

## ARTICLE

**“Free Will, If It Exists”:  
How Neuroscience Challenges Moral Responsibility**

CHARLIE COX\*

Recent developments in neuroscience have important implications for criminal justice. In shining light on the causes behind human behaviour, such developments challenge a commonly held assumption about human nature: that at any one moment, we are free to do otherwise. This article explores how that assumption is embedded in our criminal justice system in the concept of retributivism. This article argues that developments in neuroscience challenge that concept both in principle and in practice. In principle, neuroscience invites a reconsideration of how punishment is justified, pointing towards a consequentialist approach to criminal justice. In practice, the developments expose decision-makers to mechanistic explanations for criminal behaviour, which makes them more likely to see their retributivist impulses decline.

**I Introduction**

We are likely not free in the way that we commonly perceive. Indeed, there is a fundamental tension between the sense that we are free at any moment to do otherwise—known as the principle of alternative possibilities—and the materialist worldview of cause and effect that underpins modern day science.<sup>1</sup> Whether the world is dictated by the deterministic processes of cause and effect generated by the Big Bang, the indeterminacy of quantum physics, or a combination of the two, there is little room for the prevailing view

---

\* BA/LLB(Hons), Victoria University of Wellington. Many thanks to Professor Yvette Tinsley for her encouragement and thoughtful comments on earlier versions of this article.

1 Paul W Glimcher “The Neurobiology of Individual Decision Making, Dualism, and Legal Accountability” in Christopher Engel and Wolf Singer (eds) *Better Than Conscious? Decision Making, the Human Mind, and Implications for Institutions* (MIT Press, Cambridge (Mass), 2008) 343 at 357.

that we are free agents.<sup>2</sup> If our brains and thoughts are governed by prior causes over which we have no control, we cannot claim that we are free to choose how we act. There is, simply put, no compelling reason to expect that human brains are isolated from the laws of the physical universe in a way that other complex systems are not.

Whether free will exists is a debate that is no longer confined to the murky realm of metaphysics. Today, research in neuroscience increasingly seeks to shed light on the terrain which has long been contested by philosophers. While inconclusive, this research lends credence to the already compelling case that our brains are inseparable from biology and the physical laws of cause and effect that govern it.

This research has important implications for the way in which societies should respond to criminalised behaviour. Advancements in neuroscience challenge what David Eagleman terms the “fault line” model of blameworthiness.<sup>3</sup> Under this model, a fault line, the position of which is determined by the extent of current scientific knowledge, separates criminalised behaviour into two categories.<sup>4</sup> On one side of the line sits behaviour for which there is an exculpatory scientific explanation. Such behaviour is not, under the model, appropriately answered by punitive criminal sanction. On the other side is the behaviour where free will fills the gap left by an absence of scientific explanation, and which therefore calls for punishment.

As research in neuroscience progresses, however, and more biological explanations for criminal behaviour are uncovered, the fault line model of blameworthiness is challenged, and the retributive justification for punishment is undermined. Rather than shifting the fault line each time a new explanation exculpates behaviour previously thought of as criminally blameworthy, it is time to consider whether such a model is coherent. If the reasons underpinning a person’s actions stem from their neurochemistry and environment, and are thus beyond their control, it is difficult to justify not only our current understanding of moral blameworthiness, but also the fault line model itself and the libertarian theory of freedom which underlies it.

This article suggests that modern neuroscience invites us to challenge our understanding of moral responsibility, both in principle and in practice. In principle, the failure of neuroscience to substantiate what we think of when we discuss free will invites a consequentialist approach to criminal justice. The research, in its current stage, is difficult to reconcile with the penal populism which often animates criminal justice policy in Aotearoa New Zealand. In practice, even if commonly held intuitions about freedom remain static, as neuroscience reveals more about the microcircuitry of the brain that governs behaviour, it is likely that the retributivist impulses of criminal justice decision-makers will decline.

---

2 Joshua Greene and Jonathan Cohen “For the law, neuroscience changes nothing and everything” (2004) 359 *Phil Trans R Soc Lond B* 1775 at 1780.

