DRAFT COPY

DO NOT DISTRIBUTE

Current Trends in Cognitive Neuroscience and Criminal Punishment

Corey H. Allen and Eyal Aharoni[1]

Abstract: This chapter explores current trends in cognitive neuroscience research and its implications for criminal punishment. It considers what this research can (or cannot) tell us about an offender's future dangerousness as well as how brain function, and folk *perceptions* of brain function, might inform punishers' sentencing judgments. Finally, this chapter suggests some future directions for this complex and important research field.

1. Introduction

Recently, while attending a seminar on neuroscience, philosophy, and the law, one of us received an unexpected phone call: a childhood friend had shot and killed his own parents in their sleep before turning the gun on himself. When tragic events like this occur, a barrage of questions rapidly comes to mind: Was he a victim of abuse? Was he suffering from mental illness? Had he been showing signs of distress? Was he under the influence? Did he keep a diary? What the hell was he *thinking*? In such situations, we consistently ruminate on roughly the same short list of questions, as if to apprehend the contents of the perpetrator's mental state – a desperate instinct, perhaps, to form an attitude about his internal moral culpability.

The criminal law works in much the same way, but with far-reaching consequences. Beneath a complex lattice of policies and procedures is a system attempting to deliver punishment to all those – and only those – who deserve it. And to make that determination satisfactorily, the system tasks itself to somehow assess and build a representation of the defendant's mental environment. In this way, criminal law is fundamentally psychological.

To inform the question of the defendant’s mental states, trial courts have traditionally relied on subjective and qualitative judgments. Today, lawyers are increasingly querying evidence from the brain sciences (Farahany 2015). Since the mind, after all, is a product of the brain, it is reasonable to ask whether the brain contains clues about the offender’s mental health, mental capacities, and perhaps even his past intentions or future risk of reoffending. But to date, judicial opinions have not always been favorable to neuroscience evidence, which has often failed to demonstrate a legally compelling link between brain, mind, and crime (Farahany 2015[BIB-033]). According to its critics, the physicalist perspective taken by neuroscience is unlikely to tell us anything that couldn’t already be discerned using a traditional, folk psychological observation of the offender’s behavior (Morse 2016[BIB-062]).

The legal and ethical standards for evaluating neuroscientific evidence and why this evidence has often fallen short have been examined elsewhere (Brown & Murphy 2010; Buckholtz & Faigman 2014; Glannon 2011 Pardo & Patterson 2010[BIB-066]). Much of this scholarship has focused on questions bearing on the defendant’s mental health in the guilt phase of a trial, such as evaluations of legal insanity and competency to stand trial (Aharoni et al. 2008[BIB-001]; Edersheim, Brendel & Price 2012[BIB-030]; Meixner 2016. Less attention has been paid to questions arising at the sentencing phase, such as questions of the defendant’s dangerousness. Even less attention has been paid to insights to be gleaned, not just from the defendant’s brain, but from those of other players in criminal procedures, such as jurors and judges who evaluate facts and make punishment recommendations or rulings. Yet, as the brain sciences mature, it is important to anticipate the potential uses and misuses of the neuroscience evidence for such purposes.

The relevance of the brain sciences to punishment decisions can be examined from various perspectives including scientific, ethical, and legal ones. The ethical perspective asks questions like: when, if ever, is it justified to use information about a defendant’s brain in sentencing decisions? What constitutes a responsible use of neurotechnology in the court? The legal perspective raises questions about whether neurobiological evidence meets standards of evidence and admissibility, and whether we need to change these standards in light of this advancing technology. The scientific perspective includes questions like: can an understanding of the *defendant's* brain function inform questions of his dangerousness? Can sentencing decisions in any way be informed by the *punisher's* brain? And when making judgments about a defendant's responsibility, what role is played by *lay* perceptions of brain evidence? Since the ethical and legal questions surrounding brain evidence in the court depend heavily on the state of the science, the focus of this chapter will be the scientific questions.

In this chapter, we review current research trends in neuroscience research bearing on defendant dangerousness as well as how punishment judgments are made. We consider their methodological strengths and limitations, and we touch upon several ways in which these research developments may, and in some cases, may not, be relevant to criminal sentencing decisions. We conclude that while the results of current neuroscience research and application are still very inferentially limited, their potential is growing rapidly, and so efforts aiming to characterize the neural mechanisms underlying antisocial actions and punitive judgments should be taken seriously as potentially relevant contributors to our understanding of human behavior, both virtuous and vicious. Ultimately, a deeper understanding of the relationship between mind, brain, and behavior will be essential for the development of legal procedures that effectively balance the public's safety against defendants' civil liberties.

2. Can an Understanding of the Defendant's Brain Function Inform Questions of Future Dangerousness?

Is there any structure or operation of the brain whose measurement can reveal something useful about the probability that a criminal defendant will cause harm in the future? An answer to this question could be useful not just for traditional sentencing decisions but also for indictment, plea bargaining, diversion, and rehabilitation decisions. Unfortunately, the question is fundamentally flawed. It's flawed because dangerousness is a broad, normative attribution that people make based on a complex relationship between a variety of different causal factors, many of which operate in social environments outside the mind, few if any of which are necessary or sufficient for evoking the dangerous outcome. In short, there is no "danger" part of the brain (Specker et al. 2018[BIB-079]).

To break down this question into more useful components, at minimum, it would be necessary to operationally define some specific types of dangerousness, for instance, the probability that a domestic abuser will commit a new act of domestic violence within some specific period of time, holding constant various situational/contextual factors. But a narrow definition of dangerousness is not enough. It is also prudent to have a theory about which known properties of human cognition are likely to place that abuser at elevated risk of a new offense. Alas, the field of cognitive neuroscience does not have a comprehensive theory of all the cognitive mechanisms likely to play a role in every type of reoffense. Nonetheless, we can look to the existing literature to see how scholars have approached the question, and then consider some of their strengths and limitations. We can group their approaches into several methodological categories: correlation, quasi-experimentation, retrodiction, prediction, and stimulation approaches. We discuss examples of each approach and how it might contribute to discourse on dangerousness.

