Published Entry for the Encyclopedia of Human Behavior 3rd Edition (Editor in Chief VS Ramachandran) Elsevier (Oxford, UK)

Forthcoming (after substantial revisions) as a Reference Module on Neuroscience and Biobehavioral Psychology, Elsevier (Oxford, UK)

**Neuroethics**

**Glossary**

Causal Determinism: The idea that every event, including human thought and action, is necessitated by antecedent events and conditions together with the laws of nature.

Cognitive enhancement: The amplification or extension of core cognitive capacities through improvement or augmentation of information processing systems and/or decision-making.

Cognitive therapy or treatment: An intervention aimed at the correction of a specific pathology or deficit in cognitive capacities.

Compatibilism: The view that human action can be understood to be free even if it is fully determined by a physical–causal chain of prior events, if the act is linked in a particular way to a person’s desires or psychological states.

Deep Brain Stimulation (DBS): The implantation of electrodes within the brain to produce electrical impulses that may regulate abnormal impulses or impact cells or chemicals in the brain.

Dualism: The view that humans have two components: the physical self and a nonphysical mind or soul.

Electroencephalography (EEG): The recording of the electrical activity along the scalp produced by the firing of neurons within the brain. These electrical voltages, often called brain waves, vary as a function of time and type of brain activity (e.g., the EEG of a person in a coma will look different from the EEG of a normal, awake adult).

Incompatibilism (including libertarianism and hard determinism): The view that free will exists only where an act is uncaused by prior physical effects, but is not random or uncaused. Libertarians hold that free will exists and thus causal determinism is false; hard determinists hold that causal determinism is true and thus free will doesn’t exist.

Magnetic resonance imaging (MRI): An imaging technique used to visualize detailed internal structure and function of the body. MRIs use a powerful magnetic field to align the nuclear magnetization of atoms in the body. Radio frequency (RF) fields are then used to produce a signal that can construct an image of the internal structure of the body. Functional MRIs (fMRIs) measure signal changes in the brain due to changing neural activity by detecting increases and decreases in blood flow.

Mind/Brain Reduction: The notion that the mind can be understood as just the same thing in the world as some portion of the brain (ontological reduction), or the mind can be understood in terms of a theory of the brain (epistemological or theoretical reduction).

Moral enhancement: The use of cognitive treatment or enhancement to improve moral decision-making.

Neuroethics: The body of work exploring the ethical, legal, and social implications of neuroscience.

Neuroimaging: Technologies such as computed tomography (CT), anatomical and functional magnetic resonance imaging (aMRI and fMRI), and positron emission tomography (PET), which can display the structural and functional neurobiological bases of our cognitive capacities and psychological states.

Physicalism: The view that humans are made of only one component, the physical stuff (and thus there is no immaterial soul or immaterial mental state).

**Synopsis:**

Neuroethics is the body of work exploring the ethical, legal, and social implications of neuroscience. This work can be separated into two rough categories. The neuroscience of ethics concerns a neuroscientific understanding of the brain processes that underpin moral judgment and behavior. The ethics of neuroscience, on the other hand, includes the potential impact advances in neuroscience may have on social, moral and philosophical ideas and institutions, as well as the ethical principles that should guide brain research, treatment of brain disease, and cognitive enhancement. Central to the questions posed in neuroethics is the way in which neuroscience might impact our sense of self and personal responsibility, and our understanding the structure of moral judgments.

**Body:**

The term “neuroethics” was coined in 2003 by political journalist and New York Times columnist William Safire. At a conference convened by Safire and the Dana Foundation, of which Safire was the Chairman, Safire defined neuroethics as “the field of philosophy that discusses the rights and wrongs of the treatment of, or enhancement of, the human brain.” He then charged 150 neuroscientists, bioethics scholars, policy-makers, and journalists with the task of exploring the ethical, legal, and social implications of neuroscience.

Safire’s definition has since been expanded to include both the ethical implications of neuroscientific research – the ethics of neuroscience – but also the neuroscience of ethics. According to Adina Roskies, the neuroscience of ethics is concerned with a neuroscientific understanding of the brain processes that underpin moral judgment and behavior. The ethics of neuroscience, on the other hand, includes the potential impact advances in neuroscience may have on social, moral and philosophical ideas and institutions, as well as the ethical principles that should guide brain research, treatment of brain disease, and cognitive enhancement.

This entry will discuss these different aspects of neuroethics, with a special focus on the way in which neuroscience might impact our sense of self and personal responsibility, a concern that cuts across these categories. For example, I will discuss whether advancing knowledge of brain states or processes undermine common notions of free will and responsibility. I will also examine ethical implications of cognitive enhancement, including moral enhancement, and whether certain treatments of brain abnormality are ethical (is it acceptable to irreversibly ‘cure’ pedophilia or obsessive/compulsive disorder?). Finally, I’ll discuss whether the neuroscience of ethics can provide insight into who should be deemed criminally responsible via a neuroscientific analysis of intentional action.

