

Neuroscience, Free Will and Criminal Justice

Madeleine Udell

May 5, 2009

Abstract

A retributive legal system relies on the premise that people freely choose their actions and should be punished accordingly. This sits uncomfortably with growing neuroscientific evidence that moral actions are mediated by structural features of the brain that are not freely chosen.

1 Introduction

In John Locke's Second Treatise on Government, he proposes the idea that governments exists in order to help their citizens secure their rights of "life, health, liberty, or possessions." As such, it seems that the legal system in a society should be designed to prevent crimes that tend to deprive citizens of these rights, and restore those goods which can be restored to the victims.

But strangely enough, our legal system does not seem to reflect the principle that laws should be designed to prevent crime. Instead, we tend to ask whether the criminal "deserves" a given sentence. We agree that the feeble minded and the very young are less deserving of punishment for a given crime, whereas adults in full possession of their faculties are fully responsible and should be punished to the fullest extent of the law.

Some neuroscientific evidence begins to make one a bit queasy when applying this kind of logic. Neuroscientists have shown that individuals who suffer brain lesions in certain parts of their brains (in particular, the prefrontal cortex) may commit morally reprehensible actions that would have been highly out of character before they incurred a brain injury. More troubling still is evidence that some people are born with smaller PFCs or with other structural or genetic deficits that predispose them to violence and crime. Are these people fully responsible for their actions? Do they deserve to be punished for their crimes? Or did their brain make them do it?

It is right to be queasy when talking of guilt given this neuroscientific evidence. But in fact, even the kind of determinism that comes out of physics is sufficient to raise questions about the validity of a retributive system of justice. Neuroscientific evidence makes these points more vividly, but adds no real information to the philosophical discussion of free will.

If we understand our justice system as being premised on the idea of preventing crime rather than on punishing the blameworthy, we can make full sense and use of this new neuroscientific data. In this paper I attempt to point the way to a rational justice system, and I explore how neuroscientific research might fit into a system of preventative justice.

2 “My brain made me do it”: Evidence from Neuroscience

The story of Phineas Gage is well known: a mild-mannered, hard-working railroad foreman became an impudent and crude no-goodnik after a railroad accident in which an iron bar pierced his skull, leaving behind massive brain damage in his prefrontal cortex (Harlow, 1848). More recent evidence confirms the role

of the prefrontal cortex in social and moral thinking. A study of Vietnam veterans found that a lesion to the PFC in one particular area was correlated with increased verbal, though not physical, aggression (Grafman et al., 1996). Still, it is relatively easy to interpret the patients in these studies as fundamentally “good people” whose brains were damaged by the outside world and should therefore be held less “responsible” for the actions they commit thereafter.

The cases get more morally ambiguous. A case study by Anderson et al documents two children who suffered damage to the prefrontal cortex before they were 16 months old (Anderson et al., 1999). Adults who suffer lesions of the PFC tend to understand social rules, if not always to obey them; these children, on the other hand, seemed to have no understanding of social norms. Neither would respond to punishment, or expressed guilt or remorse for misbehavior. The boy “suffered from occasional brief and explosive outbursts of anger” and “occasionally engaged in physical assault.” If the region of the brain that would have allowed them to learn the social norms they are violating has been irreparably damaged, it is easy to say that these children could not have done differently - that they cannot be blamed for their bad behavior.

The symptoms of the children with PFC damage are very similar to symptoms of psychopathy or sociopathy. In fact, there is considerable evidence that these disorders are linked to dysfunction in the PFC. One study found that subjects with antisocial personality disorder had on average 11% less grey matter in the PFC than did control subjects (either normal or psychiatric or drug-addicted controls) (Raine et al., 2000). Thus we have evidence in both directions: PFC damage is associated with antisocial behavior, and antisocial behavior is associated with having an abnormally small PFC.

It is still scientifically conceivable that these sociopaths are in some sense responsible for the defects in their brains. They were perhaps criminally negli-

gible in not exercising their moral capacities. Their PFCs may have withered away from lack of use, and they should therefore be held responsible for their crimes. This is of course a scientific question; a longitudinal study could determine whether these sociopaths were born with a smaller PFC or whether they failed to develop the capacity they were born with.

But can we blame them for failing to develop this capacity? It seems that even this ability is influenced by factors outside of a person's control. For example, children with a genotype conferring high expression of MAOA, an enzyme that metabolizes neurotransmitters, were less than half as likely to develop a conduct disorder after being severely mistreated in childhood (Caspi et al., 2002). Certainly we cannot blame a person for her genotype. If the genotype strongly predisposes her to a certain action, can we blame her for the action?