3 David Eagleman *Incognito: The Secret Lives of the Brain* (Pantheon Books, New York, 2011) at ch 6.

4 At 174–176.

## II What Neuroscience Says about Free Will

### A *The readiness-potential*

Doubts regarding free will perhaps first emerged in the neuroscientific community following the results of a now-infamous experiment conducted by Benjamin Libet and colleagues.<sup>5</sup> In Libet's experiment, subjects were hooked up to an electroencephalography machine, which records brain activity, and told to perform a simple action—such as lifting their finger or flexing their wrist—at a time of their choosing. Subjects were to note the exact moment they felt the urge to move.<sup>6</sup> Professor Libet found that unconscious neural activity—the “readiness-potential”—preceded subjects' awareness of an intention to act.<sup>7</sup> The readiness-potential occurred several hundred milliseconds prior to when subjects reported experiencing conscious intention.<sup>8</sup> Based on these findings, some have argued the experience of volition is just an after-the-fact mental rationalisation of unconscious decision-making processes: the brain made the decision before the subject experienced the urge to act.<sup>9</sup> Arguably, if all decisions are caused by such prior unconscious neural activity, freedom of choice is illusory.<sup>10</sup>

Libet's findings have been controversial.<sup>11</sup> However, his experiments have been replicated and built upon to address a range of critiques. For example, a more recent and sophisticated experiment, conducted by Chun Siong Soon and colleagues, measured the neural activity of subjects through functional magnetic resonance imaging (fMRI) while a stream of letters was shown to them on a screen.<sup>12</sup> The subjects were asked to “freely decide” to press one of two buttons, operated by different fingers. At the same time, the subjects were to remember which letter was on a screen at the time they experienced deciding which button to press.<sup>13</sup> Rather than focusing on the readiness-potential as Libet did, Soon found the *outcome* of the choice could be encoded in the activity of the prefrontal and parietal regions of the brain.<sup>14</sup> The experimenters could thus predict the outcome of a choice not yet consciously made up to 10 seconds before the subject was aware of having made a decision. This result is a much longer time window between neural activity and reported conscious volition than was observed by Libet.<sup>15</sup>

---

5 See, for example, Victoria Saigle, Veljko Dubljević and Eric Racine “The Impact of a Landmark Neuroscience Study on Free Will: A Qualitative Analysis of Articles Using Libet and Colleagues' Methods” (2018) 9 *AJOB Neuroscience* 29 at 29.

6 Benjamin Libet and others “Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act” (1983) 106 *Brain* 623 at 624–625.

7 At 623.

8 At 623.

9 Emad H Atiq “How Folk Beliefs about Free Will Influence Sentencing: A New Target for the Neuro-Determinist Critics of Criminal Law” (2013) 16 *New Criminal Law Review* 449 at 456.

10 Farah Focquaert, Andrea Glenn and Adrian Raine “Free Will, Responsibility, and the Punishment of Criminals” in Thomas A Nadelhoffer (ed) *The Future of Punishment* (Oxford University Press, New York, 2013) 247 at 248.

11 See, for example, Alfred R Mele *Free Will and Luck* (Oxford University Press, New York, 2006) at ch 2.

12 Chun Siong Soon and others “Unconscious determinants of free decisions in the human brain” (2008) 11 *Nature Neuroscience* 543.

13 At 543.

14 At 543.

15 At 543.

This later study appears to address concerns expressed in relation to the Libet experiment: that the readiness-potential could simply reflect non-specific preparatory brain activity connected to concentrating on the experiment, or the human error of experiment subjects.<sup>16</sup> While concerns have been raised about the limited predictive power that the experimenters actually had (they could guess the correct button about 60 per cent of the time),<sup>17</sup> a more recent study predicted when a decision to raise a finger occurred with over 80 per cent accuracy, 700 milliseconds before the decision was consciously perceived.<sup>18</sup>

Perhaps the most enduring and challenging critique of Libet's experiment, and those of its ilk, is that such experiments are restricted to simple, morally neutral decision-making—what Alfred Mele calls “Buridan situations”.<sup>19</sup> These experiments do not, so the critique goes, relate to the morally charged, complex decisions that are of interest to criminal justice theorists. Surely, critics argue, actions like flexing a wrist, raising a finger or pressing a button are not comparable to the drawn-out, deliberative decision-making required, for example, when deciding to commit arson or stage a bank robbery. For this reason, Mele writes that Libet only “studies free will ... in the sphere of proximal decision making in Buridan situations or situations of a similar kind”, and to generalise from these simplistic situations to criminal offending, would be “extremely bold”.<sup>20</sup>