1. **Neuroscience, the Self, Free Will, and Responsibility**

Neuroscience is a scientific discipline aimed at understanding how human behavior and capacities are related to brain structure and systems. Neuroscience includes neurophysiology, neuropsychology, and neuroanatomy, and utilizes techniques such as Electroencephalography (EEG) as well as brain imagining technologies such as computed tomography (CT), anatomical and functional magnetic resonance imaging (aMRI and fMRI), and positron emission tomography (PET). These technologies can display the structural and functional neurobiological bases of our cognitive capacities and psychological states and processes, such as decision-making, desires, memories and emotions.

Other disciplines that treat human beings as the objects of scientific inquiry include anatomy, biology, and chemistry. However, neuroscience focuses upon the operations of the nervous system, which is the locale of capacities closely associated with personality, character, and the self. Thus neuroscientific findings are thought to have deep philosophical implications with regard to who we are (and possibly, what we are).

So, who are we, according to neuroscience? Neuroscience describes the brain as a causal machine. This means the brain goes from state to state as a function of prior, or antecedent, conditions: if the antecedent conditions were different, the result would be different; if the antecedent conditions were the same, the result would be the same. For example, at the beginning of some of these functional state changes is perceptual (or sensory) input, and at the end of some series of changes is a behavioral output (action).

Somewhere in between the input (perception of the environment) and output (action) thoughts, preferences, memories, and decisions seem to reside. Neuroscience provides us with new means to examine the way in which these mental states and processes are instantiated in the brain by allowing scientists to observe the brain undergoing processes in response to stimuli via functional MRI (fMRI) machines. Data from an fMRI scanner shows what sites in a subject's brain are active (indicated by an increase of the blood flow signal) during mental activity, such as the experience of particular conscious mental states. Additionally, anatomical MRI and PET scans allow scientists to view the structure of the brain, and thus allow for identification of abnormalities in brain tissue. It may then be postulated that such abnormalities impact brain function in particular ways.

Neuroscience also facilitates the discovery of new means of intervention in mind/brain processes. Although indirect cognitive interventions, such as meditation, exercise, and nutrition plans have been used to improve decision-making for a very long time, direct cognitive interventions became available fairly recently due to progress in neuroscience. Medicines such mood-enhancers (SSRIs), stimulants (Adderall), and treatments for fatigue (modinafil), as well as neuro-techniques such as deep brain stimulation (DBS), all aim to impact the mind via impact on brain function.

Neuroscience is thus thought by some to comment upon the age-old philosophical debate regarding the nature of the human mind, and the relationship between the mind and brain.

**Neuroscience and the mind**

Long before neuroscience informed our ideas of how the brain operates, many philosophers and scientists were already convinced that the position of dualism – that humans are made out of two types of stuff, a physical part and non-physical mind or soul – was false. Many feel there is little evidence in support of an immaterial mind, especially if one thinks this immaterial mind has causal effects. Neuroscience adds to the already substantial body of evidence that brain states and processes either enable or constitute our psychological states and processes, including thoughts, memories, and even higher mental functions such as consciousness and reasoning.

The tools of neuroscience thus provide further evidence that the brain constitutes the mind, and may even eventually describe which parts of the brain constitute which mental states or processes. However, neuroscience does not necessarily rule out the possibility of a non-physical aspect of the mind. Some think that conscious states, such as pain or love, are just what it feels like to undergo certain neural processes. Frank Jackson, however, has argued that the conscious feel of a mental state, such as the perception of the color red, could be an immaterial “add-on” to a physical brain state. In this case, however, it is most likely that such immaterial states, often termed “qualia”, are epiphenomenal: that is, they have no physical causal effects. This means that it would be the neural instantiation of the state of pain or thirst that causes behavioral effects, not the immaterial conscious feel.

Although neuroscience may not definitively rule out immaterial qualia, it does seem to indicate that the cognitive capacities we consider most representative of the “self”, such as decision-making capacities and modulation of emotions, can be functionally identified as discrete states or processes within the brain. fMRI studies have allowed us to isolate and locate stored representations (memories) as functionally dissociated from the processes that operate upon them (decision-making). The prefrontal lobes primarily contain the thought processes, often called executive processes by neuroscientists, which monitor and manipulate representations and are linked with conscious decision-making. Mental representations or memories, however, are thought to reside in the posterior cortical regions, in the temporal and parietal lobes. (An exception may be motor representations, and some representations involved in procedural or working memory.)

