Journal of Law and the Biosciences, 120–139
doi:10.1093/jlb/lsv051
Advance Access Publication 30 December 2015
Original Article

What does recent neuroscience tell us about criminal responsibility?

Uri Maoz[1,*,†] and Gideon Yaffe[2,‡]

1. Adjunct Assistant Professor of Psychology, University of California Los Angeles (UCLA); and Visiting Researcher in Neuroscience, California Institute of Technology (Caltech)
2. Professor of Law and Professor of Philosophy & Psychology, Yale Law School
*Corresponding author. E-mail: urimaoz@ucla.edu

ABSTRACT

A defendant is criminally responsible for his action only if he is shown to have engaged in a guilty act—*actus reus* (eg for larceny, voluntarily taking someone else's property without permission)—while possessing a guilty mind—*mens rea* (eg knowing that he had taken someone else's property without permission, intending not to return it)—and lacking affirmative defenses (eg the insanity defense or self-defense). We therefore first review neuroscientific studies that bear on the nature of voluntary action, and so could, potentially, tell us something of importance about the *actus reus* of crimes. Then we look at studies of intention, perception of risk, and other mental states that matter to the *mens rea* of crimes. And, last, we discuss

† Uri Maoz is a an adjunct assistant professor at UCLA's Department of Psychology, a Visiting Researcher for Neuroscience at Caltech, and a Bial Fellow. He teaches psychology and neuroscience at UCLA. He is a computational neuroscientist, specializing in statistical modeling of neural data, especially online and in real time. He has published in the fields of decision-making, neuro-law, motor control, and consciousness studies, among others, and has developed data-driven methods for rigorous analysis of neural data. As part of his interest in the influence of neuroscience on society, and especially moral and criminal responsibility, he has been collaborating with several law scholars and philosophers. This is in an effort to hone in on questions that are both of interest to the law community and empirically tractable using the tools of psychology and neuroscience.

‡ Gideon Yaffe is a professor at Yale Law School, where he teaches criminal law and the philosophy of law. He also holds appointments in Yale's philosophy and psychology departments, and is a Guggenheim Fellow. In recent years, he has been writing primarily about criminal responsibility and how best to understand it in light of philosophical thought about agency, freedom and moral responsibility, and in light of empirical discoveries in psychology and neuroscience. While his PhD is in philosophy (from Stanford), fellowships of various sorts while working as a philosophy professor helped him to receive some legal training (from USC) and some training in neuroscience (from Caltech). Currently, he collaborates with several neuroscientists on experiments aiming to be of relevance to the criminal law, and particularly to the criminal law's efforts to sort people on the basis of their different degrees of responsibility for wrongful, and often very harmful, behavior.

studies of self-control, which might be relevant to some formulations of the insanity defense. As we show, to date, very little is known about the brain that is of significance for understanding criminal responsibility. But there is no reason to think that neuroscience cannot provide evidence that will challenge our understanding of criminal responsibility.

KEYWORDS: Intention and perception of risk, neuroscience and criminal responsibility, neuroscience and law, self-control, voluntary action

I. INTRODUCTION

Shortly after turning 40 years old, Michael developed an interest in child pornography. This was surprising because he had had no interest in such materials up to that point in his life. He also found himself attracted to his 12-year-old stepdaughter, with whom he and his wife lived. One night, while putting his stepdaughter to bed, he fondled her. He was convicted of a crime of sexual molestation of a child. The judge gave Michael the option of avoiding jail time should he pass successfully through a treatment program. Anxious to avoid jail, Michael agreed. He did very poorly in the program. Among other things, he came on to the staff and other patients. Michael also complained of headaches, and a neurologist, working with the patients in the treatment program, recommended a structural magnetic-resonance imaging (MRI) of Michael's brain. The MRI revealed a large orbitofrontal tumor. Surgery was performed to remove the tumor, and, after recovery, Michael no longer had sexual urges directed towards children. He returned to the treatment program and passed without difficulty, meeting the judge's demands. He returned home and seemed to be having no problems, beyond those that might be expected of someone who had undergone brain surgery. Months later, however, his sexual desires for children returned. Another MRI revealed that the tumor too had returned. The tumor was again removed, and, again, Michael's urges went away.[1]

In one important sense, Michael's case is so rare as to be, perhaps, unique. It is not unheard of, although rare, for impulses to criminal behavior to be traceable to brain abnormalities. But Michael's case may be the only one to date in which the relevant brain abnormality was also correctable through medical intervention. However, in another important sense, Michael's case is just like *every* case of criminal behavior. Criminal behavior is a product of the interaction between the states and functional dispositions of the brain, on the one hand, and the environment, on the other. If we knew enough about the brain, and if our medical technology was sophisticated enough, wouldn't every case, in every courtroom, be just like Michael's? Wouldn't it always be the case that we could eliminate the impulse that gave rise to the crime through a medical procedure?[2]

Consider Michael's case in relationship to a study conducted by Kiehl and colleagues, which investigated neural activity predictive of recidivism in criminals. Ninety-six inmates, who were about to be released from jail, were instructed to press a button as quickly as possible whenever an X appeared on the screen (84 per cent occurrence probability), but not when a K was displayed (16 per cent occurrence probability).

---

[1] Jeffrey M. Burns & Russell H. Swerdlow, *Right Orbitofrontal Tumor With Pedophilia Symptom and Constructional Apraxia Sign*, 60 ARCH. NEUROL. 437, 440 (2003).

[2] The medical procedures and technologies required to eliminate such impulses might be much more complex than tumor removal, of course, and well outside the scope of our current knowledge.

Using fMRI, the researchers tracked activity in the anterior-cingulate cortex (ACC)—a limbic region associated with impulse control. They found that the odds that an offender with low ACC activity would be arrested within 4 years of release were about double those of an offender with high ACC activity. ACC activity was also a better recidivism predictor than other factors, like age at release, substance abuse, task error rate, and psychopathy score. Hence, the authors suggested ACC hemodynamic activity as a potential neurocognitive biomarker of persistent antisocial behavior, at least at the group level.[3]

A first and essential step towards appreciating the bearing of neuroscientific results on criminal responsibility is recognizing the sense in which Michael's criminal behavior, and the criminal behavior of the inmates who recidivated in the study just described, is no different from anyone else's: such behavior has its source in the person's brain and his environment. There are two natural ways to respond to this: give up on criminal responsibility all together, or, alternatively, accept that being fully responsible for bad behavior is consistent with that behavior being explicable in principle through appeal to facts about one's brain and the environment in which one finds oneself. Those drawn to the former view may be ready to stop reading this article now. If nobody is criminally responsible for his behavior, then neither neuroscience nor any other science can help us to understand criminally responsible behavior better; there is none. Those drawn to the latter view, however, can seek ways in which neuroscience can, or does, inform our understanding of the features of human beings in virtue of which they are criminally responsible for their behavior. This is the approach taken in this article. We here survey some recent neuroscientific studies that might be thought to shed light on those facts about people in virtue of which they are, and are held by our legal system to be, criminally responsible for their behavior.