It seems likely that as neuroscience and psychology progress, we will be able to explain more of human behavior in terms of environmental and genetic influences. As we grow to understand criminal behavior in terms of factors beyond the criminal's control, how can we continue to make sense of blame and punishment? For some reason, having to do with a mix of environmental and genetic influences, their brains fired in certain ways that led them to do certain things for which they are now incarcerated. Do they deserve their punishment any more than poor Phineas Gage should be blamed for his cursing?

3 Determinism in Physics and Philosophy

Philosophers and physicists have been debating free will and determinism since long before the advent of functional Magnetic Resonance Imaging. In order to understand what new information is added by neuroscientific studies, it will help to understand free will and determinism in a more abstract setting.

3.1 Physics

Let us begin with the premise that we can describe the way that objects move in the world using the laws of physics. These laws may be Newton's laws, or those of general relativity and quantum mechanics, or some unified theory yet to be discovered. In each of these theories, no non-physical entity can influence the action of a physical entity. Physical objects move only due to other physical objects. We do not know of any exception to this rule.

Consider the case of a human who commits a crime. The crime consists of the human body contorting itself in a certain way at a certain time in a context that renders the act so performed illegal. Perhaps the crime also involved a tightening of the vocal chords and concurrent expulsion of carbon dioxide from the lungs. In any case, the crime consists of the motion of physical objects. If we are to believe the laws of physics, then these physical objects can only have been set into motion by other physical objects. These causes are likely to include the contraction of certain muscles, which can be traced back to the firing of neurons in the motor cortex, whose cause seems to be the arrival of certain neurotransmitters at the receptors of the motor neuron, which seem to have come across the synapse from other neurons. So the action is due to the synapse structure in the brain, coupled with the environmental influences whose signals reached the brain through various sensory neurons and set the action in motion. The synapse structure originates, so far as we know, in a complex interaction between genetic and environmental factors that occurred from the time the criminal was conceived until the moment the crime was committed. Whether this person has an excessively large or excessively small PFC seems largely irrelevant to the question of moral blame. In either case, the criminal could not have done otherwise (Glatz, 2008).

At this point, some cunning audience member shouts out, "But I thought

that quantum mechanics was a non-deterministic theory! Perhaps that's where free will comes in!" The first response is that quantum mechanics has only been proved to be locally non-deterministic: events cannot be deterministically predicted based on local variables. There are perfectly good non-local deterministic theories of quantum mechanics — theories that allow for action at a distance — that have not been disproved (Bohm, 1952). But even a non-deterministic theory of quantum mechanics cannot save free will. Quantum mechanical theories are probabilistically deterministic: while the outcome of a given event cannot be predicted perfectly, the probability distribution of many identical events can be predicted extremely well. The idea that free will consists of choosing when to steal and when to refrain, while conforming to a deterministically dictated probability distribution over stealing and not stealing, is not a very satisfying notion of free will.

3.2 Why Free Will is still a Good Idea

If the argument from physics is so clear, then why do we believe in free will at all? The most clear reason is that we all still need to live our lives (Duus-Otterström, 2008). For whatever reason, the processes in our brain endow us with the feeling of consciousness, which leaves us as creatures who feel that we have choices and make decisions.

And from a practical perspective, we might as well. Quantum mechanics is at least nondeterministic enough to prevent any possible experimental device from being able to predict our actions perfectly. In conventional speech, free choice is generally used to mean that we are able to do what we want to do (Fingarette, 2008). In this sense we do indeed exercise free will, even though our wants may themselves be determined. Given environmental inputs, our brains calculate what to do, and it is this calculation that we experience as free

will.

It is unclear in what sense the decisions that come out of these neural processes can be considered blameworthy, but it is clear that the inputs and outputs depend on each other. One such input is, strangely enough, a belief in free will: participants in one study who read a text encouraging a belief in determinism were more likely than control subjects to cheat on a test (Vohs and Schooler, 2008). The rest of this essay is inspired by the insight that the decisions that we call “freely willed” could be well described as decisions for which different inputs might have yielded different results. This insight can help us understand why punishment for a crime should depend on the mental state of the criminal.