As we shall see, neuroscience’s increasing ability to provide a structural and functional description of the brain has far-reaching implications. First, it could allow us to designate a portion of the brain as particularly important because of the role it plays in decision-making. This, in turn, may result in a heightened need to protect the area against manipulation or even enhancement (see below). Second, neuroscience may be able to not only identify brain abnormalities, but to explain why certain disorders of the brain are experienced as particularly harmful. Third, neuroscience could provide us with a better understanding of who is fully responsible for their actions, and who is not, via analysis of the function of part of the brain most closely associated with the self and decision-making.

**Free Will**

Because neuroscience advances the idea that the brain is a causal machine, it has provided further evidence that the traditional libertarian theory of free will – the notion that human beings are uncaused causers – is false. Neuroscience appears to support what scientifically-minded philosophers had long surmised: that there are no explanatory gaps in our brain processes wherein the causal chain is broken and libertarian free will steps in to cause action. Neuroscientific evidence, like a biological understanding of the conditions under which a neuron fires (or fails to fire), indicates that one need not appeal to an immaterial self or “soul” to explain human behavior.

In particular, the work of Libet caused many outside of philosophy to think neuroscience meant human free will was an illusion. Libet conducted experiments where he had subjects make voluntary hand movements while he measured their brain activity using event-related potentials. Subjects were told to report the position of a black dot when they made the conscious decision to move their hands. Then this moment was compared to the time when the subject’s brain appeared to change in relation to the desire to move the hand.

Libet discovered that the brain appeared to initiate the sequence resulting in the hand movement before the subject was consciously aware that he or she had made a decision to move: about 300ms elapsed between the beginning of brain activity and the decision. Thus it seems our brains know we are going to make a decision, and perhaps maybe even what the decision is, before we do. (The “we” here is the conscious agent.) Libet noted, however, there is still about 100 ms after the initiation of the act for the conscious mind to either endorse or veto the decision. Thus he argued that free will was not entirely an illusion, claiming that it consists in this veto power consciousness may have over a decision.

Many have criticized both Libet’s data and the philosophical conclusions he draws from the data. Roskies, for example, has argued that one may be consciously aware of conscious intent at some point after the conscious intent is formed. Regardless of how one interprets Libet’s results, however, it seems neuroscience does not force one to abandon the concept of free will. Many philosophers now accept the position of compatibilism – that human action can be understood to be free even if it is fully determined by a physical-causal chain of prior events. Walter Stace, for example, argues that free acts are “those whose immediate causes are psychological states in the agent”, regardless of whether those states are conscious (or not), physically instantiated, or part of a larger physical-causal chain. That is, it may be that your desire to get a drink was initiated by some (initially unconscious, then conscious) physical state equal to thirst, and that your decision to get up and open your refrigerator was causally determined by this physical state. But because these states are “yours” the action should be considered free. By contrast, unfree acts are acts whose immediate cause is some state of affairs external to the agent, such as one holding a gun to your head.

**Who is responsible?**

Some scholars have raised the alarm regarding the use of neuroscience in the criminal courts precisely because they feel that criminal responsibility requires libertarian free will. But as indicated above, even if neuroscience provides a new sort of explanation of the causes of behavior, this explanation does not necessarily constitute an excuse. The current criminal justice system can hold persons responsible even if human actions have full physical explanations: that is, even if a desire to kill is neuronally instantiated and has a complete causal history, the criminal justice system is designed to attribute responsibility when that desire leads to criminal harm.

Criminal responsibility requires that an action linked to criminal harm be immediately caused in a certain way by a defendant’s mental states. If a harmful act is connected to the actor's desires or goals, and if the actor held certain beliefs about the harm that could result from the act, then he or she is criminally responsible. Thus, to be guilty of murder under the U.S. Model Penal Code, for example, one must have a mens rea (or a mental state) that includes: (1) the desire to perform an act that results in the death of another; and (2) performance of the act with the desire to kill, knowledge one will kill, or with reckless indifference to the chance that one may kill.

It is unclear that it is even theoretically possible for neuroscience to provide a complete physical causal description of the mental states required for criminal responsibility -- the desire to kill and the understanding that the act will result in a death. This is because the mental states required for guilt will be differently realized across defendants. It is more likely that neuroscience could provide a functional account of certain brain processes necessary to the capacity to commit a crime. Regardless, the categories of culpability (such as murder, rape, and theft) and excuse (diminished capacity, juvenile status, and insanity) need not be impacted by findings in neuroscience. The breadth of legal category is a policy decision determined by legislatures and judges. Neuroscientific knowledge is best understood as a tool to help us better categorize defendants as guilty or not guilty based upon the existing categories of culpability and defense.