Under what conditions is a person criminally responsible for his behavior? Understanding the answer given by our legal system sufficiently for our purposes here requires appreciating some legally defined concepts: *actus reus*, *mens rea*, and affirmative defenses. Criminal statutes define crimes. These definitions are taken to express a series of conditions that must be met, and must be shown to have been met by the prosecution, for guilt. Under typical larceny statutes, for instance, a defendant is guilty only if he (a) took something (b) that belonged to someone else (c) without permission, and while (i) knowing that he was taking it, (ii) being aware of a substantial risk that it belonged to someone else, (iii) being aware of a substantial risk that he lacked permission to take it, and (iv) intending never to return it. Roughly speaking, the non-mental facts that must be shown to establish that a defendant meets a statutory definition of a crime, such as (a), (b), and (c), are the crime's '*actus reus*'. The mental facts, such as (i), (ii), (iii), and (iv), are the crime's '*mens rea*'. These are only rough definitions of '*actus reus*' and '*mens rea*', however, because some elements traditionally classified as *actus reus* include facts about the defendant's mind. Establishing that a person engaged, for instance, in the act of taking something, requires proof that his bodily movements were guided by his mind—a person who is holding a loaf of bread when someone throws him out of the door of the bakery has not 'taken' the loaf. However, facts about the person's mind do not suffice. We also need to know that he had control of the object and moved it out

3 Eyal Aharoni et al., *Neuroprediction of Future Rearrest*, 110 Proc. Natl. Acad. Sci. 6223, 6228 (2013).

of the control of its rightful owner. So, more precisely, conditions that must be met for guilt and are *not exclusively* facts about the defendant's mind constitute the *actus reus* of the crime. By contrast, conditions (i), (ii), (iii), and (iv) are facts entirely about the defendant's mind; they constitute the *mens rea* of the crime.

Engaging in a crime's *actus reus* with its *mens rea* is necessary but not sufficient for criminal responsibility. Additionally, a defendant must lack an affirmative defense. For instance, someone who is suffering from a severe mental illness that includes deific hallucinations may delusionally believe that he must steal his neighbor's car in order to comply with God's demands. When he does so, he may very well meet all of the conditions required for commission of larceny. But he is not criminally responsible for his behavior. The insanity defense is one example of an affirmative defense that might be available in such a case. The affirmative defenses also include self-defense, defense of others, and duress, among others. Whenever a defendant has an affirmative defense, he is not criminally responsible, and so is not subject to punishment in the form of incarceration, fines, or execution, nor is he subject to other forms of related state supervision, such as probation. In the case of the insanity defense, however, he may be subjected to involuntary commitment in a mental institution, or be subject to other forms of medical intervention, through, for instance, outpatient programs, against his will. In general, any condition that is necessary for criminal responsibility is necessary also for punishment by the state, but in the absence of criminal responsibility, the state might still exercise control over the agent, just not in the form of punishment.

Potentially, a neuroscientific result could show that a class of people, or perhaps, even, an individual person, fails to meet, or succeeds in meeting, one of the necessary conditions of criminal responsibility. If it could be shown, for instance, that people with orbitofrontal tumors like Michael's typically meet the law's criteria for insanity, then such research would provide some support for the claim that Michael was not criminally responsible for his behavior by supporting the claim that he had an insanity defense. We would not know that for sure; Michael himself would need to be examined. But still, such a result would provide some support for the claim that Michael is not criminally responsible. Notice that such a result would not extend to everyone who engages in criminal behavior. Some states of the brain—eg low ACC activity—that (when in the right environment) give rise to criminal behavior might do so without it being the case that the agent meets the law's criterion for insanity. Sanity, like insanity, depends on capacities that brains have. And criminal behavior by the sane, like such behavior on the part of the insane, has its source in the brain. But, still, if the neural sources of insanity can be identified, that would be of potential use to the legal system. For example, a better understanding of the neural basis of capacities underlying criminal responsibility might shed new light on the standards that we apply for individuals having or lacking criminal responsibility.

In addition, neuroscientific studies might illuminate the neural mechanisms that underlie those features of people in virtue of which they are criminally responsible for their behavior. And so they would help us to understand criminal responsibility better, without thereby supporting an argument for or against holding any person or class of people criminally responsible. While knowing more about a problem can be a first step to solving it, knowing more might be valuable simply because it involves *knowing more*. In the same way that the knowledge that the neuroscience of memory provides is of value even before we make use of it to treat memory disorders, or to improve our memories, if neuroscience can help us to understand the neural nature of criminally responsible behavior, that would be of value, even if we cannot use such knowledge to reduce crime or increase justice. Such results would add to human knowledge not just of the brain but of one of the most socially important phenomena to which the brain gives rise: crimes for which people are responsible and deserving of punishment.

Section I.1 of this article discusses neuroscientific studies that bear on the nature of voluntary action, and so could, potentially, tell us something of importance about the *actus reus* of crimes. Section I.2 discusses studies of intention, perception of risk, and other mental states that matter to the *mens rea* of crimes. And Section I.3 discusses studies of self-control. The topic of self-control bears a less direct relationship to criminal responsibility than one might think, for reasons that are explained. However, something like self-control is of relevance to some formulations of the insanity defense, and to other issues important to the assessment of criminal responsibility. As we will see in all three sections, the work that has been done to date is just the smallest drop in the bucket. So far, very little is known about the brain that is of significance for understanding criminal responsibility.

I.1. Voluntary action

It seems abhorrent to hold people criminally responsible only for their thoughts. And it seems abhorrent to hold people criminally responsible for their status. It would be wrong, that is, to punish people for intending to do things that they take no steps towards doing—assuming, as seems plausible, that to intend something is merely to have a certain kind of thought about it; and it would be wrong to punish people for, for instance, being poor, or unemployed, or belonging to a particular race or class. These base level moral intuitions are accommodated in the law in part through a restriction on the *actus reus* of crimes: the *actus reus* must include a voluntary act. There are some exceptions to this—notably, people are also criminally responsible for omitting actions they have a duty to perform—but, as a default, criminal responsibility requires a voluntary act.

When definitions of voluntary action have been required for legal purposes, as when it is less than obvious whether the defendant's behavior amounted to voluntary action, the law has employed something like the following definition: a voluntary act is a bodily movement guided by a conscious mental representation of that bodily movement. If a hurricane's wind blows someone through a shop window, shattering it, there is no guilt for destruction of property since the bodily motion that caused property to be destroyed was not guided by the agent's mental activity. The bodily movements during an epileptic seizure, also, do not rise to the legal standard of a voluntary act since the brain activity that gives rise to them does not guide them in the way, for instance, in which a desire to move one's hand guides one's hand. Consider a more problematic case. In *People v. Newton*[4], the defendant was shot in the gut during an altercation with police. Moments later, he shot and killed one of the officers before wandering several blocks to a hospital where he collapsed, unconscious. On waking later, he reported no memory of the incident. A doctor testified that it is common for people who have

4 People v. Newton (8 Cal. App. 3d 359 (Ct. App. 1970))

just suffered severely traumatic injury, as Newton had, to engage in complex, seemingly goal-directed bodily movements while unconscious. Eventually, Newton was acquitted on the grounds that the bodily movements in which he engaged when shooting the officer dead were not voluntary acts since the mental representations that guided them were not conscious.[5]

The nature of voluntary bodily movements is amenable to investigation using neuroscientific tools. And, in fact, a substantial amount of neuroscientific work has been done in this area. Some of this work might be thought to suggest that the law's definition of voluntary action is deeply flawed. As we will see, to reach such a conclusion on the basis of the work done to date, would be extremely rash.

In a seminal experiment, Benjamin Libet and colleagues asked subjects to move their right hand at a time of their choice while some of their brain activity was recorded using electroencephalography. The subjects were seated in front of a rapidly rotating spot on a clock and were instructed to recall, after they moved, its position *when they first felt the urge to move*. On average they reported this urge about 200 ms before movement onset. Yet, the experimenters found that a negative electrical potential above motor-related areas of the cerebral cortex—termed a 'readiness potential' and known to precede voluntary action—began about 500 ms before movement onset.[6] More recently, it was shown that when subjects are instructed to press a button with the left hand or another with the right, some information about which hand they would select could be deciphered from functional magnetic-resonance imaging (fMRI) of frontopolar cortex already 10 s or so before movement onset, well before subjects reported having made up their minds (Fig. 1).[7] The authors in these studies therefore concluded that it was *unconscious* brain activity that initiated the action.

