important conceptual overlap' between two types of reasoning is a category mistake because it confuses the positive—the brain

⁶² J. W. Buckholz et al., 'The Neural Correlates of Third-Party Punishment' (2008) 60 *Neuron* 930. The study deservedly received an admiring comment in the same issue of the journal in which the study appeared. J. Haushofer and E. Fehr, 'You Shouldn't Have: Your Brain on Others' Crimes' (2008) 60 *Neuron* 738.

findings—with the normative—the concepts of moral and legal reasoning. Even if precisely the same brain regions were activated during moral and legal-reasoning tasks, it would not follow that they are not conceptually distinct.

The interesting question is the legal relevance of the study. It is unsurprising that two different types of decisions are associated with activation in different brain regions. This study does help confirm the apparent similarities between two-party and three-party norm-violation punishment decisions, and it is thus a contribution to understanding the evolution of punishment practices. But what is its current legal relevance? It does not tell us who to punish or how much. In other words, it does not contribute to the substance of crucial normative questions in criminal law.⁶³

The prepotent rDLPFC response to punish and the role of the amygdala in such decisions are also not surprising because everyone understands that virtually all human beings will experience emotional responses when making decisions about whether punishment is warranted and how much to punish norm violators. This might create concerns about whether judges may sometimes be influenced by emotional factors and then decide unfairly or with implicit bias. But we already knew they sometimes do this and deciding which decisions are not fair or impartial is a normative question of folk psychological practical reasoning that cannot be read off from the brain (unless we found a perfect neural marker for those decisions that we had already decided on normative grounds were not fair, which is a fantasy at present and perhaps for the future). Assume that a particular judge showed particularly active amygdala activity when sentencing. What follows?

In short, this is first-rate cognitive neuroscience, but it is legally inert.

28.5.2 The neural correlates of moral decision-making in psychopathy⁶⁴

In a task involving emotional decision-making with seventeen subjects, those with higher psychopathy scores showed decreased amygdala activation and those scoring particularly high on the interpersonal factor for psychopathy (manipulation, conning, superficiality, and deceitfulness) also exhibited reduced activity in what is called the general 'neural moral circuit'. Decreased activation in these areas is associated with deficits in the ability to envision harming others and more generally with deficits involving the proper use of emotions and understanding of interpersonal processes. We already possess a well-characterized understanding and operationalized measure of the emotional and interpersonal deficits of psychopaths. We understand that they cannot use conscience, empathy, and similar qualities to guide

⁶³ See Selim Berker, 'The Normative Insignificance of Neuroscience' (2009) 37 *Philosophy & Public Affairs* 293; Frances Kamm, 'Neuroscience and Moral Reasoning: A Note on Recent Research' (2009) 37 *Philosophy & Public Affairs* 330.

⁶⁴ A. L. Glenn, A. Raine, and R. A. Schug, 'The Moral Correlates of Moral Decision-making in Psychopathy' (2009) 14 *Molecular Psychiatry* 5. Publication of this article was followed by a useful exchange about it. See, S. Tassy et al., 'Do Psychopathic Patients Use their DLPFC when Making Decisions in Moral Dilemmas?' (2009) 14 *Molecular Psychiatry* 908; A. L. Glenn et al., 'Increased DLPFC Activity during Moral Decision-making in Psychopathy' (2009) 14 *Molecular Psychiatry* 909.

their conduct. Nevertheless, the criminal law provides no excuse or mitigation for this condition. Do these findings imply that we should?

First, it is again unsurprising that the brain activity of psychopaths doing moral judgement tasks differs from the activity of non-psychopathic controls. Their behaviour is different and different brain activity is consistent with different behaviours. Even assuming that the brain findings were causal and not correlational, should these data cause us to excuse or mitigate the criminal conduct of psychopaths?

All behaviour has associated brain activity and even abnormal brain activity is not per se an excusing condition. Nor is brain causation the equivalent of compulsion or lack of self-control. The criteria for excuse—roughly, lack of rational capacity or the capacity for self-control—are folk psychological and we already understood the folk-psychological differences between people with and without psychopathy. (Whether we consider the psychopathic characteristics 'deficits' is a normative question.) Moreover, most informed observers already concluded that these deficits are not the psychopath's own fault. It is how they are constituted.⁶⁵ Thus, even without the brain findings, we knew everything we needed to know to argue normatively about whether psychopaths ought to be excused or mitigated. We already knew what we had to know about psychopaths' rational and control capacities, which are properties of folk psychological people, not brain criteria. The study does provide weakly convergent evidence for the validity of psychopathy as an identifiable condition, but that only becomes important if we had already decided on normative grounds that psychopathy should have legal relevance. In short, this is excellent cognitive neuroscience, but it is legally inert once again.