2.1 Neuroimaging Approaches to Questions of Dangerousness: Correlational Methods

Correlational methods are widely used for getting a lay of the neurological land. In a typical correlational approach, neurotechnologies such as magnetic resonance imaging (MRI) and electroencephalography (EEG) serve as passive recording devices that help scientists measure associations between brain states and behaviors. Importantly, these technologies do not actively manipulate those brain states, so they cannot independently demonstrate that such states cause a particular behavior. Even so, when collected during an experimental task, such techniques are valuable for the development of hypotheses about neural mechanisms of a particular behavior, in large part, because they help to reduce the number of candidate causal factors to a manageable list.

Many of the correlational neuroimaging studies to date have focused on populations with elevated risk such as incarcerated offenders and those with psychopathic personality traits. The psychopathic personality is a research classification that includes a cluster of traits, such as impulsivity, manipulativeness, and lack of empathy, which together predict antisocial outcomes such as reoffending (Tengström et al. 2000[BIB-084]). Forensic and psychopathic populations can be particularly informative for the study of the brain's contribution to dangerous behavior.

In one study, researchers examined the neural circuitry underlying reward-related decision making in a sample of 49 incarcerated offenders who varied in psychopathic personality traits (Hosking et al. 2017[BIB-045]). In theory, some criminal offenses (i.e., impulse crimes) could result from a tendency to overvalue immediate rewards relative to delayed rewards, without regard for future consequences (e.g., the probability of arrest). The researchers found that when making such valuation trade-offs, participants higher in psychopathic personality traits exhibited greater activity in the nucleus accumbens (NAcc), an area known to underlie the processing of reward. The NAcc in these participants also showed less functional connectivity to the ventromedial prefrontal cortex (vmPFC), a region involved in regulation of the NAcc, representation of delayed value outcomes, and moral decision making (Economides et al. 2015[BIB-029]; Raine & Yang 2006[BIB-071]). Interestingly, NAcc-vmPFC functional connectivity was negatively correlated with criminal convictions.

These findings suggest that dysregulation of the neural circuitry underlying cost/benefit valuation may play an important role in the proclivity to favor short-term over long-term rewards, which in turn may contribute to impulsive and perhaps, criminal decision making.

Another potential contributor to criminal behavior is the failure to recognize moral concepts or violations. In one test of this theory, Fede and colleagues (2016[BIB-034]) employed functional MRI to examine the neural activity underlying the classification of words or phrases with moral content related to being wrong (e.g., murder, lying, stealing), not wrong (e.g., charity, kindness, saving lives), or ambiguous (e.g., animal testing, gun, control, prostitution) in nearly 250 criminal offenders who varied in psychopathic personality traits. Individuals high in psychopathy (incorrectly) rated a larger portion of the moral stimuli as being “not wrong”. Furthermore, during presentation of morally “wrong” stimuli, these higher psychopathy individuals showed less activity than their low-psychopathy counterparts in the anterior cingulate cortex (ACC), a region found to be involved in the detection of norm-violations (Güroğlu et al. 2010). While processing morally ambiguous stimuli, those high in psychopathy also showed less activity than those low in psychopathy in the temporal-parietal junction (TPJ), a region previously implicated in the processing of ambiguous moral stimuli in healthy populations (Schaich Borg et al. 2011[BIB-076]), and the dorsolateral prefrontal cortex (dlPFC), a region engaged when exerting cognitive control over emotions (Tassy et al. 2009[BIB-083]).

Some evidence suggests that brain differences associated with psychopathy may even be observable at the structural level. Using a structural, correlational technique known as voxel-based morphometry, one study examined gray matter differences in nearly 300 incarcerated men (Ermer et al. 2012[BIB-031]). The researchers reported that participants higher in psychopathy exhibited decreased gray matter in limbic, paralimbic, and orbitofrontal areas. These differences accounted for 20% of the variance in psychopathy scores and converge with previous research (e.g., Kiehl 2006[BIB-048]; Kiehl et al. 2001; Walton, Devlin & Rushworth 2004[BIB-086]).

Studies of neurobiological differences among former offenders have revealed distinct patterns underlying the processing of reward and moral norms. And these findings cohere with existing models about the cognitive functions typically associated with these networks. As noted, however, because of their correlational design, they cannot, by themselves, demonstrate that the observed brain abnormalities cause people to behave in risky ways. Instead, these effects could represent indicators of some other causal process or could even be an effect of the high-risk behavior itself. In the next subsections, we review studies that attempt to draw causal links between brain function and potentially dangerous behavior.

2.2 Neuroimaging Approaches to Questions of Dangerousness: Quasi-Experimental and Retrodictive methods

One method of uncovering causal links between brain and behavior is quasi-experimental design. Quasi-experimental designs permit researchers to rely on pre-defined groups when experimental manipulation is not feasible, such as the examination of the adverse consequences of pre-existing brain lesions.