These experiments and their interpretation have come under much criticism, both methodological and conceptual[8]. For instance, there is evidence that the timing of the reported urge to move is systematically biased[9] , retrospectively inferred from the movement time rather than independently constructed.[10] This casts doubt on the validity of subjective reports of the time of decision, and so challenges the original Libet studies, where the readiness potential precedes the reported decision time by about

5 The case was racially charged. Newton was the head of the Black Panthers and had made explicitly threatening remarks about the police on previous occasions. It is unlikely that Newton was stopped randomly. But the explicit legal grounds on which the case was decided were not racial, even if the implicit grounds were.

6 Benjamin Libet et al., *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential) the Unconscious Initiation of a Freely Voluntary Act*, 106 BRAIN 623, 642 (1983).

7 Chun S. Soon et al., *Unconscious Determinants of Free Decisions in the Human Brain*, 11 NAT. NEUROSCI. 543, 545 (2008).

8 Benjamin Libet, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, 8 BEHAV. BRAIN SCI. 558, 566 (1985); Jeff Miller, Peter Shepherdson & Judy Trevena, *Effects of Clock Monitoring on Electroencephalographic Activity Is Unconscious Movement Initiation an Artifact of the Clock?*, 22 PSYCHOL. SCI. 103, 110 (2011); WALTER-SINNOTT ARMSTRONG & LYNN NADEL (eds), CONSCIOUS WILL AND RESPONSIBILITY: A TRIBUTE TO BENJAMIN LIBET (2011); ADINA L. ROSKIES, *How Does Neuroscience Affect Our Conception of Volition?*, 33 ANNU. REV. NEUROSCI. 109, 130 (2010); Aaron Schurger, Jacoba D. Sitt & Stanislas Dehaene, *An Accumulator Model for Spontaneous Neural Activity Prior to Self-Initiated Movement*, 109 PROC. NATL. ACAD. SCI. E2904, E2913 (2012).

9 Uri Maoz et al., *On Reporting the Onset of the Intention to Move*, in SURROUNDING FREE WILL: PHILOSOPHY, PSYCHOLOGY, NEUROSCIENCE 184, 202 (Alfred R. Mele ed., 2015).

10 William P. Banks & Eve A. Isham, *We Infer Rather Than Perceive the Moment We Decided to Act*, 20 PSYCHOL. SCI. 17, 21 (2009).

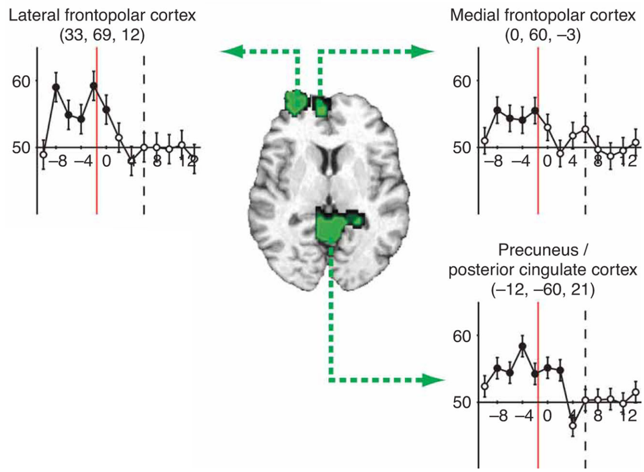

Figure 1. Decoding outcome of decisions over time before they were reported to reach awareness. The three gray patches designate regions where the specific outcome of a motor decision could be decoded before it had been made (mean ± s.e.m.; filled circles indicate significant decoding accuracy at $p < 0.05$). The vertical red line shows the earliest time at which the subjects became aware of their choices. The dashed (right) vertical line in each graph shows the onset of the next trial. (**Adapted with permission from Soon et al., 2008.**)

300 ms, on average. It suggests that this gap between readiness-potential onset and the reported time of the urge might be an artifact of the process through which the time of the urge is inferred from the time of the movement rather than indicative of a gap in time between the two events. It has also been shown that when monkeys are offered one of several queries at random, frontal and striatal brain activity *before the query is presented* may predict the monkeys' eventual resolution of it, especially when the monkeys must decide between decision options of similar values.[11] As randomly pressing a button with the left or right hand with no purpose or consequence is a prototypical situation where the values of the two decision options are very similar, the early fMRI prediction signals may reflect bias activity rather than an early formation of a decision[12]. Further, and importantly, when our concern is criminal responsibility we are interested in deliberate decisions rather than the purposeless, insignificant and unreasoned ones investigated in the Libet experiments. And so the Libet results do not generalize well to decisions of the kind that bring people into courtrooms.[13]

11 Uri Maoz et al., *Predeliberation Activity in Prefrontal Cortex and Striatum and the Prediction of Subsequent Value Judgment*, 7 FRONT. NEUROSCI. 225 (2013).

12 John-Dylan Haynes, *Decoding and Predicting Intentions*, 1224 ANN. N. Y. ACAD. SCI. 9, 21 (2011).

13 See Bruno G. Breitmeyer, *Problems With the Psychophysics of Intention*, 8 BEHAV. BRAIN SCI. 539, 540 (1985); Adina L. Roskies, *How Does Neuroscience Affect our Conception of Volition*?, 33 ANNU. NEUROSCI. 109, 130 (2010); for example, for the generalization of the Libet results to deliberate decisions. Deliberate decisions have begun to be studied in this context, Uri Maoz et al., *Predicting Action Content On-Line and in Real Time*

But even if taken at face value, to conclude from such work that the law's definition of voluntary action is flawed would be very hasty. First, some might be drawn to the idea that the experiments' evidence of the predictability of decisions undermines the voluntariness of subsequent behavior. But the law does not require that an act be unpredictable in order to be voluntary. If a resolute person announces that he is going to rob a bank tomorrow, we might be able to predict that he will do so with great reliability, but we do not therefore conclude that his conduct is involuntary when he does as he promised. We might also be able to predict quite reliably how a member of the Klu Klux Klan will act towards a Black person. And we saw that ACC activity can to some extent predict recidivism. But we would not conclude that the predicted behaviors in such cases are, because predictable, involuntary.

Second, others may be struck by the experiments' evidence that the agent carrying out the action does not consciously recognize the psychological states accompanying the neural activity that correlates with action as the source of subsequent bodily movement. However, it is far from clear that the mental activity that is the immediate cause of bodily movement has been shown by such studies to be 'unconscious' *in the sense that matters for the law*. Someone, for instance, who, in a rage, punches someone else in the face, may not have been attending to his decision to do so—perhaps he was attending only to the source of his anger—but it does not follow that his decision was 'unconscious' in a sense that would disqualify his bodily movements from being voluntary in the legal sense.

And, third, it may appear to some that the cause of subjects' actions in those experiments was that early brain activity rather than conscious volition. But these experiments provide no real evidence that conscious volition does not cause action. It is therefore possible that the bodily movements of the subjects were both predictable from early brain activity *and, also,* caused by their conscious volitions. But since causation by conscious volition is what is required for voluntary action under the law, it follows that it is possible that the subjects acted voluntarily in the legal sense, despite the fact that early brain activity predicted what they would do.[14]

Another set of experiments studied the awareness that people have of initiating, executing and controlling volitional action—termed 'the sense of agency'. While the law's definition of voluntary action does not appeal to the sense of agency, it is probably also true that the intuitive appeal of the law's conception of voluntary action derives from the fact that willed bodily movements are, at least typically, accompanied by the sense of agency. So, if there were reasons to believe that the sense of agency is an unreliable guide to the voluntariness of the behavior it accompanies, then that would give us some ground for concern about the law's definition of voluntary action.

before Action Onset–an Intracranial Human Study, in 25 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 881, 889, and there is preliminary evidence that deliberate and random decisions may utilize different brain processes; Liad Mudrik et al., *Dissecting Different Types of Decision-Making: An ERP Study of Reasoned vs. Unreasoned voluntary decisions*, 43 SOC'Y NEUROSCI (2013).