Nonetheless, the research suggests our unconscious brain is causally responsible for the outcome of decisions we consider to be “paradigmatically voluntary”.<sup>21</sup> Intuitively, it is difficult to envisage a decision more replete with free will—that is, less constrained by biology and environment—than deciding between which of two different buttons to press. It seems, then, the proper interpretation of the Professor Libet experiments balances on the edge of Ockham's razor. Should we view the limited predictions of the experiments as a function of limited computing power and imperfect methods of observing the brain? Or should such limitations support a yet unsubstantiated claim that, despite all that we know about the laws of cause and effect that govern the physical world, human brains are somehow exempt from them? This article suggests that the former has both simplicity and science on its side.

### B *Brain or biology?*

Because behaviour is a function of biology, “it no longer makes sense to ask, ‘To what extent was it his *biology* and to what extent was it *him*?’”<sup>22</sup> As Joshua Greene and Jonathan Cohen observe, in a naturalistic world, there is “no ‘homunculus’, in the brain that is the real you behind the mass of neuronal instrumentation”.<sup>23</sup> And yet, on questions of moral responsibility, such “doublethink” is not uncommon.<sup>24</sup> Unless someone's problematic behaviour can be attributed to adverse environmental influences—which in turn affect the

---

16 Focquaert, Glenn and Raine, above n 10, at 249.

17 Adina L Roskies “How Does Neuroscience Affect Our Conception of Volition?” (2010) 33 *Annu Rev Neurosci* 109 at 116.

18 Itzhak Fried, Roy Mukamel and Gabriel Kreiman “Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition” (2011) 69 *Neuron* 548 at 548.

19 Alfred R Mele *Effective Intentions: The Power of Conscious Will* (Oxford University Press, New York, 2009) at 85.

20 At 85.

21 Atiq, above n 9, at 456.

22 Eagleman, above n 3, at 176–177.

23 Greene and Cohen, above n 2, at 1779.

24 At 1779.

mental processes that govern behaviour—or some discernible neural pathology, the law assumes that behaviour results from free will, rendering a person blameworthy. For this reason, while laypeople tend to believe in the possibility of “doing otherwise”,<sup>25</sup> retributivist urges are diminished when exposed to mechanistic explanations for human behaviour.<sup>26</sup>

Free will scepticism is therefore reinforced by findings in neuroscience that link behaviour to biology. Where neuroscience can trace behaviour to its organic origins, an actor’s agency is diminished, and moral exculpation appears necessary. For example, where epilepsy was once seen as a sign of demonic possession, the seizures it causes are attributable to “uncontrolled and synchronous neuronal discharges”.<sup>27</sup> Huntington’s disease, which typically manifests around the age of 40, can involve drastic changes in personality, such as impulsiveness, aggression and hyper-sexuality, but results from the mutation of a single gene.<sup>28</sup>

However, these examples bely the multifactorial causes of behaviour. A wide range of environmental factors also influence the neurobiology which ultimately causes behaviour. For example, childhood stress can permanently affect the brain’s ability to prevent secretion of glucocorticoids. An excess of glucocorticoids, especially during development, can negatively affect cognition, impulse control and empathy.<sup>29</sup> Similarly, by age five, the lower a child’s socio-economic status, the higher their basal glucocorticoid levels, and the thinner their prefrontal cortex.<sup>30</sup>

As neuroscientist Robert Sapolsky notes, damage to the prefrontal cortex provides another biological explanation for aberrant behaviour. There are strong links between damage to the prefrontal cortex and increases in impulsiveness, antisocial behaviour and criminality.<sup>31</sup> Even for criminals lacking obvious prefrontal cortex damage, research suggests their prefrontal cortex function is different: for instance, the basal metabolic rate of a sociopath’s prefrontal cortex tends to be lower than average.<sup>32</sup> Further, when experimental subjects with damaged prefrontal cortexes are presented with a “right” and “wrong” option, they may verbalise an intent to pick the right one but nonetheless impulsively choose the neurologically easier “wrong” option.<sup>33</sup> Thus, those with prefrontal cortex damage may know the “right” thing to do—but their neurophysiology renders them unable to do it.<sup>34</sup>

Writing in this context, Sapolsky argues the law adheres to a “false dichotomy” between blameworthy and excused behaviour.<sup>35</sup> Biology exculpates only to an extent (for example, if an epileptic hits someone during a seizure):<sup>36</sup>

---

25 Atiq, above n 9, at 480.

26 See Azim F Shariff and others “Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution” (2014) 25 *Psychological Science* 1563.