Schofield and colleagues (2015[BIB-077]) used this method to determine whether there was an association between criminal convictions and pre-existing brain lesions. The researchers retrospectively analyzed criminal and medical records spanning over 30 years on nearly 8,000 individuals with brain lesions, and compared these records to sibling and general population matched control data. The investigators reported that the presence of a brain lesion was associated with increased conviction rates (for any crime as well as violent crimes specifically) in both men and women relative to the general population, and that these relationships largely held even when limiting the analysis to those with sibling controls ($n \approx 2{,}400$). The researchers interpreted these findings as evidence that brain lesions (of any sort) may increase behavioral dysregulation, aggression, or impulsivity, which may, in turn, increase the risk of criminal offending. One known limitation of quasi-experimental methods is that the pre-defined groups may vary in a variety of unknown ways, and Schofield and colleagues acknowledged this inability to rule out the influence of a third variable: a particular phenotype, perhaps, that could place some individuals at risk of both criminality and impulsive, aggressive behavior.

Another method of uncovering causal links between brain and behavior is retrospective prediction, or "retrodiction". Retrodiction refers to the retrospective assessment of a pre-existing explanatory factor given a known outcome. In a similar fashion, Darby and colleagues (2018[BIB-024]) evaluated brain scans of convicted offenders with pre-existing lesions in order to understand how these lesions might precipitate changes in the functional connectivity (an analysis known as lesion network mapping) between regions commonly engaged in moral decision-making tasks. The researchers reported two main findings: First, lesions that were temporally associated with criminal activity were confined to a single brain network (consisting primarily of the inferior orbitofrontal cortex, anterior temporal lobes, vmPFC, NAcc, intraparietal sulcus, and dlPFC), which was distinct and separate from networks affected by lesions that were not temporally connected to criminal activity. Second, the locations of these lesions were functionally connected to areas engaged in moral decision making, value-based decision making, and theory of mind tasks, domains that are potentially relevant to decisions to commit a crime. These results raise the possibility that this unique pattern of brain connectivity can help to explain why some individuals with lesions are likely to commit crimes whereas others are not, while at the same time potentially demarcating candidate causal factors shaping criminal activity.

These studies suggest potential candidates for causal mechanisms in the development of antisocial behavior. However, we caution against over-interpretation because quasi-experimental and retrodictive methods carry risks of misattributing the source or overestimating the strength of their effects, largely due to an inability to remove and control for confounding variables during the sampling process. One way to reduce these risks is using prospective methods.

2.3 Neuroimaging Approaches to Questions of Dangerousness: Prospective Prediction

Prospective prediction is achieved by tracking changes in the outcome variable over time and employing an a priori hypothesis to test whether information at Time 1 predicts outcomes at Time 2. It asks: of a sample of people with a varying attribute, what proportion will later exhibit a particular outcome? Evidence of a time-dependent relationship between a potentially explanatory factor and outcome increase confidence that the relationship is a causal one. Prospective prediction has been validated in the context of violence risk assessment using structured, actuarial instruments that measure a variety of person-level characteristics, including behavioral ones (e.g., Monahan 2008; Nicholls, Ogloff & Douglas 2004[BIB-064]). These instruments rely on normative samples – large groups of former offenders that have been observed over time – to serve as a basis to estimate a new individual's prospects by comparing the two along predefined dimensions. The accuracy with which these classification tools predict individual violent reoffending has been shown to significantly exceed that of traditional unstructured risk assessment methods (Grove & Meehl 1996[BIB-041]; Monahan 1981).

The notion that structured risk assessment could be further improved by functional brain metrics seems ambitious. But a growing line of research in "neuroprediction" lends credence to this possibility. Indeed, there are strong empirical reasons to expect the existence of substantial causal links between brain function and violent behavior (see Liu 2011[BIB-056]). For example, prospective studies have demonstrated that early lead exposure can increase risk of violence by affecting critical brain networks. In one longitudinal study, Wright and colleagues (2008[BIB-090]) measured blood lead levels in pregnant women and their children throughout the first six years of life. Years later, the investigators tracked the children's violent behavior using criminal arrest records and found a predictive relationship between early lead concentrations and arrests for violent crimes, even after controlling for other contributing factors. In a separate longitudinal study, investigators found a significant association between early childhood lead exposure and brain volume (Cecil et al. 2008[BIB-019]), specifically gray matter loss in the ACC, a region formerly associated with the ability to control aggressive impulses (Devinsky, Morrell & Vogt 1995[BIB-027]). The evidence for the damaging effects of lead exposure is strong enough that it has been used to argue that it should qualify as a potentially mitigating factor in criminal sentencing decisions (Kittilstad 2018[BIB-051]).

In another longitudinal study, Pardini and colleagues administered a battery of psychological assessments to young boys, and then assessed their brain volumes, violence and delinquency approximately 20 and 23 years later (Pardini et al. 2014[BIB-065]). Using this design, the researchers were able to correlate brain volume with violence and aggression both retrospectively and prospectively. Retrospectively, they found that lower amygdala volumes were associated with more aggression and psychopathic tendencies in childhood and early adulthood. Prospectively, lower amygdala volumes were associated with increased risk for aggression, violence, and psychopathic traits. Their findings are consistent with previous research (Bobes et al. 2013[BIB-012]), including evidence that the amygdala plays a crucial role in emotional processing and the development of antisocial traits (Phelps & LeDoux 2005[BIB-069]).