4 The Legal Perspective

We have seen that the idea of blameworthiness is philosophically questionable, and that its application is growing progressively more difficult as neuroscientific evidence provides examples of the deterministic (or at least probabilistic) causes of human behavior. Reunderstanding the idea of blame and guilt in light of the philosophical and neuroscientific evidence is a laudable goal, but one that is beyond the scope of this paper. Instead, I concentrate here on the practical consequences these insights should have for our legal system. At what point does the idea of blame enter into the law, and why? Is there a way to reformulate the law to preserve its function without invoking the dubious idea of blame?

An investigation of the grounds for the insanity defense will provide us with a window onto the role of blameworthiness in legal proceedings, and on how evidence that undermines the logic of blame might yet be profitably incorporated into legal thinking.

4.1 The Insanity Defense

The insanity defense entered the legal cannon in the M'Naghten case of 1843, when Daniel M'Naghten killed the secretary of the Prime Minister of England while under the delusion that the Prime Minister himself was at the forefront of a “murderous conspiracy,” but was acquitted on the grounds of insanity (Mobbs et al., 2007). Legal theorists have been struggling to justify the insanity defense ever since. It has been interpreted primarily in two different ways (Goldstein and Katz, 1963). First, it may act as evidence that “precludes establishing a crime by leaving in doubt some material element of an offense;” namely, the *mens rea*, or guilty mind, of the criminal. Second, it may serve as a defense to a crime that may change the sentence invoked, even though all the elements of the crime have been established.

Many justifications of the insanity defense have invoked the idea of guilt and blameworthiness. The Royal Commission on Capital Punishment of 1953 noted that “it has for centuries been recognized that, if a person was, at the time of his unlawful act, mentally so disordered that it would be unreasonable to impute guilt to him, he ought not to be held liable to conviction and punishment under the criminal law” (Goldstein and Katz, 1963). However, they do not address just what makes it “unreasonable” to impute guilt to a mentally disordered person any more than to a mentally ordered person. Likewise, in the 1954 *Durham* case, the court found with equal vacuity that “our collective conscience does not allow punishment where it cannot impose blame” (Goldstein and Katz, 1963). The court goes on to invoke the idea of *mens rea*: an insane person cannot possess the guilty mind that would confer criminal responsibility. This formulation indicates that the court sees insanity as being a piece of evidence that prohibits the establishment of a crime.

4.2 *Mens Rea* and the Insanity Defense

In *US v. Currens* (1961), the court continues to elaborate its understanding of the insanity defense as showing that the criminal act did not occur with *mens rea*, and thus is not a crime. The court understand *mens rea* to rest “on the assumption that a person has a capacity to control his behavior and to choose between alternative courses of conduct” (Goldstein and Katz, 1963). An insane person cannot be charged with a crime “because the totality of his personality is such, because of mental illness, that he has lost the capacity to control his acts in the way that the normal individual can and does control them...he must be found not to possess the guilty mind, the *mens rea*, necessary to constitute his prohibited act a crime” (Goldstein and Katz, 1963).

At this point, Goldstein and Katz point out, the court logically ought to have noted that since an insane defendant lacks the “essential element of *mens rea*, there is no crime from which to relieve the defendant of liability and consequently, since no crime has been committed, there is no need for formulating an insanity defense” (Goldstein and Katz, 1963). This stance would leave neuroscientists in the awkward legal position of trying to determine whether a human being were sufficiently capable of controlling his or her actions to be guilty of the crime in question (Buchanan, 2008). This binary question would grow more difficult to answer as evidence accumulates that measurable genetic and environmental factors do influence a person’s likelihood to commit a crime (and, we infer, ability to refrain from doing so). Consider the children whose pre-frontal cortices were damaged in their infancy. Would we find them criminally responsible or not?

However, the court does not merely merge the insanity defense into the existing *mens rea* doctrine. An insane defendant who is not guilty by reason of insanity does not walk free; rather, “the acquittal which accompanies the

insanity defense means deprivation of liberty for an indefinite term in a ‘mental institution’ ” (Goldstein and Katz, 1963). Thus, while the language used to justify the insanity defense is often couched in moral terms, the insanity defense is in fact used to select for restraint individuals who are nevertheless considered free of guilt. While we often justify legal punishments using the idea of guilt and blame, the insanity defense is a clear sign that guilt and blame are not in fact the only criteria that we use to establish the penalty due to a crime.