However, there is some chance that neuroscience will cause us to alter or abandon the categories we use to attribute criminal guilt. Some have suggested that the vast majority of offenders on death row suffer from severe prefrontal lobe damage. What if it turned out that the vast majority of murderers had severely compromised executive function, such that, using the delicate tools of neuroscience, we determined they couldn’t either control or understand their violent acts in the way that those without compromised brain function can? In other words, what if they didn’t form desires or understand the nature or consequences of their acts in a “normal” way? Would punishment still be justified?

In the US, criminal punishment tends to be justified on the basis of one or several of the “principles of punishment” such as deterrence, incapacitation, or retribution. Neither retribution nor deterrence would seem to justify severe punishment of persons with brain function compromised in a way relevant to criminal responsibility (e.g. with regard to intentional or rational action). Such persons would not be deserving of retribution if they truly didn’t understand the nature of the act they were performing or were unable to stop themselves from performing it. Similarly, an understanding of the act as criminal and the ability to stop oneself from performing the act is crucial to the effectiveness of deterrence.

However, just as we incapacitate those deemed legally insane or “guilty but mentally ill” despite the fact that they are ineligible for punishment, we might also decide to incapacitate violent offenders who lack the capacity to understand or prevent their criminal acts. Thus in this case, neuroscience might alter our theoretical handling of offenders with abnormal brain processes, but our actual treatment of the offenders – that they are incapacitated – may not change.

**Cognitive treatment and enhancement**

In addition to its implications for traditional philosophical concepts of mind, will and responsibility, there are other ethical concerns raised by advances in neuroscientific research. Neuroscience provides new methods to treat and enhance core brain functions and decision-making. Cognitive therapy or treatment aims to correct a specific pathology or deficit in cognitive capacities. Cognitive enhancement, on the other hand, is any intervention meant to amplify a person’s core cognitive capacities where they are not suffering from a defect. When the intervention is in the form of a drug, the drug sometimes is called a “nootropic”.

Pharmacological enhancement of brain function is already widely practiced via drugs such as selective serotonin reuptake inhibitors (SSRIs, prescribed for mood enhancement), Adderal (for improved attention and concentration), Modafinil (for wakefulness), and Aricept (for improved memory). Although originally conceived to make up for deficits in cognitive ability due to disorders such as depression and attention deficit hyperactivity disorder, such drugs are now often used by those with highly functioning, and even “normal” cognition, such as those diagnosed as having a vulnerability to depression (but without any current depressive symptoms), travelers to overcome jet-lag, and students trying to improve their academic outcomes.

Enhancement may also be achieved by brain implants. Brain implants have been used to lessen the effects of several common disorders, such as Parkinson’s and chronic pain. For example, in the case of Parkinson’s, a technique called Deep Brain Stimulation (DBS) uses a brain implant to deliver electrical impulses to a region in the center of the brain that controls and coordinates movement, in an attempt to regulate abnormal impulses. DBS has also been used in an attempt to stimulate various parts of the brain which appear dysfunctional in those who suffer from major depression and Tourette’s syndrome. In the future, it is possible that DBS could be used to enhance cognitive function on those without psychological disorders.

Scientists are currently making progress using both carbon nanotubes and conventional silicone implants in the brain. Ultimately, such implants may be used to manipulate thought, possibly via transmission or implantation of thoughts. One possible use of this technology could be to allow soldiers to fight better under combat situations, or to allow athletes to better focus on their sports.

Although neurological treatment programs seem to involve the same ethical implications as any medical treatment, brain enhancement is thought by some to involve unique worries. One obvious issue is the way in which enhancement will be distributed. It is likely that the wealthy and privileged will have more access to such enhancements, thus creating an even greater gap between socioeconomic classes. For example, a drug that assists in memory may mean that those who can afford the drug may have an advantage in school or the job market, and athletes with neuro-enhancements will be able to out-compete those without enhancement.

Some have argued that in order for humans to fully enjoy or appreciate life, they must experience both joy and pain. Overuse of SSRIs or other brain enhancing drugs might thus blunt a person’s ability to enjoy the full range of human experience. This is turn could have an impact on creativity or ones capacity to compose good art. Another worry has to do with a normative idea of the optimal human being or life. First, there is much disagreement about the capacities we may need to enhance to create the best life. Second, even capacities we might all agree are beneficial with regard to specific tasks could turn out to have negative outcomes overall. For example, there is some evidence that the average human level of forgetfulness actually helps us attend to the things that are important, instead of being distracted by unnecessary details. In general, there is certainly no evidence that having better mental capacities will make human beings happier.