A growing number of prospective prediction studies have demonstrated the utility of neurocognitive models in the prediction of antisocial outcomes including substance abuse relapse (Janes et al. 2010[BIB-046]; Paulus, Tapert & Schuckit 2005), treatment completion (Fink et al. 2016[BIB-035]; Steele et al. 2014[BIB-080]; Steele et al. 2018[BIB-081]), and rearrest (Aharoni et al. 2013[BIB-002]; Aharoni et al. 2014[BIB-003]; Delfin et al. 2019[BIB-026]; Kiehl 2018[BIB-049]; Steele et al. 2015[BIB-082]). Importantly, evidence of prospective statistical relationships between brain and behavior do not necessarily imply that the relationship is strong enough to be used as biomarkers for diagnostic uses in individual cases. High diagnostic specificity would be useful for the provision of tailored treatment to individuals at risk, and arguably for legal decisions such as involuntary civil commitment, but it requires that these models achieve classification accuracy levels determined by some pre-existing normative standard. In legal contexts, this standard is often likely to be prohibitively high given the highly multivariate nature of criminal behavior. At least one group of prospective neuroimaging studies, however, has evaluated the classification accuracy of a neurocognitive model in predicting antisocial outcomes. In a sample of 96 offenders, the investigators found that activity within the ACC prospectively predicted rearrest up to three years following release. Furthermore, this neurocognitive model correctly classified rearrest at levels that rival validated actuarial risk tools (Aharoni et al. 2013[BIB-002]; Aharoni et al. 2014[BIB-003]; Steele et al. 2015[BIB-082]).

Although these results provide a proof of concept about the potential of neuroprediction models, much work remains to be done to demonstrate whether such effects are reliable and specific enough for legal and treatment applications. For example, as others have noted (e.g., Poldrack et al. 2018[BIB-070]), the gold standard for estimating the predictive utility of a model is to test it on a second, independent sample – a technique known as out-of-sample validation. Prospective imaging studies have only recently risen to this challenge. For instance, a recent study by Kiehl and colleagues (2018) successfully used a structural model of brain age to predict rearrest in an independent sample. Their model implicated many of the same regions reported in previous studies, including the ACC and amygdala.

The growing body of research on neuroprediction suggests that the function and structure of the brain plays an instrumental role in the expression of many harmful behaviors. So, while predictive modeling is unlikely to become accurate enough for use in criminal sanctioning decisions, efforts to prospectively predict risk outcomes in lower stakes domains, such as the provision of treatment resources to those most likely to benefit from it, may be scientifically meaningful and empirically tractable (see also Faigman et al. 2013[BIB-032]; Heilbrun 2009).

2.4 Neuroimaging Approaches to Questions of Dangerousness: Brain Stimulation

Another promising method of building causal, potentially predictive, models of antisocial behavior is the use of brain stimulation. Depending on the stimulation technique used, researchers can increase or decrease the excitability of areas in the brain, through the targeted application of positively or negatively charged ions. This approach has been used to elucidate neural contributors to antisocial behavior as well as ways to regulate such behavior.

For example, Gilam and colleagues (2018) employed concurrent transcranial direct current stimulation (tDCS) and fMRI to test whether facilitation of vmPFC activity aids in anger regulation – as measured by the acceptance of unfair offers in an economic game – and aggressive behavior in subsequent reactive aggression tasks. As expected, active stimulation to the vmPFC during the economic task led to a larger acceptance rate of unfair offers and a decrease in self-reported anger following the task, yet there were no differences between active and sham stimulation in reactive aggression. Similarly, there was increased vmPFC activity during the processing of unfair offers for those undergoing active stimulation. These results obtained by combining correlational and causational methodologies suggest a causal role for vmPFC functionality in the expression and regulation of anger.

Similarly, Choy and colleagues (2018[BIB-021]) utilized tDCS to test whether a brain area known as the dlPFC plays a causal role in the ability to inhibit aggressive behavior (Anderson et al. 1999[BIB-005]). After receiving active or sham stimulation, the participants (members of the general public) read hypothetical vignettes involving a physical and sexual assault and rated how likely they were to behave as the criminal protagonist in the scenario had behaved and how morally wrong that behavior was. A secondary task was also administered to measure implicit aggression behaviorally. On a computer workstation, participants were given an opportunity to stick virtual needles into an image of a doll that represented someone for whom the participant harbored ill will (as validated by Dewall et al. 2013[BIB-028]). Consistent with their hypotheses, the researchers found that the active stimulation group reported significantly lower likelihood of committing physical and sexual assault compared to the mock stimulation group, and found sexual assault to be more morally wrong than the mock stimulation group. However, stimulation of the dlPFC did not change the rate of aggression as measured in the behavioral task. These findings support the hypothesis that the dlPFC functions to modulate wrongfulness judgments and the self-reported likelihood of committing such acts, though is not necessarily sufficient to control aggressive behavior.

A study by Dambacher and colleagues (2015[BIB-023]) provides converging evidence for the dlPFC as an inhibitor of aggression. After receiving either active or sham tDCS stimulation, 32 male and female participants completed questionnaires and behavioral tasks designed to assess transient states of reactive and proactive aggression. The researchers found that active stimulation of the dlPFC selectively decreased proactive aggression in men, but had no effect on reactive aggression. These results suggest a potential gender difference in receptivity to clinical interventions and also suggest a neural dissociation between different types of aggression.

In another stimulation study (Riva et al. 2015[BIB-073]), researchers sought to test the function of the right ventrolateral prefrontal cortex (rvlPFC), a region believed to be involved in the regulation of aggressive impulses (Aron & Poldrack 2005[BIB-006]). In their experiment, 79 participants played a violent or non-violent video game while receiving active or sham stimulation over the rvlPFC. They then completed a task designed to assess both provoked and unprovoked aggression. The task pits participants against each other in a competitive reaction-time game, one that they are able to deliver provoked and unprovoked aversive noises to the opponent. For participants receiving sham stimulation, the experimenters found that unprovoked aggression was higher in those assigned to play violent (as opposed to nonviolent) games, as might be expected. Importantly, active stimulation reduced unprovoked aggression such that there were no group differences according to game type. These results suggest an important role for the rvlPFC in mediating psychologically aggravated levels of unprovoked aggression.