28.5.3 Executive functioning

Executive function is the term applied to a wide range of abilities that enable purposive, goal-oriented, successful behaviour.⁶⁶ These include the capacities to initiate and plan behaviour, to focus attention, and to self-monitor and self-regulate, including inhibition of inappropriate desires. Defects in these functions, which stem from lateral prefrontal cortex (LPFC) malfunction, tend to be global and to affect the person's behaviour generally. Not all functions need be impaired, however. The extent of impairment depends on the specific pathology involved. A person with such defects may be excitable, impulsive, and erratic, or, in the alternative, avolitional, perseverating, and with flattened affect. Narrowly construed

⁶⁵ I will not address whether psychopathy is a disorder that exists and can be properly diagnosed among children and adolescents. I simply assume that the characteristics adult psychopaths exhibit are not their fault but are the product of causal variables beyond their control.

⁶⁶ Jeffrey L. Cummings and Bruce L. Miller, 'Conceptual and Clinical Aspects of the Frontal Lobes' in Jeffrey L. Cummings and Bruce L. Miller (eds.), *The Human Frontal Lobes: Functions and Disorders*, 2nd edn (New York: Guilford Press, 2007) 12, 15–18; Joaquin M. Fuster, *The Prefrontal Cortex*, 4th edn (Boston: Academic Press, 2008) 178ff.; Joel H. Kramer and Lovingly Quitania, 'Bedside Frontal Lobe Testing' *ibid.* at 2279, 279–85; Muriel D. Lezak et al., *Neuropsychological Assessment*, 4th edn (New York: Oxford University Press, 2004) 35–7. There is some variation among writers concerning the characterization of executive functions, but the description I will give is common and sufficient for our purposes.

cognitive functioning may not be impaired, but, for example, the person's ability to use the intelligence and knowledge he possesses is diminished. These problems can impair the capacity for normal, independent life, including the ability to have successful interpersonal relations and to avoid inappropriate and unlawful behaviour.

One might well expect to find defects in such functions, whether or not associated with clear organic pathology, in large numbers of people who violate the criminal law or who have trouble meeting various competence criteria. Such defects, especially of the type involving disinhibition and poor planning, surely might play a causal role because they make it more difficult to behave well. Nevertheless, how are these impairments legally relevant. Simply having defects in rational or control capacities, no matter how they are caused, is not per se an excusing or mitigating condition. Indeed, there is no criminal law excuse for having control difficulties generally, but such difficulties may be relevant to legal insanity if they are a product of mental illness and to various forms of involuntary civil or quasi-criminal commitment.⁶⁷ Such defects may also be relevant to sentencing.

All these cases must be decided behaviourally, however, so what does localization in the LPFC of many of these defects add? In cases in which the defect is arguably sufficient to warrant a conclusion concerning mitigation, excuse or incompetence, the behavioural evidence will typically be so manifest that no neuroscience evidence will be needed.⁶⁸ Moreover, there is empirical reason to question whether tests of executive function explain much of the variance in the real world expression of executive function.⁶⁹ Therefore, neuroscience evidence concerning executive function is not likely to help resolve close cases for the general reasons discussed above concerning how neuroscience data are obtained, but that may change in the future if neuromarkers become more sensitive. Similarly, if the neuroscience becomes more sensitive, it might help decide cases in which there is concern about malingering. Most important, criminal responsibility and competence are normative legal criteria. Neurodata must therefore be considered in light of the behavioural data and normative considerations. It cannot resolve the legal question to which it may potentially be relevant.⁷⁰

It also does not seem that the neuroscience evidence presents a challenge to existing doctrine. It could, however, motivate expanding the categories of people who might be mitigated or excused if the neuroscience taught us that larger numbers of people have serious difficulty controlling their behaviour than we thought previously or if it showed that the behavioural evidence is misleading concerning control capacities. We do not have such data at present, but we might in

⁶⁷ E.g., *Kansas v Crane* 534 US 407 (2002) (requiring for mentally abnormal sexually violent predator commitments that the subject has 'serious difficulty controlling himself').

⁶⁸ See Lezak et al., note 66, above, at 36 ('Many of the behavior problems arising from impaired executive functions are apparent even to casual or naïve observers.')

⁶⁹ See, 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.