27 Robert M Sapolsky “The frontal cortex and the criminal justice system” in Semir Zeki and Oliver Goodenough (eds) *Law and the Brain* (Oxford University Press, New York, 2006) 227 at 239.

28 At 227.

29 Robert M Sapolsky *Behave: The Biology of Humans at Our Best and Worst* (Penguin Press, New York, 2017) at 187.

30 At 187.

31 Sapolsky, above n 27, at 238.

32 At 238.

33 At 238.

34 At 239.

35 At 239.

36 At 239.

But it is as if, with that area of organic impairment identified and given credence, we expect it to be bounded, and for the rest of our ‘us-ness’, replete with free will, to now shoulder the responsibility of keeping that organic impairment within the confines of its boundaries.

Sapolsky suggests this dichotomy is false, because the very *containment* of such organic impairments can also be explained in terms of “reductive, materialistic neurobiology”.<sup>37</sup> Neuroscientific research on the prefrontal cortex thus sits uncomfortably with the strict legal categorisation embodied, for example, in the rules for insanity set out in *M’Naghten’s Case*, which require that the defendant, by reason of some mental incapacity, be unable to understand the nature and quality of their behaviour.<sup>38</sup> In law, compulsion exculpates, but causation does not.<sup>39</sup> In reality, a person may know the difference between right and wrong, but be biologically incapable of doing the right thing. As Sapolsky observes, there is no justification for this distinction that does not “tacitly require a homunculus that is outside the causal universe, a homunculus that can be overwhelmed by ‘compulsion’ but that can and should handle ‘causation’”.<sup>40</sup>

In Aotearoa New Zealand, the Misuse of Drugs Amendment Act 2019 reflects an ongoing shift in the “fault line” of criminal responsibility. The Act affirms the existing police discretion to prosecute for the possession and use of controlled drugs.<sup>41</sup> The amended Misuse of Drugs Act 1975 now invites decision-makers to consider whether a “health-based” or “therapeutic” approach would be more beneficial to the public interest.<sup>42</sup>

The reform was undoubtedly enabled by developments in neuroscience that explain sources of human behaviour. While drug addiction was once thought of as an “evil” to be condemned,<sup>43</sup> the Misuse of Drugs Amendment Act reflects the emerging view that “[d]rug addiction is a health condition: people who take drugs need medical help, not criminal retribution”.<sup>44</sup> This shift was aided by insights provided by neuroscience about the neural architecture of addiction. Repeated drug use causes physical changes in the brain, reconfiguring the neural circuitry of reward and decision-making, which in turn undermines voluntary control.<sup>45</sup> While some contend drug addicts have access to rational capacity when they are not “in the throes of peak desire or craving”,<sup>46</sup> another view is that addiction impacts on a variety of neural circuits which enable self-monitoring and social functioning.<sup>47</sup> It is the “functional overlap” in the neural systems that addiction impacts, together with genetic and environmental factors, that explain the cycle of drug abuse

---

37 At 239.

38 *M’Naghten’s Case* (1843) 10 Cl & Fin 200 at 210, 8 ER 718 (HL) at 722.

39 Stephen J Morse “Excusing and the New Excuse Defenses: A Legal and Conceptual Review” (1998) 23 Crime and Justice 329 at 350.

40 Sapolsky, above n 29, at 600.

41 Misuse of Drugs Amendment Act 2019, s 6.

42 Misuse of Drugs Act 1975, s 7(6).

43 *Robinson v California* 370 US 660 (1962) at 669–670.

44 United Nations Office of Drugs and Crime *World Drug Report 2009* (United Nations, New York, 2009) at 2.

45 Nora D Volkow and Ting-Kai Li “Drug Addiction: the neurobiology of behaviour gone awry” (2004) 5 Nature Reviews Neuroscience 963 at 963.