One common characteristic of these studies is the tendency to sample from the general public, whose base rate of dangerous, criminal behavior is fairly low. This limits the ability to generalize to those with elevated risk of criminal behavior. A further limitation is the ecological validity of laboratory aggression measures. Since experimental manipulation of criminal risk is ethically problematic, this increases experimenters' reliance on more distal proxies for real-world aggression. This heavy reliance on distal proxies for aggressive behavior may help to explain the apparent disconnect in the evidence identifying neural contributors to outward aggressive behavior versus mental states theorized to motivate that behavior (such as anger). Though both types of evidence have been demonstrated, just how these mental states give rise to aggressive and criminal behavior remains relatively unknown.

Despite these limitations, these early advancements will help future investigators develop more refined hypotheses that enable them to selectively target the neural circuitry that most facilitates aggressive behavior. Such efforts would not necessarily have direct implications for legal procedure, but could potentially yield therapeutic insights, which in turn, may become relevant during legal processes such as forensic evaluation, adjudication, and sentencing.

3. How Can Sentencing Decisions Be Informed by the Punisher's Brain Function?

The intersection of law and neuroscience has focused primarily on the brains of criminal offenders. Much less attention has been paid to the role of those called upon to assign blame and to make punishment judgments or recommendations, such as judges and jurors. Although specific legal applications of this research are still ill-defined, investigating the punisher's brain could help to uncover some of the proximate mechanisms shaping typical punishment decisions. A small set of experiments provide converging evidence that typical punishment decisions might be subserved by a host of dynamic brain networks. We discuss what these studies suggest about typical punishment decision making in the brain including the effects of both endogenous factors (e.g., attributions of the defendant's mental state at the time of the crime) and exogenous factors (e.g., contextual cues) in this process.

In an early attempt to elucidate the neural mechanisms of third-party punishment, Buckholtz and colleagues (2008[BIB-015]) conducted an event-related fMRI experiment investigating brain areas that are selectively associated with ascriptions of criminal responsibility and punishment. In the experiment, participants made punishment judgments after reading about different protagonists performing criminal and non-criminal actions with varying degrees of mitigating circumstances (such as a lack of requisite knowledge of the risk of harm). For the crimes with mitigating circumstances, participants recommended smaller punishments, as to be expected. During these trials, the rTPJ was engaged, an area commonly attributed to the process of mentalizing others.[2] During trials lacking mitigating circumstances, and trials in which participants chose to punish, the rdlPFC, an area found to play an integral role in norm enforcement, was uniquely engaged (Sanfey et al. 2003[BIB-075]). The investigators attributed the activation of the rTPJ to the period when participants made judgments about the offender's level of responsibility for the act, and the rdlPFC, in contrast, was attributed to the actual decision to punish.

Buckholtz and colleagues (2008[BIB-015]) also observed engagement of the right amygdala during trials lacking mitigating circumstances, an area credited as a hub for social and affective processing generally (Murray, Brosch & Sander 2014[BIB-063]). This engagement persisted even when the investigators controlled for the emotional arousal inherent to the scenarios. This signal was positively correlated with the magnitude of punishment recommended, suggesting that the amygdala might play a distinct role in the formation of punishment judgments beyond the processing of emotional arousal, such as the processing of harm, as proposed by two of the authors in a later commentary on this experiment (Buckholtz & Marois 2012).

Similarly, Yamada and colleagues (2012[BIB-091]) varied whether crime vignettes included mitigating circumstances intended to evoke sympathy (e.g., the offender felt pity for his terminally sick wife and thus killed her) or lacked mitigating circumstances. The investigators then solicited sympathy ratings and punishment recommendations in an fMRI setting. As expected, participants reported more sympathy and less punishment towards the offender for crimes committed with mitigating circumstances. The investigators found three brain regions that were associated with sympathy ratings and the mitigation of punishment: the TPJ, precuneus, and dorsomedial prefrontal cortex (dmPFC) – all regions previously identified to be involved in mentalizing and moral conflict (Yamada et al. 2012[BIB-091]).

Ginther and colleagues (2016[BIB-037]) sought to characterize the neural processes underlying evaluations of harm and criminal intent. They systematically varied these constructs and solicited punishment recommendations in an fMRI setting. They also measured self-reported difficulty of the decision. Attributions of harm were negatively associated with activity in the right lateral prefrontal cortex (rlPFC), and difficulty attributing harm was associated with the orbitofrontal cortex (OFC). These results align with previous literature suggesting that the rlPFC is commonly involved in the processing of other's pain, and the OFC in cost benefit analysis (Janowski, Camerer & Rangel, 2013[BIB-047]; Lamm, Decety & Singer, 2011[BIB-055]). Attributions of intent were positively associated with activation of the posterior cingulate cortex (PCC), which the authors suggest is indicative of the negative valence associated with increasing culpability (Maddock, Garrett & Buonocore, 2003[BIB-057]). Increasing difficulty of attributing intent was associated with activity within the TPJ, dmPFC, and superior temporal sulcus, a network active during mentalization (Ginther et al. 2016[BIB-037]). This pattern of results suggests separable mechanisms for assessing the magnitude of these concepts, and the difficulty with which they are conceived.

The researchers were also interested in how information about the crime (such as the level of harm and intent) interact to produce a punishment judgment. Following participants' initial appraisal of harm and intent, they observed sequential activation of the bilateral amygdalae and right dlPFC which anticipated the judgment, similar to previous research by Buckholtz and colleagues (2008[BIB-015]).

In a similar study, researchers sought to understand how participants respond to conflicting instances of sentencing decisions: instances where acts committed were deemed morally right but legally wrong (Yang et al. 2019). Perhaps unsurprisingly, moral acceptability was correlated with punishment decisions, and crimes of good intention were punished less so than those of bad intention. Consistent with previous research, activation of the rTPJ was found to be higher in the evaluation of well-intentioned acts compared to those of bad intentions. Furthermore, crimes of good intent elicited greater activation in the dlPFC, and correspondingly, a greater connectivity from the dlPFC to the mPFC. In both previous experiments, the researchers interpret this pattern as evidence that these brain regions work in concert to integrate relevant information, such as the level of harm and intent, and then select a level of punishment scaled to that integrated information.