⁷⁰ Dean Mobbs et al., 'Law, Responsibility and the Brain' (2007) 5 *PLoS Biology* 0693, 96–7.

the future. Finally, neuromarkers of executive function defects may have marginal utility in helping to make prediction decisions for sentencing, parole and the like. Whether such markers are sufficiently helpful beyond behavioural methods to justify the cost of collecting the data is a normative question. For now, no such data exist, but it may in the future and it is possible that obtaining that data will be cost-benefit justified.

Understanding the role of brain abnormalities in producing executive function deficits may of course lead to treatments or other interventions that would prevent crime and other untoward behaviours. For now, however, neuroscience evidence of executive function deficits simply provides some hope for evidentiary assistance in the future. It suggests no major reforms of doctrine or practice.

28.5.4 The case of Mr Oft

This case was first reported in a neurology journal.⁷¹ Oft was a forty-year-old school teacher who was married and had a step daughter. He had an interest in pornography dating to his adolescence, but at the time in question he experienced a growing sexual interest in children and he collected child pornography and visited child pornographic Internet sites. He also solicited prostitution at 'massage parlours', which he had not previously done. Oft tried to conceal his activities because he knew that they were unacceptable. Nevertheless, he continued to act on his sexual impulses because, he said, the 'pleasure principle overrode' his restraint. Oft began to make subtle sexual advances to his prepubescent stepdaughter, who informed her mother.

Oft was convicted of child molestation and ordered to undergo an inpatient rehabilitation programme instead of prison. Despite his desire to avoid prison, he solicited sexual favours from staff and other patients in his programme and he was expelled.

The evening before his prison sentencing, Oft was admitted to a hospital emergency room complaining of headache. Although no physiologic cause was suspected, he was admitted on psychiatric grounds with a diagnosis of paedophilia. He expressed suicidal ideation and a fear that he would rape his landlady. During neurologic examination he solicited female staff for sexual favours and was unconcerned that he had urinated on himself. He had various neurological signs, including problems with his gait. Oft was alert and completely oriented. His memory was intact, his speaking and reading skills were unimpaired, and he was able to inhibit motor responses on a standard test of this ability. Word generation was somewhat impaired. He did suffer from constructional apraxia, the inability to assemble a coherent whole from its constituent elements, as demonstrated by his inability to draw a clock or to copy figures. He also could not write a legible sentence. A magnetic resonance imaging (MRI) test was performed.

⁷¹ Jeffrey M. Burns and Russell H. Swerdlow, 'Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign' (2003) 60 *Arch. Neurology* 437. The name Mr Oft is a pseudonym.

Oft had a large orbitofrontal tumour. The orbitofrontal cortex is involved in the regulation of social behaviour. Lesions acquired in this region later in life are associated with impulse control problems and antisocial conduct, but previously established moral judgement is preserved. The tumour was surgically removed and Oft quickly recovered bladder control and normal walking activity. Two days post surgery, his neurologic examination was essentially normal. Oft then successfully completed an outpatient treatment programme for his sexual disorder. He was no longer considered a threat and returned home. About a year later, he experienced a persistent headache and again began secretly collecting pornography. MRI showed tumour regrowth and the new tumour was successfully removed.

In their discussion of Oft's case, the authors said that Oft 'could not refrain from acting on his paedophilia despite the awareness that the behavior was inappropriate'.⁷² They hypothesized that the problem was caused by a disruption of his somatic marker system, which lead to a preference for short-term reward and thus impaired the 'subject's ability to appropriately navigate social situations'.⁷³

Although paedophilia is not a sufficient mental disorder to support an insanity defence, it is not absurd to think that perhaps Oft deserved mitigation or excuse for his sexual deviance on the ground that he could not control himself. With respect, however, we do not know whether Oft could not—that is, lacked the ability to—control his sexual behaviour, or whether he simply did not. Given the timing of the appearance of the sexual deviance and the tumour growth, we can be quite confident that the tumour played a causal role in producing and heightening his sexually deviant urges and in undermining his inhibitory processes.

The general legal question is how Oft is relevantly different from any other paedophile with similar urges and similar inhibitory controls? One assumption is that the sexual behaviour is a mechanistic product of the tumour and is thus just like the mechanistic sign of any other disease. This assumption begs the question of responsibility, however. Oft's desires may have been mechanistically caused, but acting on them was intentional action. An abnormal cause for his behaviour does not mean that he could not control his actions. This must be shown independently. We can reasonably infer that Oft had difficulty controlling behaviour that harmed himself because he acted in ways he knew would have negative consequences. But this is true of all paedophiles and we do not excuse them. He may have had impaired executive function, but this may also be true of many paedophiles and would again need to be established independently. Although there is reason to question whether Oft differs substantially from paedophiles generally, the temptation to respond to Oft differently is strongly influenced by the lure of mechanism.