**Moral enhancement**

Some have argued that we ought to use brain interventions to enhance moral decision-making. This could be done in at least two different ways. First, an intervention may increase or diminish the salience of moral emotions – the emotions that either drive immoral behavior, or work to support moral behavior. Second, an intervention might enhance the reasoning processes to allow agents to better perceive and weigh consequences of their actions, and inhibit immoral acts. In general, use of direct brain interventions would seem to be less worrying in the case of moral enhancement than in some other cases, because possessing good moral judgment would seem to increase a person’s self-control and agency.

But what if an agent wishes to continue to make immoral decisions and commit immoral acts, or wants to maintain a vicious trait? In this case undesired moral enhancement would seem to be an infringement on a person’s wishes to be a certain sort of self, and thus a violation of their agency. Involuntary moral enhancement is lawful under a handful of US states’ chemical castration laws (as well as the laws of a few other nations), which sentence some subset of criminal offenders to chemical castration as a part of their sentence. MPA (medroxyprogesterone acetate) is the drug most often used for chemical castration. The drug is an analogue of the female hormone progesterone, which reduces the normal level of testosterone in a male, thus reducing sex-drive, and often diminishing seminal ejaculator fluid to zero. MPA has to be continually administered, which requires non-incapacitated offenders to submit themselves to weekly shots.

One might imagine neurological castration as a replacement to this drug regimen. Hypothetically, neurological castration might directly inhibit activity in certain parts of the brain (e.g., within the hypothalamus), and block connectivity between areas of the brain (e.g. between representations of children and sexual arousal). A surgical approach to castration may be more successful, and have far fewer side effects, than current methods. All types of castration – traditional surgical, chemical, and neuro-surgical – may be less expensive than giving sex offenders long prison sentences.

There are many other ways that cognitive interventions can be used in criminal sentencing. Drug courts sometimes require defendants get therapy, including medicine, to assist with their drug addiction as part of probation or parole. There is some evidence that SSRI drugs can work to decrease aggression in violent offenders. It is possible some neuro-surgical technique could achieve the same effect. If it became possible to neurologically (via medicine or surgery) inhibit strong violent responses to stimuli, the state might offer violent offenders the chance to submit to treatment in exchange for a shortened or commuted sentence.

Neuroscience thus creates new means to achieve both rehabilitation and incapacitation of offenders, by providing new ways to impact their decision-making and behavior. Up to recently, persons who violate the law are allowed to remain the sort of person they are (even if that person was a pedophile), although the space within which they are allowed to be that person is limited to a jail or prison. And after an offender served their time, persons are released to continue to pursue their individual desires. However, neuroscience instead might allow us to permanently change a person’s character – even involuntarily – such that she no longer commits criminal acts.

Any neuroscientific manipulation or enhancement that changes a person’s loci of decision-making represents nothing less than changing their identity. Most philosophers believe that one’s psychological states – our beliefs, desires, memories and emotions – are a crucial component of one’s identity. Our psychological states are the source of our behavior: we act a certain way because we desire and believe certain things. If neuroscience changes what we want, what we know, or what we believe, it changes who we are.

Involuntary manipulation of an offender’s psychological states via neuroscience may thus be a severe violation of human agency in that it would infringe upon his ability to choose, and be responsible for, his own acts. However, use of brain interventions on criminal offenders may raise serious ethical concerns even if offenders participated in such programs voluntarily. Would we really be giving offenders a choice with regard to neurological alteration if they must choose between thirty years in prison and freedom after a “simple” operation? There are real worries that such an option would be no option at all.

**Privacy and the self**

Neurological interventions also have implications for privacy and the way in which we understand the “self”. Take, for example, the use of neuroscience as a “lie-detector”. The first lie detector consisted in a polygraph test. A polygraph is an instrument that measures and records physiological indices such as blood pressure, respiration, pulse, and skin conductivity while the subject is asked and answers questions. Although many dispute the accuracy of polygraph tests, some argue that lying produces a physiological response that can be differentiated from true answers.

A recently developed lie-detector technique, however, uses neuroscience. “Brain fingerprinting” devices use EEG readings in an attempt to identify changes in brain activity when a subject is presented with familiar information. An EEG records the electrical activity along the surface of the scalp produced by the firing of neurons within the brain. These electrical voltages, often called brain waves, vary as a function of time and type of brain activity (e.g., the EEG of a person in a coma will look different to the EEG of a normal, awake adult). In brain fingerprinting, changes in EEG readings are compared to the subject’s verbal report regarding what he knows. The device is thus thought to be able to determine whether a criminal suspect is familiar with some aspect of a crime, and hence whether the suspect is telling the truth concerning their involvement. Such a device could also be used by potential employers, or even insurance companies, to determine if a job candidate or person seeking insurance was being honest about their past work experience, or any preexisting conditions.