46 Stephen J Morse “Psychopathy and Criminal Responsibility” (2008) 1 Neuroethics 205 at 211.

47 Nora D Volkow, Ruben D Baler and Rita Z Goldstein “Addiction: Pulling at the Neural Threads of Social Behaviors” (2011) 69 Neuron 599 at 602.

prevalent even amongst those who wish to become drug-free.<sup>48</sup> It is for this reason, some contend, policy underpinned by the assumption that drug use is a “choice”, is problematic.<sup>49</sup>

### III Implications for the Criminal Justice System

Modern neuroscientific findings reinvigorate the age-old debate regarding free will and moral responsibility. Some argue that if neuroscience is able to disprove free will, it would spell “the greatest intellectual catastrophe in the history of our species”.<sup>50</sup> For others, “neuroscience contributes nothing more than new details” to then fit within the existing framework of the legal system.<sup>51</sup>

The truth, this article argues, is somewhere in the middle. In principle, the idea that free will has an inconsequential role to play in human behaviour only challenges the retributivist rationale for punishment—not the existence of punishment itself. In practice, because a layperson’s “folk understanding” of free will—embodied by the libertarian theory of freedom—is irreconcilable with modern neuroscience, developments may affect the exercise of discretion in the criminal process, as well as the moral credibility of retributivist approaches.

#### A *In principle*

Developments in modern neuroscience pull towards the conclusion that we lack libertarian free will. Libertarianism (of no relation to the political philosophy) is incompatibilist: it holds that free will and determinism are incompatible.<sup>52</sup> Libertarians believe that because determinism is false, we have free will in the sense of possessing the ability to do otherwise.<sup>53</sup> Studies indicate this view most closely reflects the “folk” notion of freedom.<sup>54</sup>

Many scientists and philosophers, however, write off libertarianism as “panicky metaphysics” because it involves the idea that human brains operate outside of the currently understood laws of physics.<sup>55</sup> In a deterministic world, which envisages a predetermined chain of causes and effects set in motion by the universe’s beginning, there is no room for a view of human beings as “uncaused causers”, existing independently from the “deterministic flow of the universe”.<sup>56</sup> Brains are made up of physical matter, which is governed by the laws of physics. For that reason, there is little sense in the idea that our brains—and the bodies they animate—can cause events to take place, without themselves being caused by prior causes that are outside of our control. And yet that is the very view of human nature implied by libertarianism. Consider an everyday example: if it were true

---

48 At 599–600.

49 See William R Kelly “Substance Abuse, Neuroscience, and Crime” (26 April 2018) Psychology Today <[www.psychologytoday.com](http://www.psychologytoday.com)>.

50 Jerry Fodor *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (MIT Press, Cambridge (Mass), 1987) at xii.

51 Greene and Cohen, above n 2, at 1775.

52 At 1776.

53 At 1776.

54 Shaun Nichols “Folk Intuitions on Free Will” (2006) 6 *Journal of Cognition and Culture* 57 at 65–67.

55 Greene and Cohen, above n 2, at 1777.

56 At 1781.

that, when deciding whether or not to have a coffee in the morning, a person could have instead chosen tea, then their decision-making process would need to be (at least in part) uncaused—because if it were wholly caused, it would be determined.

Furthermore, the randomness of quantum indeterminacy provides little aid to the libertarian cause. If the outcome of choice is dictated by (1) the prior state of the universe; (2) the laws of physics; and (3) subatomic uncertainty, then it is still not free.<sup>57</sup> As Eagleman observes, “[i]t’s either coin flips or billiard balls”—neither of which provides the ability to do otherwise.<sup>58</sup> Libertarian freedom thus requires human agency to transcend the limits of the (currently understood) natural world, which must be doubted.<sup>59</sup>