14 For further discussion, see the essays collected in WALTER-SINNOTT ARMSTRONG & LYNN NADEL (eds), CONSCIOUS WILL AND RESPONSIBILITY: A TRIBUTE TO BENJAMIN LIBET (2011); Adina L. Roskies, *How Does Neuroscience Affect Our Conception of Volition*?, 33 ANNU. REV. NEUROSCI. 109, 130 (2010); Stephen J. Morse & William T. Newsome, *Criminal Responsibility, Criminal Competence, and the Prediction of Criminal Behavior*, PRIMER CRIMIN. LAW NEUROSCI. 150, 178 (2013); and Stephen Morse, *Lost in Translation?: An Essay on Law and Neuroscience*, 13 LAW NEUROSCI., CURR. LEGAL ISSUES 529 (2011).

Experiments on the sense of agency relied on patients in need of brain surgery (eg for tumor removal, or for treatment of intractable epilepsy), who often undergo invasive brain mapping to minimize the impact of the surgery on their later everyday life. This includes electrical stimulation of various brain regions to localize their involvement in functionality (eg by eliciting hand movement or arresting speech). In one study, patients who underwent such stimulation of the supplementary motor area sometimes reported an irrepressible urge to perform an action, or reported an anticipation that action is about to occur, which was often followed by overt motor activity at higher stimulation intensities. Some patients reported a subjective experience of movement accompanied by no overt motor activity.[15] In another study, careful chronicling of stimulation locations suggested that stimulating the right inferior-parietal regions triggers a strong endogenous intention and desire for contralateral movement, while left inferior-parietal stimulation provokes an intention to move the lips and to talk. At higher stimulation intensities, participants reported having performed these movements, though no movement occurred. In contrast, premotor-region stimulation triggered overt contralateral limb and mouth movements, while patients firmly denied having moved (Fig. 2).[16] This suggests that the sense of agency may not be as strongly coupled with voluntary movement as humans generally experience them to be, and that—at least under rather abnormal circumstances—humans may experience agency over phantom actions and carry out actions with no accompanying sense of agency. It appears, that is, that the brain circuitry involved in the production of bodily movement may not be as closely related to the circuitry involved in the sensation that one has planned, initiated or even engaged in bodily movement as we might have thought.

Note that in the above experiments the sense of agency and action were never aroused simultaneously. But imagine that a patient both carried out an act and thought he was acting voluntarily due to artificial brain stimulation. Even if that were the case, direct brain stimulation of an exposed brain is a highly unnatural condition. And it is not clear to what extent such results could be generalized to more everyday decision-making situations. The crucial question for our purposes here is this: do these experiments show that bodily movements are not guided by conscious mental activity, as required by the law's voluntary act requirement? It is far from clear. At worst, a brain stimulation experiment could be developed in which subjects are shown to be under a certain kind of illusion: they think that they are acting voluntarily when they are not. But the possibility of illusion in one case does not establish its ubiquity in all cases. Your eyes can deceive you (think of optical illusions, for example), but that does not mean that they always do. The existence of optical illusions does not render all vision illusory. Further, it is not clear whether guiding one's bodily movements through a conscious mental state necessarily results in a sense of agency, or whether the sense of agency is either necessary or sufficient for consciousness of the kind that the law takes to be involved in voluntary action. These non-trivial questions might be answerable, at least in part, using empirical means. But without the answers, there is no reason to take such experiments to establish the inadequacy of the law's definition of a voluntary act.

[15] Itzhak Fried et al., *Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation*, 11 J. NEUROSCI. 3656, 3666 (1991).

[16] Michel Desmurget et al., *Movement Intention after Parietal Cortex Stimulation in Humans*. 324 SCIENCE 811, 813 (2009).

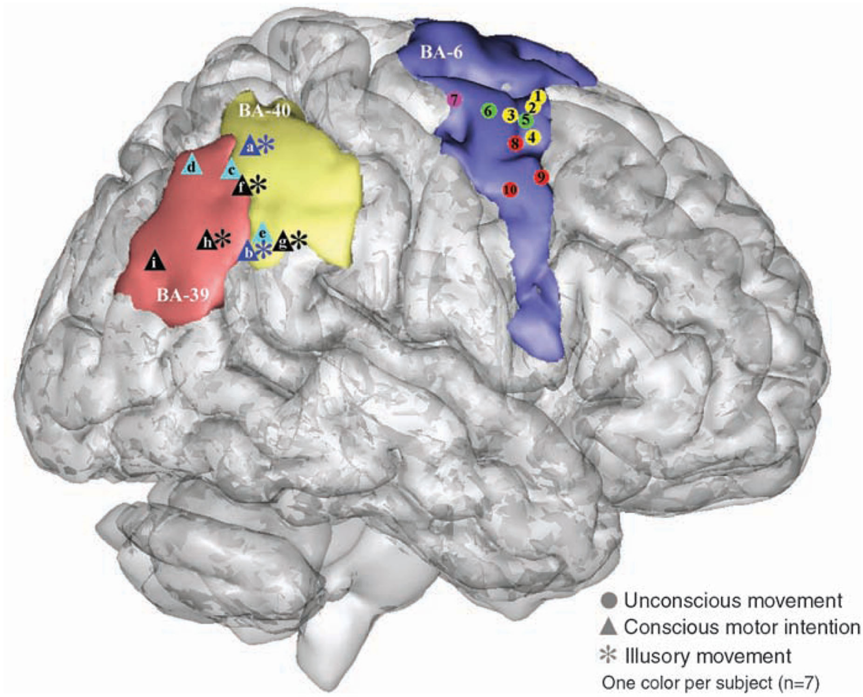

Figure 2. Premotor and parietal responsive sites shown after registration of the individual MR images to the MNI template. Left stimulations have been reported on the right hemisphere. Colored areas define the anatomical boundaries of Brodmann Area (BA) 40, BA 39, and BA 6. (**Adapted with permission from Desmurget et al., 2009.**)

I.2. Intention and the perception of risk

Imagine that, as you are passing by, someone opens the trunk of his car revealing a stack of new iPads still in their shrink-wrapped boxes. He offers to sell one to you for $200—far below the typical retail price. You ask no questions and jump at the chance. Later, you are arrested for the crime of receipt of stolen property. 'But I didn't *know* it was stolen', you say. This might be true. Perhaps you are just very naïve and so believed that the iPad was not stolen. Or perhaps you were not sure one way or the other, although you recognized that there was a good chance that it was stolen. Whether you are guilty of the crime depends on the exact *mens rea* standard set by the statute defining the crime. Does the statute prohibiting the crime require, for guilt, knowledge that the property is stolen? Or will awareness of a large enough risk of that suffice? Or is it enough that a reasonable (and, so, not naïve) person in your situation would have known, even if you did not?

The law recognizes that there are a variety of mental states that a person can be in with respect to a particular fact, such as the fact that the property is stolen. The law also recognizes that people vary in their culpability in accordance with these variations in mental state. These are viewed as different kinds of *mens rea*. For our purposes here, the most important types of *mens rea* are *intent* and *recklessness*. The *Model Penal Code*, which has had tremendous influence on the law in the United States, defines 'intent' or 'purpose' as having a particular act or causal result of that act as one's 'conscious object'.[17] This language is aimed at capturing the idea that those who intend acts or results *aim* at them, or make them their goal. By contrast, the *Model Penal Code* defines 'recklessness' as 'awareness of a substantial and unjustifiable risk' that one is acting in a certain way, or that one will bring about a particular result through one's act.[18] As is clear from the example of receipt of stolen property, it is possible to be reckless with respect to a particular fact—in that example, the fact that the property is stolen—without intending to bring it about. We can assume, in that example, that you did not intend that the property you were receiving would be stolen; you would have been perfectly happy to receive an iPad for $200, stolen or not. But you still may have been aware of a substantial and unjustifiable risk that the property was stolen.