5 Towards a Preventative Justice

From a practical perspective, this makes sense. A community wants neither a rational cold-blooded murderer nor a murderer with a damaged or undersized PFC wandering the streets. A community is designed to safeguard the life, health, liberty, and possessions of all its members. It must do so, paradoxically, by depriving certain of its members of their life or liberty. Goldstein and Katz note that “because of the inherent conflict between the values ultimately to be preferred and the sanctions authorized, the criminal law has sought to minimize the consequences of this paradox” (Goldstein and Katz, 1963). This includes establishing penalties for crimes in order to discourage them, and also establishing rules that exempt from punishment those whom punishment could not restrain.

This perspective gives us a new understanding of the *mens rea* defense. For example, suppose a man brought a peanut butter sandwich to work for lunch, only to find, when he returned from the bathroom, that the fumes had killed his cubicle-mate — a woman whose peanut allergy had plagued her since childhood. This man should not be punished not only because he bears no “guilt,” but because punishing him would serve no purpose for the community. Imprisoning this man is unlikely to have any effect on future peanut murders.

The idea of preventative justice could allow us to evade the difficult question of guilt and free will entirely. If imposing and enforcing a certain penalty for a certain crime will increase societal bliss by decreasing the incidence of that kind of crime more than it decreases societal bliss by punishing society members, then it makes sense for society to legally mandate that penalty. This maxim seems to be a reasonable way to promote the values for which society exists, and we also find legal precedent for its application in the insanity defense. In the next section we explore some possible consequences of its application.

6 Brain-o-typing

The example of the insanity defense points to a way that scientific evidence might be used to in the creation of penalties (Aharoni et al., 2008). The insanity defense effectively sets two different penalties for the same crime depending on whether a person is sane or insane. The rationale for this differential sentencing is clear: the harm to society due to rational murderers can be best prevented by locking them up until they show that they are harmless and ready for parole (if ever); the harm to society due to insane murderers can be best prevented by restraining them in a mental institution until they are well (if ever). Perhaps this logic should be extended from two types of brains (normal and insane) to more finely tuned classes of people. Trials might first determine whether a crime had been committed, and then determine the penalty based on the brain-o-type of the criminal.

We already have evidence for a way this kind of sentencing might work. Raine et al looked at functional differences between impulsive affective murderers and premeditated psychopaths. Impulsive murderers showed less activation in the PFC, while psychopaths showed normal function in the PFC but increased activation in right subcortical areas such as the amygdala and hip-

pocampus (Raine et al., 1998). Since the PFC has been linked very strongly with executive control and the regulation of impulses, one could interpret these results as indicating that the predatory psychopaths think about their decision to kill, while the impulsive murderers have murdered before they stop to realize what they've done. In the language of free will, the predatory psychopath is able to regulate his impulses, whereas the impulsive murderer cannot (Mobbs et al., 2007). Perhaps, then, it would make sense to mandate life imprisonment or the death penalty for psychopaths, while recommending shorter sentences coupled with anger management training for impulsive murderers. Thus psychopaths, who think about their actions before committing them, might be deterred from their crimes by the larger penalty, while the impulsive murderers, who do not think before they murder and therefore cannot be deterred by the idea of a penalty, might learn to control their impulses and go on to become productive members of society.

The day when this type of legal sentencing makes sense is far in the future. Before we should trust our lives to a brain image, we should make sure that the brain image will do what it claims to. That is, neuroscience would have to understand much better than it does today how to correctly infer mental processes from brain activation patterns, and how to understand the mental capacity of an individual rather than the characteristics of the group of which the individual is a part (Nichols and Newsome, 1999; Van Horn and Poldrack, 2008).

6.1 A Thought Experiment

But suppose we have surpassed these scientific difficulties. Imagine a world in which neuroscientists have divided humanity into ten brain types (the continuum of brains having been divided up using some kind of clustering analysis).

Every few years your brain is scanned in order to determine your brain-o-type. Scientists have conducted experiments to determine the sentences that will best prevent people of each brain-o-type from committing a given kind of crime. These sentences are ideally tuned to deter each type of person from committing crimes and to prevent those who do from committing more. The sentences for murder range from execution, imprisonment, and institutionalization to banishment and public humiliation. Neuroscientists are called into courtrooms not to determine whether the defendant is sane or insane, but whether she is type 4 or type 7.

Do you want this to be your world?