Granted, no science, including neuroscience, can conclusively demonstrate the existence or not of free will.<sup>60</sup> This fact is especially true in the face of libertarian theories of freedom, which explain free will as an “unmeasurable force that cannot be studied with physical methods”.<sup>61</sup> However, if a materialist view is taken, the above-mentioned research in neuroscience supports free will scepticism. If neuroscientific progress continues, the shift between blame and biology will also continue as more “meaningful biological explanation[s]’ for deviant behaviour are uncovered.<sup>62</sup> The “fault line” sits at the “limits of current technology”.<sup>63</sup> But, as Sapolsky writes in relation to the prefrontal cortex, “[i]t would be the height of hubris to think that we have already learned how to detect the most subtle ways in which [prefrontal cortex] damage impairs volitional control.”<sup>64</sup> As science progresses, our understanding will move to the microscopic neural activity that precedes conduct. Rather than simply shifting the fault line each time neuroscience sheds light on behaviour previously thought of as blameworthy, perhaps it is time to consider whether the fault line model should be abandoned altogether.

Indeed, neuroscience need not definitively answer the question of free will to impact social policy. Summarising neuroscientific research thus far, Eagleman proposes the principle of sufficient automatism: “free will, if it exists, is only a small factor riding on top of enormous automated machinery”.<sup>65</sup> Even if libertarianism proves to be true, it will not affect the fact that biology and environment determine human behaviour almost completely without regard to “volition’s invisible hand”.<sup>66</sup>

Given the scientific speciousness of libertarianism and the “blistering” pace of neuroscientific progress, culpability should not be defined by the limits of current neuroscience.<sup>67</sup> Instead, the criminal justice system should treat every offender as incapable of having acted otherwise, assuming the inseparability between biology and decision-making.<sup>68</sup> The system should thus concede to the “fatuous tautology”<sup>69</sup> that

---

57 At 1777.

58 Eagleman, above n 3, at 168.

59 Greene and Cohen, above n 2, at 1777.

60 Stephen Morse “Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty” (2011) 3 Law, Innovation and Technology 209 at 216.

61 Glimcher, above n 1, at 357.

62 Eagleman, above n 3, at 176.

63 At 176.

64 Sapolsky, above n 27, at 240.

65 Eagleman, above n 3, at 170.

66 At 171.

67 Brian TM Mammarella “An Evidence-Based Objection to Retributive Justice” (2016) 16 YJHPL 289 at 321.

68 Eagleman, above n 3, at 177.

69 Sapolsky, above n 27, at 239.



there is always a neurobiological explanation for abnormal behaviour.<sup>70</sup> For Eagleman, deviant behaviour should be treated in the same manner as other physical conditions “such as diabetes or lung disease”.<sup>71</sup> It no longer makes sense to excuse behaviours which fits into existing legal defences, but treat other behaviours as empowered by free will.

This conclusion, however, is somewhat hasty. It assumes libertarian free will is a foundational pillar of criminal justice. The retort of compatibilists, who argue free will and determinism are compatible,<sup>72</sup> is that the criminal justice system only assesses the kinds of free will “worth wanting”.<sup>73</sup> That is, “[c]riminal law presupposes a ‘folk psychological’ view” of human behaviour, which perceives mental states such as desire and belief as fundamental to a full causal picture.<sup>74</sup> Because the criminal law assigns moral responsibility to actors that are “conscious, intentional, rational and uncompelled”, it is irrelevant that individuals do not possess the ability to do otherwise.<sup>75</sup> Criminal law thus operates notwithstanding the truth of determinism.

Given that folk psychology of the mental state is normative, and neuroscience research regarding the mental state is still emerging, Stephen Morse suggests we should be wary of the role of neuroscience. To radically revise the system of responsibility would, given current knowledge, amount to “neuroarrogance”.<sup>76</sup> Moreover, while neuroscience may provide new explanations for behaviour, it does not necessarily provide exculpation.<sup>77</sup> Explanations are relevant only insofar as they inform the folk psychology model: “syndromes and other causes do not have excusing force unless they sufficiently diminish rationality in the context”.<sup>78</sup>

However, it does not follow from Morse’s work that libertarian free will is absent from the law’s philosophical assumptions. Libertarian free will continues to be a prerequisite of a retributive theory of justice.<sup>79</sup> Retributivism maintains “that punishment is justified because the offender has voluntarily committed a morally wrong act”.<sup>80</sup> As the touchstone of retributivism is desert, the non-existence of free will poses a fundamental challenge.<sup>81</sup> If criminals are victims of neuronal circumstances beyond their control, as neuroscience currently suggests, they have not voluntarily committed a wrongful act. As Michael Gazzaniga argues, if determinism is true, retributive punishment is both nonsensical and immoral.<sup>82</sup> If libertarianism is false (or mostly false), justice can no longer be defined in retributivist terms.