In this respect, some neuroscientific work purporting to be investigating the neural mechanisms underlying intention are investigating a psychological state that is importantly different from the one to which the law gives that label. For instance, in a series of experiments using single-cell recordings and direct brain stimulation, Shadlen and colleagues had monkeys discriminate the direction of motion of moving dots, some percentage of which moved coherently in one direction while the rest jittered randomly.[19] When a large percentage of dots move coherently toward a certain direction, while the rest jitter back and forth randomly, discriminating the direction of motion is easy. In contrast, if just a few percent of the dots move coherently, the task is more difficult. They suggested that neurons found in the middle temporal (MT) areas code the noisy evidence used for the decision, while neurons involved in integrating multiple pieces of information in order to reach a decision are located at the lateral intraparietal area (LIP).[20] This LIP process seems to be involved in the integration of evidence until it reaches a neuronal firing rate threshold associated with one of the decision alternatives, which then leads to action.

While important for elucidating the neurophysiological process that accompanies perceptual judgments, it is not clear to what extent these results can be thought to illuminate the neural substrates of intent as defined in the *Model Penal Code*. The buildup of neuronal firing rate towards a threshold possibly reflects the accumulation of evidence towards a finalized *perceptual judgment,* rather than an intention. Once the neural firing rate threshold is reached, the monkey recognizes that the dots are moving in one direction or another, and thus the direction toward which saccading would bring the greatest reward. But such a conclusion need not be an intention. The monkey's intention to get the juice reward, for instance, may combine with this perceptual judgment in order to produce action, but the intention itself may be coded in neural activity

17 American Law Institute, Model Penal Code and Commentaries, Philadelphia: American Law Institute, § 2.02(2)(a)(i).

18 American Law Institute, Model Penal Code and Commentaries, Philadelphia: American Law Institute, § 2.02(2)(c).

19 Importantly, Shadlen tends to portray these experiments as investigating decisions rather than intentions. But this is not the universal interpretation of their results.

20 Kenneth H. Britten et al., *The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance*, 12 J. Neurosci. 4745, 4765 (1992); Michael N. Shadlen & Willaim T. Newsome, *Motion Perception: Seeing and Deciding*, 93 Proc. Natl. Acad. Sci 628, 633 (1996); Joshua I. Gold & Michael N. Shadlen, *Representation of a Perceptual Decision in Developing Oculomotor Commands*, 404 Nature 390, 394 (2000); Joshua I. Gold & Michael N. Shadlen, *The Neural Basis of Decision Making*, 30 Annu. Rev. Neurosci. 535, 574 (2007).

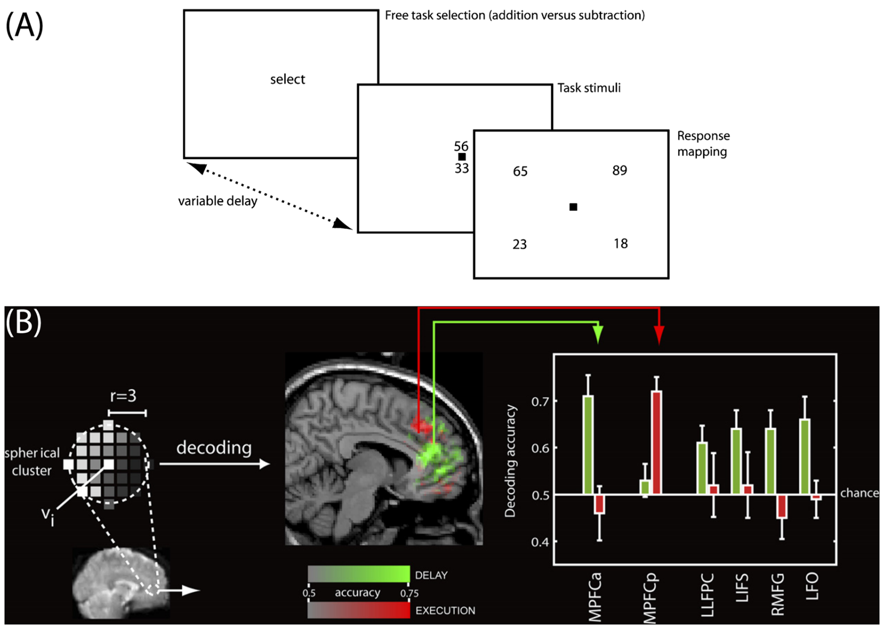

Figure 3. (A) an illustration of the task progress. (B) Brain regions encoding the subjects' specific intentions during either the delay (light gray) or execution (dark gray) periods. (MPFCa, anterior medial prefrontal; MPFCp, posterior medial prefrontal cortex; LLFPC, left lateral frontopolar cortex; LIFS, left inferior frontal sulcus; RMFG, right middle frontal gyrus; LFO, left frontal operculum.) (**Adapted with permission from Haynes et al., 2007.**)

distinct from what is found in LIP or MT. The neural activity that such studies associate with an 'intention' to move towards a target could, instead, underlie the reaching of a factual conclusion—a conclusion about which direction the dots are moving, and so which direction promises the greater juice reward. But beliefs about matters of fact are not *intentions* in the sense identified by the *Model Penal Code*.

Studies in which subjects are induced to have intentions of a sort that are closer to those that are of legal relevance can be found.[21] For instance, Haynes and colleagues had subjects decide whether they would prefer to add or subtract two double-digit numbers, without at that point knowing what the numbers were. After a variable delay period lasting a few seco nds, they were presented with the two numbers for two seconds. Then followed a screen containing the sum and difference of the two numbers as well as two distractors, randomly placed, and the subjects indicated their choice (Fig. 3A).[22] The experimenters strived to decode whether subjects intended to add

21 The law is concerned both with so called 'distal' intentions, or intentions to act in the future, and 'proximate' intentions, or intentions to act now. However, distal intentions matter to criminal liability only to the degree to which they are executed through the formation of later proximal intentions. In the study described in the main text, the focus is on the neural activity underlying the formation of the distal intention, but the experimental setup assures that subjects that follow the experimenter's instructions will act on that intention later, presumably through the ultimate formation of a proximate intention.

22 The random placement of the sum, difference, and distractors meant that the intentions decoded in this study were not simply for the upcoming motor activity, as the subjects could not know the mapping of the response buttons to the answer alternatives until the final screen was presented.

or subtract the numbers from brain data recorded using fMRI during the delay period, before the numbers were presented. They achieved an accuracy of 71 per cent from the anterior medial prefrontal cortex and above 60 per cent from other regions of the lateral prefrontal cortex, all at significantly above-chance levels. None of these areas contained addition/subtraction-divergent information after the numbers were revealed (Fig. 3B).[23]

In the Haynes study above, the subjects had no reason to keep their choices hidden from the experimenter. Yet, to what extent could intentions be decoded from neural activity in a competitive situation where revealing them is detrimental? Maoz and colleagues decoded the motor intentions of participants in a matching-pennies game. Sitting across from their opponent, the two players had to raise one hand at the go signal. And it was agreed that the subject would win $0.10 from his opponent if the hand he raised was a mirror image of his opponent's. Otherwise, he would lose $0.10 to his opponent. Both players started with $5, and if the subject was the overall winner over the 50 trials of the game, he received his final winnings in cash. The subjects—consenting epilepsy patients implanted with intracranial electrodes for clinical purposes—therefore had every incentive to keep their intentions hidden from the experimenter (Fig. 4A), though they were not explicitly informed that their brain activity would be used to decipher their actions. Nevertheless, subjects' intentions could be decoded at rates of about 70 per cent correct, on average, online and in real time, and used against them in the game. The prediction accuracy rose to 83 per cent, on average, in more rigorous offline analysis, and up to 92 per cent, on average, when the system was allowed to make predictions only on the 70 per cent of the trials on which it was most confident (Fig. 4B).[24]