Other studies have focused on contextual factors likely to influence attributions of responsibility and punishment. Capestany and Harris (2014[BIB-018]), for example, examined punitive responses to emotionally evocative language and the evidence presented about the defendant's character. Using hypothetical crime scenarios, the researchers systematically varied disgust evoking language of the crime (low vs. high disgust) and the method by which the defendant's personality was clinically assessed (behavioral vs. neurobiological). As expected, behavioral explanations of personality yielded greater ascriptions of responsibility than neurobiological explanations. Furthermore, disgust language and behavioral explanations of personality yielded more severe punishments than descriptions lacking disgust language and neurobiological personality assessments. During responsibility judgments, the investigators observed increased activation in the lingual gyrus and PCC when the defendant's personality was assessed behaviorally. These regions have been shown to be involved in logical deduction and emotional salience, respectively (Barack, Chang & Platt, 2017[BIB-008]; Goel et al. 2004[BIB-039]). During the actual decision to punish, the superior frontal gyrus and the superior parietal lobule were increasingly engaged in the presence of disgust language. These regions have been shown to be involved in executive function and the manipulation of working memory (Boisgueheneuc et al. 2006[BIB-013]; Koenigs et al. 2009[BIB-052]). Engagement of the amygdala, inferior parietal lobule, middle frontal gyrus, and insula were associated with the magnitude of punishment recommended – all regions associated with affective processing (Capestany & Harris 2014[BIB-018]).

Using a similar approach, Treadway and colleagues (2014[BIB-085]) examined the neural mechanisms underlying attributions of criminal intent and harm with and without the use of graphic language to describe the crime, as well as those underlying the decision to punish or not. Similar to the findings of Buckholtz and colleagues (2008[BIB-015]), when scenarios lacked mitigating circumstances and when participants chose to punish, the researchers found that the dlPFC was more engaged when the crime was intentional compared to unintentional, as well as when participants chose to punish the offender compared to when they chose not to do so. For intentional crimes, the left amygdala was more engaged when the crime was described using graphic language rather than plain language, and, correspondingly, participants selected harsher punishments in this condition, suggesting that the amygdala's role in emotional arousal may also operation in punitive decision making. A similar pattern emerged for level of harm: the amygdala was engaged increasingly with level of harm, yet only when the crime was intentional.

Despite their strengths, traditional fMRI study designs do not permit the inference that a particular cognitive function is directly caused by a given neural pattern. In attempt to overcome this limitation, Treadway and colleagues (2014[BIB-085]) also utilized the post hoc statistical technique of Granger Causality Modeling (GCM). GCM is designed to determine whether certain fMRI signals from particular brain regions explain variance in others, allowing the researchers to create a temporally defined map of which signals give rise to others in the brain. They found that the functional connectivity between the amygdala and dlPFC was strengthened in the presence of graphic language, but *only* when the crime was intentional. In cases where the offender committed a crime unintentionally, there was greater functional connectivity from the lTPJ to the dorsal ACC (dACC), which has been previously implicated in the detection of norm violations (Güroğlu et al. 2010), as well as from the dACC to the amygdala. Together, these results suggest that in cases of intentional crimes, graphic language can serve to aggravate punishment recommendations via the functional connectivity from the left amygdala to the dlPFC, yet this aggravating effect can be down-regulated by a top-down connection from the lTPJ-dACC pathway to the amygdala, consistent with motivations to withhold punishment from less-blameworthy offenders.

Another attempt to elucidate the causal mechanisms of punishment decisions was made by Bellucci and colleagues (2017[BIB-009]). While participants were deciding how much to punish a defendant for a hypothetical crime (compared to a control task of estimating the number of syllables in each vignette), several regions were found to be engaged within what they describe as the mentalizing network – including the vmPFC, dmPFC, PCC, left temporal pole, and the lTPJ – and the central executive network, particularly, the left dlPFC. In a GCM analysis, the investigators identified two regions (the left dmPFC and left temporal pole) that sent output to every other region reportedly involved in their punishment task. Further, the dmPFC was found to receive input only from the temporal pole, suggesting that it may serve as a communication hub in the mentalizing network. The investigators interpreted this pattern as evidence that the dmPFC plays an important causal role in the integration of all relevant signals in third-party punishment decisions. More generally, these results suggest a broad and complex network of brain regions underlying third party punishment.

Across studies on the punisher's brain, we find some consistency among brain regions during punishment judgments, including activation of the TPJ, dmPFC, PCC, amygdala, temporal pole, and dlPFC. These regions have been formerly implicated in cognitive functions including theory of mind, affective processes, and higher order cognitive functions in this process (Bellucci et al. 2016; Buckholtz et al. 2008[BIB-015]; Capestany & Harris 2014[BIB-018]; Ginther et al. 2016[BIB-037]; Treadway et al. 2014[BIB-085]; Yamada et al. 2012[BIB-091]; Yang et al. 2019; Young et al. 2010), and it would not be surprising if such cognitive functions are recruited in the formation of typical punishment judgments. Future research would benefit from going beyond study designs optimized to localize cognitive functions in the brain and directly test how these functions unfold in terms of stages of processing. Understanding the stages of cognitive processing could potentially reveal, not just where, but how punishment judgments are likely to be made and when they are likely to conform to or depart from normative preferences or expectations. Until the research can move from questions of "where" to how and when, the direct relevance of this research to legal decision making seems limited, at best. However, research on the punisher's brain may at least retain indirect relevance as part of a broader, basic research question about the contributors to punishment decisions.