Now let us be highly specific. Recall that the neurologists who examined him concluded that Oft could not refrain from acting on his paedophilic urges. Oft was clearly not responsible for having those urges in the first place. We do not know how strong his paedophilic urges were, however, nor do we know which inhibitory functions, if any, were compromised. We do know that Oft did not control his

⁷² *Ibid.* at 440.

⁷³ *Ibid.*

paedophilic and other sexual urges, including in circumstances in which it was unlawful or completely inappropriate to express them. Moreover, Oft understood that his behaviour was unacceptable and he reported that the pleasure principle overrode his inhibitions. It is reasonable to conclude based on common-sense inferences that Oft experienced substantial difficulty controlling himself, but how do we know that he lacked sufficient control capacity to deserve mitigation or excuse? Oft was firmly in touch with reality and fully understood the moral and legal rules. He understood that important interests were at stake and that he should not violate them. We do not know how firmly Oft resolved not to yield to his impulses or whether he took steps to restrain them. There is a hint in his comment about the pleasure principle that he took no such steps.

At least in the beginning when the paedophilic urges emerged, Oft was able to conceal many of his unacceptable activities, which often requires inhibiting one's response to desires, and his spouse evidently noted nothing amiss in his behaviour. Oft's paedophilia and pornographic interests were discovered only when his step-daughter reported his sexual advances. Later on, however, the picture becomes more mixed. Most of the neuropsychological testing done in the hospital prior to surgery that bears on his executive functioning was normal. His memory was intact, and Oft 'verbally shifted between letter and number sets, conceptualized, performed sequential hand movements, and inhibited motor responses on the Luria go-no go test'.⁷⁴ On the other hand, complete neuropsychological testing of frontal lobe function was not performed because prompt surgical intervention was necessary. Such testing might have disclosed other executive function deficits. Further, during his neurological examination, Oft appeared unconcerned that he had urinated on himself, which suggests some type of irrational disconnection from his situation, and he solicited hospital personnel for sexual favours. Perhaps most importantly, Oft's sexual behaviour in his treatment programme threatened his ability to avoid prison. In short, Oft's ability to control himself apparently deteriorated quite rapidly and became markedly worse by the time of hospitalization compared to the time of the offence.

We know a great deal about the cause of Oft's paedophilic desires and have a plausible causal account of how his executive function might have been undermined. Brain causation, even by such a manifest abnormality, still does not answer our question about how difficult it was for him to control himself. The case study authors proposed a 'somatic marker' hypothesis to explain behaviour that is similar to acquired sociopathy. If true, this helps explain why a hitherto continent agent began to act for immediate gratification and with insufficient regard for future consequences. Put another way, it helps explain why his judgement was impaired and poor judgement makes controlling oneself more difficult. Nevertheless, it does not tell us how impaired Oft's judgement was or the role such impairment played in explaining his inappropriate and criminal behaviour.

⁷⁴ Burns and Swerdlow, note 71, above, at 438. The Luria test examines how well the subject is able to inhibit a prepotent response.

Finally to decide whether Oft deserves mitigation or excuse we must use all the considerations just reviewed and come to an all-things-considered normative evaluation about his capacities at the time of the crime. This is a standard question of deciding when a person had sufficient capacity to resist temptation. To use the capacious Model Penal Code language, did Oft lack substantial capacity to understand the wrongness of his actions or to conform his conduct to moral and legal norms? These capacities range along a continuum. At the time of the crime, Oft was clearly in the grey area between the kinds of situation in which virtually every person would have trouble controlling undesirable behaviour and those in which virtually no one would fail to exercise control. I will leave readers to decide the question for themselves, but the answer cannot be based on the presence of abnormal brain causation *per se*. Oft's case might elicit sympathy for his plight, which was terrible luck, and it certainly suggests that a medical rather than punitive response might be cost-benefit justified. But these considerations are distinct from whether Oft deserved mitigation or excuse.

28.6 Conclusion

The relation between neuroscience and legal doctrine and practice is conceptually fraught. Neuroscience has the potential to make internal contributions to legal doctrine and practice if the relation is properly understood. For now, however, such contributions are modest at best and neuroscience poses no genuine, radical challenges to concepts of personhood, responsibility, and competence.