But correlation does not entail causation. Is the brain activity on the basis of which the subjects' intentions could be decoded the same as that through which the intention is stored, or does it merely accompany it? In addition, there is mounting evidence that the intuitively appealing model of serial decision making guiding experiments of this kind may not be accurate. Under this intuitive model, decision-making involves a three-step process: (1) gathering information from the senses to form a percept, (2) forming an intention to act in accordance with a decision made on the basis of the percept, and (3) executing the action. Instead, some evidence suggests that sensory information is used to continuously specify several potential actions in parallel, and often in the same brain regions that later control the chosen behavior.[25] These processes do not stop once

[23] John-Dyaln Haynes et al., *Reading Hidden Intentions in the Human Brain*, 17 Curr. Biol. 323, 328 (2007); Chun S. Soon et al., *Predicting Free Choices for Abstract Intentions*. 110 Proc. Natl. Acad. Sci. 6217, 6222 (2013).

[24] Uri Maoz et al., *Predicting Action Content On-Line and in Real Time before Action Onset – an Intracranial Human Study*, in 25 Advances in Neural Information Processing Systems (2012).

[25] Paul Cisek & John F. Kalaska, *Neural Mechanisms for Interacting With a World Full of Action Choices*, 33 Annu. Rev. Neurosci. 269, 298 (2010); David Freedman & John Assad, *A Proposed Common Neural Mechanism for Categorization and Perceptual Decision*, 14 Nat. Neurosci. 143, 146 (2011); Joshua I. Gold & Michael N. Shadlen, *The Neural Basis of Decision Making*, 30 Annu. Rev. Neurosci. 535, 574 (2007); Joseph W. Kable & Paul W. Glimcher, *The Neurobiology of Decision: Consensus and Controversy*, 63 Neuron 733 (2009); Michael N. Shadlen et al., *Neurobiology of Decision Making: An Intentional Framework*, in Better Than Conscious? Decision Making, the Human Mind, and Implications for Institutions (Engel C. & Singer W., eds, 2008), 71, 102; Steven P. Wise et al., *Premotor and Parietal Cortex: Corticocortical Connectivity and Combinatorial Computations 1*, 20 Annu. Rev. Neurosci. 25, 42 (1997).

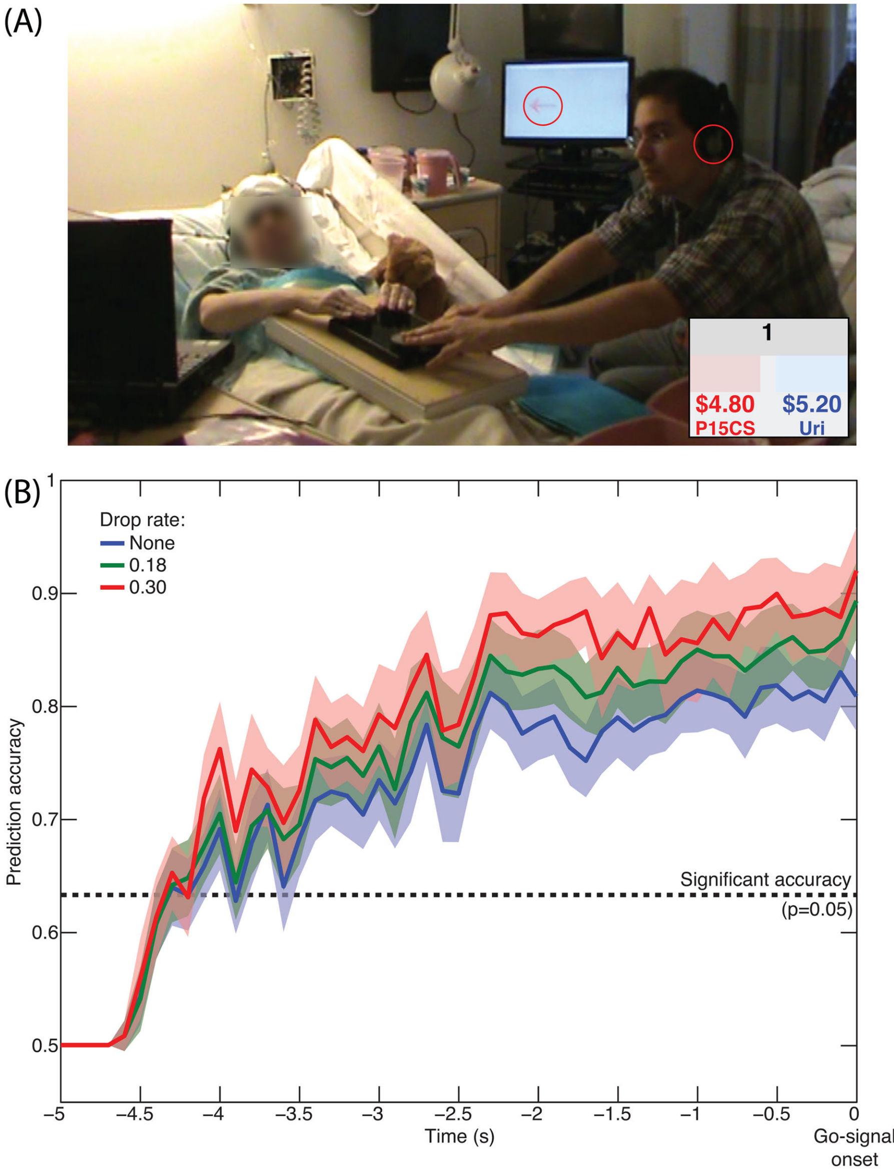

Figure 4. (A) The experimental setup in the clinic. The patient and experimenter are watching the game screen (inset on bottom right) on a computer (bottom left) and still pressing down the buttons of the response box. The real-time system already computed a prediction, and thus displays an arrow on the screen behind the patient and plays a tone in the experimenter's ear ipsilateral to the hand it predicts he should raise to beat the patient. (B) Across-subjects average of the prediction accuracy (mean ± s.e.m. shaded) versus time before the go signal. Values above the dashed horizontal line are significant at $p = 0.05$. (**Adapted with permission from Maoz et al., 2012.**)

the movement begins, even when sensory input is then suspended, facilitating potential changes of mind.[26] Therefore, intention might not be clearly mappable onto a single neural process, and the brain may contain representations of multiple, conflicting action plans. If so, then further conceptual work will be needed to determine what it is, exactly, that distinguishes an intention, in the sense that matters to criminal responsibility, from other mental states representing action plans in accordance with sensory information.

Turn now to recklessness—awareness of a substantial and unjustifiable risk of a fact of legal importance; for example, risk that the iPad was stolen, in the example above. Many neuroscientific studies examine the way in which people process probabilistic information and use it in decision-making. One famous set of studies use the Iowa Gambling Task, developed by Damasio and colleagues. Normals and patients with bilateral ventromedial prefrontal cortex (vmPFC) damage picked cards from four decks randomly bearing positive or negative monetary rewards. The total positive reward in decks A and B were double that in decks C and D. But A and B were stacked so that selecting cards from them would result in a loss overall, while selecting from C or D would result in a gain. The players, who were instructed to gain as much money as possible, knew nothing about the setup of the decks or when the game would be stopped. Normals began to choose advantageously—ie selecting mainly from C and D—before they could report any knowledge about the advantageous strategy. They also generated anticipatory skin-conductance responses (SCRs) before selecting from the riskier A or B decks, and 70 per cent of them could spell out how the decks were set up before the end of the game. The patients neither chose advantageously—generally preferring cards from A and B—nor generated SCRs, not even the 50 per cent who eventually could explicitly explain how the decks were set up.[27] It therefore seems that, at least when people are asked to make decisions such as those involved in the Iowa Gambling Task, the vmPFC is required for collecting information about risks to be used in the guidance of reward-seeking action.