4. How Can Sentencing Decisions Be Informed by the Punisher's *Perceptions* Of Brain Function?

Another way that neuroscience might influence sentencing decisions is in the way neuroscientific evidence is interpreted by non-experts, such as judges and jurors. A growing body of research suggests that laypeople may place undue confidence in neuroscientific evidence, and this can affect their ascriptions of responsibility and punishment.

In two studies, Weisberg and colleagues (2008[BIB-088], 2015[BIB-089]) found that when scientific explanations were offered for a behavior, laypeople's ability to distinguish between good and bad quality explanations was hindered when irrelevant neuroscientific information was included. More specifically, explanations that included irrelevant neuroscientific information were judged as stronger and more satisfying than those without it. Further research has suggested that particular aspects of neuroscientific arguments, such as brain images in particular, are especially persuasive when it comes to credibility judgements (McCabe & Castel 2008[BIB-058]; but see Farah & Hook 2013; Michael et al. 2013).

Is this "seductive allure" of neuroscientific information likely to impact legal decisions? In a vignette study with US trial court judges, Aspinwall, Brown, and Tabery

(2012[BIB-007]) found that including a neurobiological description of a defendant's mental illness reduced ratings of responsibility and punishment. Similarly, in a study of mock jurors, Greene & Cahill (2012[BIB-040]) found that neuroscientific evidence of the defendant's psychosis reduced death sentence recommendations compared to a diagnosis of psychosis sans neuroscientific evidence. One explanation for these mitigating effects is the perception that the presence of biological causes implies that the defendant had less control over his actions.

Although biological evidence of behavior seems to reduce judgments of a defendant's responsibility, some research has suggested that this perception can be double-edged if that evidence increases perceptions of future dangerousness (Aspinwall et al. 2012[BIB-007]; Berryessa 2017[BIB-010]; Chandler 2015[BIB-020]; Fuss 2016[BIB-036]; Hardcastle & Lamb 2018[BIB-043]). In support of this prediction, McCabe and colleagues (2011) found that inculpatory evidence (that is, evidence suggesting guilt) from ostensible lie detection technologies produced more guilt recommendations when the technology was described as fMRI as opposed to a traditional polygraph, thermal facial imaging, or no evidence at all. Similarly, Saks and colleagues (2014[BIB-074]) found that when the prosecutorial side presented arguments that included neuroimaging evidence, there was an increase in death sentence recommendations.

These studies suggest that perceptions of neurobiological evidence might be consequential for legal proceedings. Yet, the effects across studies have been somewhat inconsistent. And despite the observation of both mitigating and aggravating effects of neuroscientific evidence on sentencing decisions, no single study has observed both patterns side by side, and several studies have reported null results (Blakey & Kremsmayer 2018; LaDuke, Locklair & Heilbrun 2018[BIB-054]; Remmel, Glenn & Cox 2018[BIB-072]; Schweitzer et al. 2011[BIB-078]). These inconsistencies raise questions about the robustness of the effects across different methods. One commonality of these studies is the use of fairly simple punishment measures such as a single prison sentencing scale. Such scales, though, are not equipped to distinguish between distinct motives for punishment such as retribution, rehabilitation, or deterrence. This fact may be important because the theories about aggravation and mitigation imply the interaction of distinct punitive motives.

In support of this interpretation, recent research by Allen and colleagues (2019[BIB-004]) conducted the first quantitative demonstration of the theorized double-edge effect by including multiple measures designed to separately elicit distinct motives for punishment. In their mock trial study, the researchers found that while neurobiological evidence of an impulse control disorder mitigated prison sentence recommendations, it also aggravated involuntary hospitalization recommendations compared to equivalent psychological evidence. The implication is that fact-finders may be unduly influenced by neurobiological framing of the evidence, and that this framing can cut both ways depending on the type of decision at hand.

Further validation of these effects are needed before gleaning confidence in exactly how and under what conditions they are likely to play out in real criminal punishment decisions. Even so, it raises questions regarding how policy makers can manage these effects (see de Kogel & Westgeest 2016[BIB-025], and Meynen 2013[BIB-060] for further discussion). For example, should judges and juries be trained on the interpretation of neurobiological evidence? Should the presentation of neurobiological evidence be accompanied by a statement about its potentially biasing effects? When is neurobiological evidence adequate to stand alone, and when must it be accompanied by behavioral evidence? Further scholarship is needed to explore the answers to such questions, but, as a whole, the growing body of research on how people perceive neuroscience evidence does suggest that such perceptions will be consequential for criminal law practice.

5. Conclusion

In this chapter, we sought to examine current developments in the brain sciences that pertain to risk of reoffense and to punitive decision making. While comprehensive neurobiological models of criminal behavior and punitive decision making remain a distant goal, the existing literature on these subjects suggests important progress. Across studies, several brain regions make repeat appearances. The ACC and amygdala, for example, appear to play consistent roles in the impulsivity and development of criminal behavior (Aharoni et al. 2013[BIB-002]; Aharoni et al. 2014[BIB-003]; Kiehl 2018[BIB-049]; Steele et al. 2015[BIB-082]). Likewise, the TPJ, dmPFC, PCC, amygdala, temporal pole, and dlPFC show some consistency in punishment judgments, suggesting a possible role of theory of mind, affective processes, and higher order cognitive functions in this process (Bellucci et al. 2016; Buckholtz et al. 2008[BIB-015]; Capestany & Harris 2014[BIB-018]; Ginther et al. 2016[BIB-037]; Treadway et al. 2014[BIB-085]; Yamada et al. 2012[BIB-091]; Yang et al. 2019; Young et al. 2010).