Because brain fingerprinting detects brain activity directly, it is thought by some to be a more accurate lie detector than use of a polygraph. This is because it does not entail detection of emotional responses, which can be consciously manipulated by some subjects. However, there is controversy over whether brain fingerprinting returns consistently accurate results. In addition, brain fingerprinting can only indicate whether a subject is familiar with certain information; the technique cannot indicate how the subject gained this information (e.g. via being present at the crime scene, or reading about the crime in the newspaper).

There are other ways that neuroscience may provide access to previously private mental states. Quantum dot technology is being used to gather information in the brain at the level of the neuron. Nano-sized functional quantum dots can help build data-capture devices that are easy to use by neuroscientists. Many feel that nanotechnology in conjunction with neuroscience will eventually allow for targeted interactions between neurons, the cells responsible for signal transmission in the brain. Ultimately, it seems that nanotechnology will allow us to visualize and track functional responses in neurons, and this means we may be able to learn information about a person’s thoughts remotely.

Several ethical issues are raised by the possibility of use of brain fingerprinting devices or functional quantum dots. Normally, an employer or a police officer must rely upon information voluntarily offered: regardless of what one is thinking, the information actually spoken is what counts, and most persons have control over the words they chose to speak. In so far as neuroscience allows others direct access to one’s consciousness or mental states, it would appear to entail a serious violation of privacy. And in the cases mentioned above, the consequences of the violation are severe: one might not get a job or qualify for health insurance, or in the case of the criminal law, one could lose their liberty or life.

1. **A Neuroscientific Understanding of Moral Judgment and Action**

Traditionally, philosophers such as Immanuel Kant argued that higher-level reasoning was crucial to moral judgment and action. However, it has become increasingly clear that moral judgments are not the sole product of reason or introspection isolated from emotion. Indeed, it turns out that emotional input is crucial to moral action and the judgments we make regarding whether others’ actions are right or wrong.

Neuroscientific research attempting to isolate the regions involved in moral reasoning and behavior support this conclusion. Neuroimaging indicates that there is no particular “moral center” of the brain. Widely distributed brain areas have been found to be activated during moral reasoning tasks, including those regions associated with higher cognitive functions and those involved in emotional responses. There appears to be a complex interplay between reasoning and emotions in the formulation of a moral decision.

It has proven difficult to parse brain regions associated with moral judgments from those regions associated with behavior that we might judge immoral or “bad”. This is because being able to determine that a situation requires a moral judgment, and treating a decision as a moral decision, may be important to moral behavior. If I don’t see my decision to steal my neighbor’s TV, or not to give money to the Red Cross, as a moral judgment with ethical implications, then it seems I am less likely to make a decision others will see as morally right.

**Moral Judgments**

Studies have found significant neurological differences between subjects making judgments on moral personal dilemmas, and those contemplating impersonal moral dilemmas and non-moral dilemmas. Moral dilemmas involving a personal component – where one is asked to consider whether they would personally perform some moral or immoral act – seem to activate the orbital, ventromedial and dorsolateral prefrontal cortex (PFC), the medial frontal gyrus, and possibly the posterior cingulate cortex, amygdala and inferior parietal lobe. These regions are associated with integration of emotions, mental imagery, and memories into decision-making; representation of rewards and punishments; and control of risky or inappropriate behavior. In addition, the medial frontal gyrus, anterior and posterior cingulate cortex, and temporal areas (such as the superior temporal sulcus and temporal pole) are also associated with moral judgment. These areas are correlated with theory of mind, the ability human beings have to attribute mental states to others as a means to understand and predict their behavior.

Antonio Damasio and colleagues found that patients with damage to the posterior ventromedial cortex have impaired ability to make moral judgments. However, their impairment was primarily emotional; they appeared to have an inability to generate and effectively use what Damasio calls “somatic markers”: neural representations of body states that give potential bodily actions a (sometimes unconscious) emotional feel making them more or less appealing. If you ask a subject without ventromedial damage to imagine running over another person in their car, he or she tends to have a negative emotional response in association with the mental image. Asked to imagine the same scenario, subjects with ventromedial damage have no problem forming a mental representation of performing the act and its consequences, or understanding that society would judge such as act as wrong or criminal. However, they fail to have a normal emotional response to the mental image.

It seems the more personal a moral dilemma, the stronger ones’ potential emotional response. Greene and Haidt scanned subjects using an fMRI while the subjects responded to a series of personal and impersonal moral dilemmas, as well as non-moral dilemmas. They determined a moral violation is personal if it is clear that the subject was hurting some particular other person, whereas it is impersonal if it didn’t require the subject directly harm someone else. A classic philosophical thought experiment, for example, asks whether the subject would pull a switch to move a trolley car from one track, where five people are likely to be hit by the trolley and killed, to a second track, where only one person is going to be hit and killed. This is an impersonal moral dilemma, because it does not require that the subject’s act directly cause harm, but only deflect harm from one party to another. On the other hand, when subjects are asked to push a fat man onto the tracks to derail the trolley to keep the five persons from being killed, the dilemma becomes personal, because it requires direct harm be caused by the subject to a person who would otherwise be safe.