In another study, Glimcher and colleagues tested for neural differences in the processing of *risk* and *ambiguity* using fMRI. Both are situations where the outcomes associated with decision alternatives are not certain. Under risk, the probabilities of the different outcomes can be estimated, whereas under ambiguity these probabilities remain unknown. In the main experiment, subjects were instructed to blindly extract one chip from an urn containing overall 60 red and blue chips, and told which color was associated with a cash reward; the other color was associated with no reward. They were also informed of the size of the reward—5, 9.5, 18, 34 or 65 dollars. In risk trials, the subjects were told the proportion of red and blue chips in the urn (12.5, 25 or 37.5, 62.5, 75 or 87.5 per cent red). In the ambiguity trials, the subjects were only informed of the range of the proportion of red and blue (37.5 to 62.5 per cent, 25 to 75 per cent, or 12.5 to 87.5 per cent red). The subjects had to decide between either the risky or

[26] Arbora Resulaj et al., *Changes of Mind in Decision-Making*, 461 NATURE 263, 266 (2009); Luc P. J. Selen, Michael N. Shadlen & Daniel M. Wolpert, *Deliberation in the Motor System: Reflex Gains Track Evolving Evidence Leading to a Decision*, 32 J. NEUROSCI. 2276, 2286 (2012).

[27] Antoine Bechara et al., *Deciding Advantageously Before Knowing the Advantageous Strategy*, 275 SCIENCE 1293, 1295 (1997); Antoine Bechara et al., *Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex*, 50 COGNITION 7, 15 (1994).

ambiguous choice on the one hand and a 50 per cent chance of winning $5 on the other hand. Three of their choices were selected at random, and subjects took out chips from urns with the parameters in those trials at the end of the game, winning the cash reward if they took out the right chip. The researchers found that the subjective value of the choice under both risk and ambiguity activated the same brain areas (up to fMRI resolution)—medial prefrontal cortex and striatum—although the activation for ambiguity was greater than for risk.[28]

The law does not distinguish between risk and ambiguity. A defendant who knows when he fires a five-chamber gun at another that the gun contains two bullets, is acting while aware of a risk. A defendant who fires the same gun while knowing that it contains either one, two or three bullets acts while aware of ambiguity. But both defendants would be classified as reckless under the law. The fact that Glimcher's study seems to indicate a neural commonality in the representation of risk and ambiguity suggests that the law may be right not to distinguish the two when it comes to criminal responsibility. This is not to imply, however, that the law would need to be changed were it discovered that neural representations of risk and ambiguity are distinct. The possibility would remain that the different neural activities have the same import for criminal responsibility.

While the representation of probabilistic information is a part of the legal notion of recklessness, it is not exhausted by that idea. Even setting aside the fact that for recklessness the probabilities in question must be 'substantial' and conduct in light of them must be 'unjustifiable', a crucial aspect to the legal notion of recklessness is *conscious awareness* of probabilistic information. The thought is not just that the reckless agent represents information about the risks of harming others but also that he acts in a way that imposes those risks while *aware* of the possibility, although not the certainty, of harm. Awareness of risks is crucial to the thought that the reckless agent is *criminally culpable*. Where there is awareness of risk, and action that imposes the risk, there seems to be a disregard of the importance of the harms that are being risked. To date, few if any neuroscientific studies have investigated the distinctive nature of *conscious awareness* of risk, distinguishing its neural basis, and role in decision making, from tacit, or unconscious representations of probabilistic information.

One area where the law could particularly use assistance from neuroscience concerns the impact of mental disorders on mental states crucial to criminal responsibility. Broadly speaking, for instance, we know that addiction involves disruption of dopamine signals and also that brain areas mediated by dopamine, such as the striatum, are crucial to processing and learning from probabilistic information.[29] We would therefore expect that addicts would be consciously aware of risks in different ways, and in different patterns, from non-addicts. How exactly addiction modulates conscious awareness of risk would be important information for the legal system. A large number of addicts find their way into courtrooms, and a crucial question in many of their trials is whether

28 Ifat Levy et al., *Neural Representation of Subjective Value under Risk and Ambiguity*, 103 J. Neurophysiol. 1036, 1047 (2010).

29 Howard C. Cromwell & Wolfram Schultz, *Effects of Expectations for Different Reward Magnitudes on Neuronal Activity in Primate Striatum*, 89 J. Neurophysiol. 2823, 2838 (2003); Bernard W. Balleine, Mauricio R. Delgado & Okihide Hikosaka, *The Role of the Dorsal Striatum in Reward and Decision-Making*, 27 J. Neurosci. 8161, 8165 (2007). See also [ref to Volkow et al. chapter in this volume].

they were reckless with respect to the harms they caused when they acted. But at the moment, the legal system does not incorporate any empirical information about the ways in which addicts represent and process information about risks when reaching judgments about their criminal responsibility. In part, this is because the legal system is slow to incorporate scientific information, and with good reason. The gap between the lab and the world is often too wide to warrant changing the way legal judgments are made. But it is also because scientific studies of decision in the face of risk, and the ways in which decisions are modulated by mental disorders, have not been guided by the conception of risk perception that matters to the criminal law, namely the legal conception of recklessness. Significant progress could be made in this area— progress that might point the way to substantial legal reform—by neuroscientific research that would be guided from the outset by the entrenched legal concepts, like that of risk perception. Similar points can be made about the bearing of mental disorders on intention, and other mental states that matter to *mens rea*. More work in this area can be, and should be done.

I.3. Self-control

The thought that the capacity to control oneself is crucial for responsibility is a deep-seated feature of our moral lives. Intuitively speaking, of those who harm others, some are in control when they do so; others have lost control, but could have been in control had they, for instance, stopped to think at an earlier time; and still others have lost control, and it was impossible for them to have maintained it. And these intuitive differences seem to matter to responsibility. These three kinds of agents, that is, seem to differ from a moral point of view.

For the most part, the criminal law is insensitive to these moral differences, although it is a further question whether it should be. It is very rare for a difference in treatment under the criminal law—punishment rather than release, or more punishment rather than less, for instance—to turn on a finding to the effect that the defendant was or was not in control, or could or could not have been. There are some exceptions. For instance, a defendant who is very upset when he kills another person will, if various other conditions are met, be guilty not of murder but of the lesser crime of manslaughter, which can bring substantially lower penalties. It is possible that the law grants mitigation in such cases through a recognition of the fact that at least some who kill while in a state of extremely heightened emotion are less than fully in control of what they are doing. However, similar mitigation is not available to those who, for instance, commit the crime of destruction of property while very upset, as when a person smashes his girlfriend's windshield after an argument. Mitigation in such cases is specific to homicide, which is a far less common crime than many others. This indicates the rather stingy attitude in the law towards basing differences in treatment on differences in control. Still, there are other important exceptions to this. Under one formulation of the insanity defense, although by no means the most common formulation, it is sufficient for excuse from criminal responsibility that, due to mental illness that gave rise to the relevant conduct, the defendant 'lacked substantial capacity to conform his conduct to the requirements of law'.[30] Under this so-called 'volitional prong' of the insanity defense, those who

[30] American Law Institute, Model Penal Code and Commentaries, Philadelphia: American Law Institute, § 4.01(1).