---

70 Wolf Singer “Keiner kann anders, als er ist” *Frankfurter Allgemeine Zeitung* (Frankfurt, 8 January 2004) as cited in Eagleman, above n 3, at 177.

71 Eagleman, above n 3, at 170.

72 Greene and Cohen, above n 2, at 1777.

73 See generally Daniel C Dennett *Elbow Room: The Varieties of Free Will Worth Wanting* (2nd ed, The MIT Press, Cambridge (Mass), 2015).

74 Stephen J Morse “Neuroscience and the Future of Personhood and Responsibility” in Jeffrey Rosen and Benjamin Wittes (eds) *Constitution 3.0: Freedom and Technological Change* (Brookings Institution Press, Washington, 2011) 113 at 115.

75 At 120.

76 At 122.

77 At 119.

78 Stephen J Morse “New neuroscience, old problems” in Brent Garland (ed) *Neuroscience and the Law: Brain, Mind, and the Scales of Justice* (Dana Press, New York, 2004) 157 at 180.

79 Derk Pereboom *Free Will, Agency, and Meaning in Life* (Oxford University Press, Oxford, 2014) at 155.

80 CL Ten *Crime, Guilt, and Punishment* (Oxford University Press, New York, 1987) at 46.

81 Greene and Cohen, above n 2, at 1777.

82 Michael S Gazzaniga “The Law and Neuroscience” (2008) 60 *Neuron* 412 at 413.

This logic has led commentators to argue that punishment should instead be grounded in consequentialism.<sup>83</sup> Consequentialism is necessarily forward-facing: it justifies punishment through its future effects, such as public protection or deterrence.<sup>84</sup> As such, it can be applied notwithstanding one's view of free will.<sup>85</sup> Therefore, consequentialism does not require dispensing with a compatibilist view of free will. Indeed, as Robert Wright notes, free will has acted as a "rough proxy" for consequentialist justice.<sup>86</sup> Following a consequentialist approach, if Person A's criminal behaviour is thought to be freely willed (in the sense of being rational and unconstrained), it makes sense to punish them more than Person B, whose behaviour is captured by the *M'Naghten* rules for insanity, for example. In Person A's case, punishment is justified because, so the assumption goes, their rationality with respect to their criminal behaviour means they are responsive to incentives and can therefore be deterred or rehabilitated to prevent future offending. In Person B's case, this logic of punishment and reward is inapplicable, as they are incapable of understanding that their behaviour was wrong.

Along these lines, Eagleman argues that criminal sentences should be based on the "modifiability" of the deviant behaviour.<sup>87</sup> Punishment can be grounded in neuroplasticity: the biological capacity for change. So, for example, it would make sense to punish a child who draws on the wall while awake, but not a child who was sleepwalking at the time. This example, given by Eagleman, reflects the consequentialist intuition that in the latter case, the child cannot change their behaviour.<sup>88</sup> In a similar vein, in order to achieve the best consequences, different approaches are needed between those who commit crimes in the spur of the moment, and those who choose to cause harm while fully aware of the consequences of their behaviour.

Another example is white-collar crime. There is little evidence that the possibility of imprisonment specifically deters individuals as compared to noncustodial sanctions, or that individuals are affected by prison sentences increasing in severity.<sup>89</sup> Some argue this does not hold true for white-collar criminals, whose rational, profit-seeking orientation suggests they are likely to be attuned not only to the likelihood of apprehension but also to the severity of punishment.<sup>90</sup> It makes sense, therefore, to cater punishment to each respective offender type and to what is required to minimise harm to society. While increasing prison sentences for white-collar criminals may be justified under a consequentialist lens, the same cannot be said for other crimes.