Greene and Haidt found that responding to personal moral dilemmas, as compared with impersonal and non-moral dilemmas, produced increased activity in the medial frontal gyrus, posterior cingulated gyrus, and the angular gyrus. These areas are associated with social and emotional processing. Impersonal and non-moral dilemmas, on the other hand, produced increased activity in areas associated with working memory, such as the dorsolateral prefrontal areas. These findings again indicate the importance of emotional input in moral decision-making: subjects faced with a moral dilemma tended to be quick to moral condemnation but slow to approval of personal violations, apparently due to the salience of negative emotions, whereas approvals and disapprovals took equally long when subjects were faced with impersonal and non-moral dilemmas.

In sum, neuroscience indicates that many different capacities are important to moral reasoning and judgment, including normal decision-making and theory of mind capacities; however, the importance of emotional input may be particularly important to moral judgments (as opposed to many other types of judgments).

**Moral behavior**

As indicated above, there is significant overlap between the brain regions associated with moral judgment, and the regions associated with immoral or antisocial behavior. Regions common to both include the ventral and medial PFC, the amygdala and the angular gyrus/superior temporal gyrus. One difference, however, is that persons who tend toward antisocial behavior show hippocampal and anterior cingulate impairment, whereas moral judgment studies fail to indicate such activation is important to moral reasoning tasks. However, psychopaths also show frontotemporal-limbic gray matter reductions and reduction in connections between amygdala and OFC.

Adrian Raine and collegues have argued that some of the neurological impairments found in antisocial individuals disrupt moral/emotional decision making, which in turn acts as a factor predisposing such individuals to antisocial behavior. Some antisocial personalities, and psychopaths, appear to show normal or excellent moral reasoning ability, but appear to fail to apply the outcome of such reasoning processes due to a lack of emotional input.

As an example, some argue that the ventromedial frontal cortex (VM cortex) mediates between the neural systems for arousal and emotion, and thus is important to moral behavior. Patients with VM cortex damage tend not to show deficiency in moral reasoning: that is, in their ability to identify potential moral conflicts and potential solutions to such conflicts. However, at least some VM patients feel differently about the identified solutions than those without VM damage. Going back to the trolley thought experiment, both VM patients and persons without VM damage realize that killing the one fat man will save five others lying in wait on the other trolley track. However, persons without VM damage tend to judge it wrong to kill him. Persons with VM damage, however, are more likely to judge that pushing the fat man onto the tracks is the right thing to do.

Not surprisingly the difference between the two judgments seems to be emotional: as indicated above, many think psychopaths act immorally partially due to a lack of normal affect. Thus, persons with VM damage may be more likely to engage in antisocial or immoral behavior, precisely because they do not feel badly about such action. In conclusion, many argue that it is predominately the failure to feel of what is moral, rather than the knowing of what is moral, that leads many to antisocial or immoral action.

**Neuroscientific evidence and criminal responsibility**

From the perspective of the criminal justice system, it is difficult to understand what the conclusion that many antisocial persons lack emotional insight should mean. Thus far neuroscience has been introduced in criminal trials primarily to help the court determine whether a defendant had either minimal capacity to form the intent to commit a crime, or if he was capable of forming the highest level of intent (e.g. the intent to kill in the first degree or “purposely”, for which the defendant might qualify for capital punishment). In this way neuroscience has been understood to be a rough tool for commenting upon the basic decision-making capacities necessary to understand the nature and quality of one’s actions in a normal way.

For example, in 2003, Burns and Swerdlow reported a case of a 40-year-old schoolteacher in an otherwise normal state of health who developed an increasing interest in pornography, including child pornography. The patient also began soliciting prostitution at "massage parlors," which he had not previously done. Magnetic resonance imaging revealed a large tumor in his right orbitofrontal lobe. The patient’s symptoms disappeared after the tumor was removed. The tumor in the frontal lobes had impacted several important higher cognitive functions (executive processes), including inhibition (denial of immediate gratification), the moderation of emotional responses, and consideration of the potential future consequences of actions. Burns and Swerdlow hypothesized that the tumor had not given the patient new goals or desires, but had rather impacted the patient’s ability to conform his behavior to societal norms and laws.

The patient was arrested for sexual assault of his stepdaughter. If the case had gone to trial, it would seem that identification of the man’s tumor would seem to be extremely helpful to an accurate understanding and categorization of his behavior. Indeed, it would appear to be unjust to deny a judge or jury a chance to consider the neuroscientific evidence of the presence of a tumor in the man’s orbitofrontal lobes as evidence of diminished legal capacity.