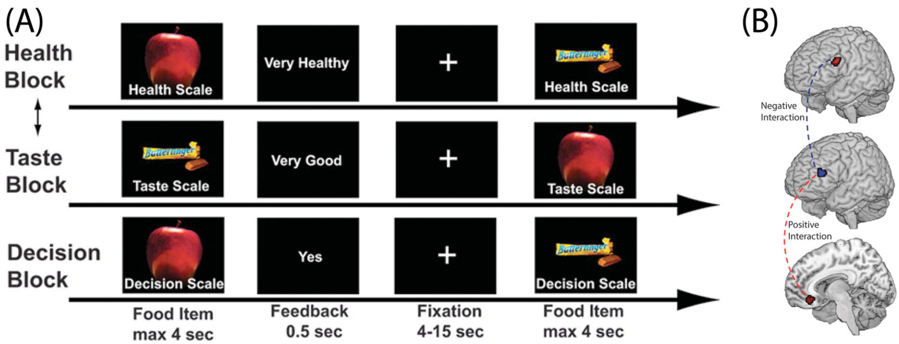

Figure 5. (A) Depictions of the experimental progress during the health, taste and decision blocks (in grayscale). (B) The vmPFC (top) may control the DLPFC (bottom) through an intermediate brain region like IFG (inferior frontal gyrus, BA 46; middle). (**Adapted with permission from Hare et al., 2009.**)

cannot, or find it extremely difficult, to do what the law requires of them are not subject to criminal punishment (although as noted in the introduction they may be subject to commitment in a mental institution). In addition, the Supreme Court has recently ruled that adolescents who have committed very serious crimes, including murder, ought not be punished as severely as otherwise identical adults in part on the grounds that adolescents, as a group, are more impulsive than adults.[31] So, the court seems to predicate an important difference in legal treatment of adolescents, although not adults, on a difference in the capacity for self-control. It is also possible that a deeper scientific understanding of control and its limits could come to inform the law. Perhaps if more were known about the factors that influence self-control, and impose limits on it, and more were known about how self-control can be measured, the legal system would come to predicate more differences in treatment under the law on differences in control. Only time will tell.

Important neuroscientific work on self-control has emerged in recent years, although it remains uncertain how, if at all, it bears on criminal responsibility. For instance, Rangel and colleagues first instructed hungry dieters and non-dieters to make choices about either the healthiness or taste of various foods on a five-point scale (Fig. 5A). One item that was rated as neutral on both scales was then taken as the reference food for each subject, and subjects were told to repeatedly choose between different foods and that reference food. To make the choices concrete, one trial was selected at random and the subjects had to eat the food selected on that trial. They then divided the subjects into self-controllers (SC) and non-self-controllers (NSC) based on their behavior during the experiment (eg declining unhealthy, liked items).

Using fMRI, the researchers found that activity in the vmPFC was correlated with the subjects' choices, regardless of the self-control they exhibited. Also, vmPFC activity was correlated with both taste and health for SC, but only with taste for NSC. In contrast, dorsolateral prefrontal cortex (DLPFC) activity increased during successful self-control trials and was then also correlated with

31 Roper v. Simmons (543 U.S. 551 (2005)), Graham v. Florida (130 *S.Ct.* 2011 (2010)), Miller v. Alabama (132 *S.Ct.* 2455 (2012)).

vmPFC activity. They therefore suggest that vmPFC reflects short-term goals (taste), which are then modulated by long-term considerations (health) using the DLPFC. And that the extent to which DLPFC can modulate vmPFC activity accounts for much of the difference between successful and unsuccessful self-control (Fig. 5B).[32]

Work can also be found on the neural substrates of psychological tools that people can employ, with varying degrees of success, in resisting temptation. For instance, Ochsner and colleagues investigated the neural bases of two emotion regulation strategies: attentional distraction and cognitive reappraisal. Distraction draws selective attention away from evocative aspects of an event (eg telling your child a story while she receives a shot). Reappraisal involves cognitive effort to reinterpret a situation's affective meaning (eg learning to reinterpret criticism as constructive rather than a threat to self-esteem). Before showing subjects negatively affective pictures for 8 s, the researchers instructed them to attend to them (control), decrease the affect (reappraisal) or remember a string of letters (distraction).

They found that both strategies successfully reduced emotional experience and amygdala activity, while engaging prefrontal and cingulate regions implicated in cognitive and emotional control. Reappraisal was a more effective strategy, and preferentially activated medial prefrontal and anterior temporal regions associated with affective meaning. Distraction decreased amygdala activity more while increasing prefrontal and parietal activation, associated with selective attention, to a greater extent. The researchers therefore hypothesized that while distraction simply resulted in decreased processing of affective meaning, reappraisal required the regulation of attention and processing of affective meaning.[33]

How work of this kind bears on criminal responsibility is far from clear. Self-control appears to be highly variable within subjects, and highly context-sensitive. Someone with grave difficulty in resisting tempting food might have no problem resisting temptations to steal. And someone who has trouble resisting the temptation to mildly harm another person while playing a game in a lab might never be tempted, or might have no trouble resisting such temptations, when not playing games, or when outside the lab. In addition, how much, and what sort, of self-control is required for moral responsibility, much less criminal, is itself difficult to specify. So very little can be said with confidence about what recent work on the neuroscience of self-control implicates, if anything, about criminal responsibility. Potentially, however, work of this kind could provide the first steps in helping us to determine in what ways particular psychological disorders do indeed result in the absence of 'substantial capacity to conform one's conduct to law' in the sense that is sufficient for insanity under some legal definitions. Much work, however, would need to be done before such steps could be taken.

[32] Todd A. Hare, Colin F. Camerer, & Antonio Rangel, *Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System*, 324 SCIENCE 646, 648 (2009).

[33] Kateri McRae et al., *The Neural Bases of Distraction and Reappraisal*, 22 J. COGN. NEUROSCI. 248, 262 (2010).

II. CONCLUSION

As indicated in the introduction to this article, many of the most difficult, and deepest, questions about criminal responsibility and the brain are no more, or less, tractable in light of recent neuroscientific experiments. Is there something about the dependency of our mental life on the state of a physical organ that is incompatible with criminal responsibility? If so, then descriptions of how exactly the mind, and the behavior to which it gives rise, are dependent on that organ will likely tell us nothing about criminal responsibility. But on the assumption that criminal responsibility is not just here to stay—as it surely is—but *should be* here to stay, there is room to ask to what extent neuroscience can illuminate the nature and underlying mechanisms of those features that contribute to and constitute criminal responsibility. The question is very much worth asking, and while some work has already been done in this direction, as described in the previous sections of this article, there is much more work ahead for a confidently positive answer.[34]

ACKNOWLEDGEMENTS

This work was funded by the Ralph Schlaeger Charitable Foundation and Bial Foundation for UM, Templeton funded 'Big Questions in Free Will Initiative' at Florida State University for UM and GY and the Mellon Foundation New Directions Fellowship for GY. In addition, preparation of this article was supported, in part, by a grant from the John D. and Catherine T. MacArthur Foundation to Vanderbilt University. Its contents reflect the views of the authors, and do not necessarily represent the official views of either the John D. and Catherine T. MacArthur Foundation or the MacArthur Foundation Research Network on Law and Neuroscience (www.lawneuro.org).

A shorter and less detailed version of this manuscript was previously published as the 'Cognitive Neuroscience and Criminal Responsibility' chapter in Gazzaniga et al., 'Cognitive Neuroscience'.[35]

[34] The authors thank Adina Roskies and Walter-Sinnott Armstrong, and other participants in a session discussing a draft in July 2013, for helpful comments.

[35] Uri Maoz & Yaffe Gideon, *Neuroscience and the Law*, in COGNITIVE NEUROSCIENCE, THE BIOLOGY OF MIND (4th ed., Michael S. Gazzaniga et al. eds., 2013) 1025, 1033.