One difficulty with this approach concerns the overlap between retributivism and restoration.<sup>91</sup> Eagleman writes that a criminal incapable of change "should simply be warehoused".<sup>92</sup> Care must be taken to balance the consequentialist goals of protecting

---

83 Greene and Cohen, above n 2, at 1783.

84 At 1776.

85 At 1777.

86 Robert Wright *The Moral Animal* (Pantheon Books, New York, 1994) at 347.

87 Eagleman, above n 3, at 188.

88 At 188.

89 Daniel S Nagin "Deterrence in the Twenty-First Century" (2013) 42 Crime and Justice 199 at 201.

90 See Paul H Robinson and John M Darley "The Role of Deterrence in the Formulation of Criminal Law Rules: At Its Worst When Doing Its Best" (2003) 91 Geo LJ 949 at 956. Contrast Peter J Henning "Is Deterrence Relevant in Sentencing White-Collar Criminals?" (2015) 61 Wayne L Rev 27.

91 Elisabeth McDonald and Yvette Tinsley (eds) *From "Real Rape" to Real Justice: Prosecuting Rape in New Zealand* (Victoria University Press, Wellington, 2011) at 417.

92 Eagleman, above n 3, at 189.

society from the offender with the prevention of disproportionality and maintaining the overall credibility of the justice system.

### B *In practice*

Even if the findings of modern neuroscience do not challenge all norms underpinning our criminal justice system, the research remains of practical import. Although compatibilism can ostensibly save criminal justice folk psychology from deterministic critique, the same cannot be said for the layperson's folk psychology. Research suggests that laypeople predominantly subscribe to a belief in libertarian freedom,<sup>93</sup> the very freedom challenged by the neuroscientific findings. This may distort their views of criminal behaviour.<sup>94</sup> As neuroscience continues to marginalise such freedom, the retributive impulses of decision-makers will proportionately decline.<sup>95</sup>

This hypothesis is grounded in reality. Individuals tend to exhibit a “fundamental attribution error”, where they downplay the extent to which situational causes, such as biology and environmental circumstances, cause behaviour, while exaggerating dispositional (that is, deterministic) causes, which “originat[e] from the person's character”.<sup>96</sup> However, while beliefs in libertarian freedom “appear to remain sturdy in the face of deterministic explanations for human behavior”,<sup>97</sup> retributive impulses decline as people are exposed to such deterministic explanations.<sup>98</sup> Even though neuroscience cannot comprehensively dismiss libertarian freedom, as it “becomes increasingly vivid and reductionistic, laypeople will become proportionally less willing to attribute blame, and these shifting societal intuitions will ultimately diminish criminal law's moral credibility”.<sup>99</sup>

It is difficult to reconcile these findings with the “tough on crime” punitiveness that pervades Aotearoa New Zealand and international criminal justice discourse. The extent to which such discourse may impede changing perceptions about behaviour requires its own empirical research. However, given that the rhetoric of free will and blameworthiness is often used as a justification for punitive reforms,<sup>100</sup> the political appeal of such rhetoric may diminish as “scientific explanations ... crowd out the possibility of human agency”.<sup>101</sup>

## IV Conclusion

Neuroscience challenges moral responsibility both in principle and in practice—but not to the extent that some suggest. While libertarian thinking struggles to maintain its intuitive logic in the face of the inextricable link between behaviour and biology, the folk psychology of criminal law remains—for now—undisturbed.

Nevertheless, neuroscience invites a reconsideration of the way punishment is justified. According to conventional retributivist theory, actors do not morally deserve punishment for behaviour that is outside of their control. A world devoid of libertarian

---

93 Mammarella, above n 67, at 294.

94 Atiq, above n 9, at 474.

95 Mammarella, above n 67, at 323.

96 Atiq, above n 9, at 476.

97 Mammarella, above n 67, at 321.

98 At 314.

99 At 289.

100 Atiq, above n 9, at 479.

101 Mammarella, above n 67, at 317.

freedom thus calls for a consequentialist approach to criminal justice. Of course, retributivist purists might brand this as a disincentive to moral responsibility. However, this is a consequentialist justification for retributivism—it is not grounded in moral desert. Ultimately, as Sapolsky observes, “although it may seem dehumanizing to medicalize people into being broken cars, it can still be vastly more humane than moralizing them into being sinners”.<sup>102</sup>

Finally, neuroscience challenges the layperson’s understanding of free will. Thus, even if belief in free will remains static, with neuroscience providing a more mechanistic explanation for human behaviour, it is likely that retributivist impulses will decline.

---

102 Sapolsky, above n 27, at 240.