There are grave problems with this state of affairs. Coupling the thought experiment in the language of science obscures the immense potential for abuse of the system. It is highly conceivable that brain-o-types might predict things other than response to punishment. For example, the vast majority of politicians might turn out to be type 5, and might pass laws imposing low penalties for crimes by individuals with type 5 brains.

But there is something more deeply troubling about this world. It poses some basic affront to our idea of fairness. Is it that in this world, individuals are treated according to statistical data on groups? Is it that individuals receive different punishments for the same crime? Both of these problems already occur in the case of the insanity defense. Is it that we don't believe that neuroimaging data will ever be reliable enough to trust in this grave a situation?

Or is it because we cannot believe that we, like the legally insane, are no more than our brains?

7 Conclusion

New evidence from neuroscience is rapidly undermining the retributive thinking that most people use to understand why criminals ought to be punished. If people commit crimes because certain neurons in their brain fired while others did not, and if those neurons fired because of a complex combination of genetic and environmental influences, then it is unclear in what sense we can say that anyone is exercising free will. However, the conclusion of this paper is that there still is a reason to punish criminals. We can shape behavior by changing the environmental influences impinging on the brains — a brain that might decide to commit murder under one legal regime might refrain in another. Neuroscientific evidence might be able to suggest better ways to design penalties so as to minimize the suffering of the community due to crime and of the criminals due to legal punishments. But mandating different penalties for different kinds of brains — or acknowledging that we have already begun to do so, in the usage of the insanity defense — violates our intuitive notion of fairness, and is dangerously susceptible to abuse.

Should we continue to punish based on notions of guilt that will become progressively less scientifically justifiable? Or can we find an alternative grounding for our legal system, such as preventative justice?

How comforting, when faced with decisions like these, to remember that we've got no free will anyway.

References

- J. Harlow, Passage of an iron bar through the head, Boston Medical Surgery Journal 13 (1848) 389–393.

- J. Grafman, K. Schwab, D. Warden, A. Pridgen, H. R. Brown, A. M. Salazar, Frontal lobe injuries, violence, and aggression: A report of the Vietnam head injury study, *Neurology* 46 (5) (1996) 1231–1238.
- S. W. Anderson, A. Bechara, H. Damasio, D. Tranel, A. R. Damasio, Impairment of social and moral behavior related to early damage in human prefrontal cortex, *Nature neuroscience* 2 (11) (1999) 1032–1037.
- A. Raine, T. Lencz, S. Bahrle, L. LaCasse, P. Colletti, Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder, *Archives of General Psychiatry* 57 (2) (2000) 119–127.
- A. Caspi, J. McCray, T. E. Moffitt, J. Mill, J. Martin, I. W. Craig, A. Taylor, R. Poulton, Role of genotype in the cycle of violence in maltreated children, *Science* 297 (5582) (2002) 851–854.
- R. M. Glatz, The (near) necessity of alternate possibilities for moral responsibility, *Philosophical Studies* 139 (2) (2008) 257–272.
- D. Bohm, A suggested interpretation of the quantum theory in terms of "hidden" variables., *Physical Review* 85 (2) (1952) 166–179.
- G. Duus-Otterström, Betting against hard determinism, *Res Publica* 14 (3) (2008) 219–235.
- H. Fingarette, Free choice, *Current Psychology* 27 (1) (2008) 1–5.
- K. D. Vohs, J. W. Schooler, The value of believing in free will: Encouraging a belief in determinism increases cheating, *Psychological Science* 19 (1) (2008) 49–54.
- D. Mobbs, H. C. Lau, O. D. Jones, C. D. Frith, Law, responsibility, and the brain., *PLoS biology* 5 (4).

J. Goldstein, J. Katz, Why an "Insanity Defense", *Daedalus* 92 (3) (1963) 549–563, ISSN 00115266.

A. Buchanan, Commentary: Freedom and function, *Journal of the American Academy of Psychiatry and the Law* 36 (1) (2008) 25–26.

E. Aharoni, C. Funk, W. Sinnott-Armstrong, M. Gazzaniga, Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience, vol. 1124, 2008.

A. Raine, J. Meloy, S. Bahrle, J. Stoddard, L. e. a. LaCasse, Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affective murderers, *Behavioral Science and Law* 16 (1998) 319–322.

M. J. Nichols, W. T. Newsome, The neurobiology of cognition, *Nature* 402 (6761 SUPPL. 1) (1999) C35–C38.

J. D. Van Horn, R. A. Poldrack, Functional MRI at the crossroads, *International Journal of Psychophysiology* .