However, this specific sort of use of neuroscientific evidence by the criminal law is quite different from imagining the implications of a finding that some high percentage of antisocial behavior is linked to brain abnormalities. In the above case, the tumor in the brain of the patient described by Burns and Swerdlow led him to behavior that was abnormal for the patient (and disappeared after the tumor was removed), indicating that less responsibility for an act linked to the tumor is appropriate. However, as stated above, if a high percentage of those who commit murders have certain common brain abnormalities, does that mean we have a better indicator of the sort of person we feel is dangerous and want to punish, or does it mean such persons are less culpable? This question does not appear to have an empirical answer, and seems to fall instead within the realm of public policy.

Katrina L. Sifferd, Elmhurst College (sifferdk@elmhurst.edu)

**Further Reading**

Bostrom, N., & Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics*, 15(3), 311-341.

Burns JM and Swerdlow RH (2003) Right orbitofrontal tumor with pedophilia symptom and constructional apraxia sign. *Archives of Neurology* 60: 437–440.

Casebeer WD and Churchland PS (2003) The neural mechanisms of moral cognition: A multiple-aspect approach to moral judgment and decision-making. *Biology and Philosophy* 18(1): 169–194.

Churchland PS (2006) Moral decision-making and the brain. In: Illes J (ed.) *Neuroethics: Defining the Issues in Theory, Practice and Policy*, pp. 3–16. Oxford: Oxford University Press.

Craig MC, Catani M, Deeley Q, et al. (2009) Altered connections on the road to psychopathy. *Molecular Psychiatry* 14: 946–953.

Dennett D (2003) *Freedom Evolves*. Baltimore, MD: Penguin Books.

DeOlivera-Souza R, Igna´cio FA, Moll J, and Hare RD (2008) Psychopathy in a civil psychiatric outpatient sample. *Criminal Justice and Behavior* 35(4): 427–437.

Douglas, T. (2008). Moral enhancement. *Journal of Applied Philosophy*, 25(3), 228-245.

Gazzaniga M (2008) The law and neuroscience. *Neuron* 60: 412–415.

Glannon W (ed.) (2007) Defining Right and Wrong in Brain Science: Essential Readings in Neuroethics. *The Dana Foundation Series on Neuroethics*. New York: Dana Press.

Greene J and Haidt J (2002) How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6(12): 517–523.

Illes J (ed.) (2006) *Neuroethics: Defining the Issues in Theory, Practice and Policy*. Oxford: Oxford University Press.

Libet B (2002) The timing of mental events: Libet’s experimental findings and their Implications. *Consciousness and Cognition* 11(2): 291–299.

Morse SJ (2006) *Moral and legal responsibility and the new neuroscience*. In: Illes J (ed.) Neuroethics: Defining the Issues in Theory Practice and Policy, pp. 33–50. Oxford: Oxford University Press.

Raine A and Yang Y (2006) Neural foundations to moral reasoning and antisocial behavior. *SCAN* 1: 203–213.

Roskies AL (2002) Neuroethics for the new millenium. *Neuron* 35: 21–23.

Roskies AL (2006) A case study of neuroethics: The nature of moral judgment. In: Illes J (ed.) *Neuroethics: Defining Issues in Theory, Practice and Policy*, pp. 17–32. Oxford: Oxford University Press.

Roskies, A. (2011). Why Libet studies don’t pose a threat to free will. In L. Nadel & W. Sinnott-Armstrong (eds.), *Conscious Will and Responsibility. A Tribute to Benjamin Libet*. Oxford University Press 11—22.

Silva G (2006) Neuroscience nanotechnology: Progress, opportunities and challenges. *Nature Neuroscience* 7: 65–74.

Smart J (1959) Sensations and brain processes. *Philosophical Review* 66: 141–156.

Smith KL (1998) Making pedophiles take their medicine: California’s chemical castration law. *The Buffalo Public Interest Law Journal* 17: 1–42.

Stace WT (1952) The Problem of Free Will. In *Religion and the Modern Mind*, pp. 248–258. New York: Harper Collins.

**Relevant Websites**

http://www.neuroethics.upenn.edu/ – The Center for Neuroscience and Society at the

University of Pennsylvania.

http://www.lawandneuroscienceproject.org/ – The MacArthur Foundation’s Law and

Neuroscience Project.

http://neuroethics.ubc.ca/National\_Core\_for\_Neuroethics/Home.html – The National

Core for Neuroethics at the University of British Columbia.

http://www.neuroethicssociety.org/ – The Neuroethics Society.

<http://link.springer.com/journal/12152> - The journal Neuroethics