Across these subject matter domains, we identified several priorities for future research. First, this research should aspire to move beyond traditional correlational study designs toward methods that permit stronger causal inference between brain functions and particular behavioral outcomes. A key metric for validating these models is the extent to which they make accurate predictions. Accurate predictive models might, in turn, be valuable for the development of tailored treatments for high risk offenders. Second, this research should further examine how disparate cognitive functions are integrated neurobiologically into a final punishment judgment (e.g., Bellucci et al. 2016; Ginther et al. 2016[BIB-037]). To achieve this objective, however, it may be necessary to specify more precise, conceptual cognitive models of crime and punishment. Part of this effort should entail probing the relevant stages of cognitive processing that culminate in a punishment judgment, such as the sequence in which we appraise information about an actor’s intentions, or the harmfulness of the act, and the integration of such information in service of the judgment. Third, further research is needed to address how fact-finders interpret neuroscientific evidence. This effort should show consideration for the plurality of motivations that drive fact-finders punishment recommendations.

As we consider future directions, it is important to acknowledge the narrowness of this chapter's scope. For example, we did not discuss the neurocognitive effects of incarceration on inmates. We also ignored a large literature on cognitive heuristics and biases, as well as gender and racial biases (see Peer & Gamliel 2013[BIB-068] for a review of cognitive heuristics and biases relevant to criminal sentencing, and Ward, Hartley & Tillyer, 2016[BIB-087] for an in depth analysis of potential drivers in gender and racial biases in drug sentencing decisions). Likewise, we were silent on the many contextual and structural factors that undoubtedly shape criminal behavior, sentencing, and associated neurobiological function. Although such topics were beyond this scope of this chapter, their growing contribution to the field of neuroscience and law will bring theoretical clarity and nuance to neurobiological models of criminal and legal decision making and cannot be overstated.

Just how this laboratory evidence may or may not bear on real legal punishment decisions remains an open question. As this research develops, some people may be tempted to stretch the findings for questionable uses, such as to justify a sentencing increase, diagnose judicial bias, or predict the outcome of a particular trial. But, in order to be of any use to trial procedures, neuroscience evidence – like any scientific evidence – must first demonstrate that it will be relevant to the legal question. For many legal questions, neuroscientific explanations do not meet this basic standard (Morse 2011[BIB-061]). Even if they are deemed relevant, they would have to meet levels of specificity and reliability that are not observed in the current scientific literature. Amidst these legal challenges are a variety of unresolved ethical questions that must be addressed in stride with the scientific advances.

Despite the current limitations of the brain sciences in informing specific legal practices, these sciences nonetheless hold promise in addressing humbler goals. Neuroscience methods carry the potential to validate behavioral findings about antisocial behavior and punishment behavior, generate new hypotheses about cognitive mechanisms, and lend incremental utility to models attempting to explain and predict that behavior for clinical rather than punitive purposes (see also Coppola 2018[BIB-022]). Whether or not the contribution of neuroscience will be probative, let alone pragmatic, are open questions. Until then, studying the factors that drive crime and punishment decisions remains an important objective of basic and clinical research.

Acknowledgements

We thank Nathaniel Anderson, Morris Hoffman, Eddy Nahmias, Saskia Verkiel, and the Cooperation, Conflict, and Cognition Lab for helpful comments.

References

BIB-001 Aharoni, E., Funk, C., Sinnott-Armstrong, W. and Gazzaniga, M. (2008) “Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience,” *Annals of the New York Academy of Sciences*, 1124, pp.145–160.

BIB-002 Aharoni, E., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., Gazzaniga, M. S. and Kiehl, K. A. (2013) “Neuroprediction of future rearrest,” *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), pp.6223–6228.

BIB-003 – – et al. (2014) “Predictive accuracy in the neuroprediction of rearrest,” *Social Neuroscience,* 9(4), pp.332–336.

BIB-004 Allen, C. H., Vold, K., Felsen, G., Blumenthal-Barby, J. S. and Aharoni, E. (2019) “Reconciling the opposing effects of neurobiological evidence on criminal sentencing judgments,” *PLOS ONE*, 14(1), p.e0210584.

BIB-005 Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. and Damasio, A. R. (1999) "Impairment of social and moral behavior related to early damage in human prefrontal cortex," *Nature Neuroscience,* 2(11), pp.1032–1037.

BIB-006 Aron, A.R. and Poldrack, R.A. (2005) "The cognitive neuroscience of response inhibition: relevance for genetic research in attention-deficit/hyperactivity disorder," *Biological Psychiatry*, 57(11), pp.1285–1292.

BIB-007 Aspinwall, L.G., Brown, T.R. and Tabery, J. (2012) "The double-edged sword: does biomechanism increase or decrease judges' sentencing of psychopaths?" *Science*, 337(6096), pp.846–849.

BIB-008 Barack, D.L., Chang, S.W.C. and Platt, M.L. (2017) "Posterior Cingulate Neurons Dynamically Signal Decisions to Disengage during Foraging," *Neuron,* 96(2), pp.339–347.e5.

BIB-009 Bellucci, G., Chernyak, S., Hoffman, M., Deshpande, G., Dal Monte, O., Knutson, K. M., Grafman, J. and Krueger, F. (2016) "Effective connectivity of brain regions underlying third-party punishment: Functional MRI and Granger causality evidence," *Social Neuroscience*, 12(2), pp.124–134.

BIB-010 Berryessa, C.M. (2017) "Jury-Eligible Public Attitudes Toward Biological Risk Factors for the Development of Criminal Behavior and Implications for Capital Sentencing," *Criminal Justice and Behavior*, 44(8), pp.1073–1100.

BIB-011 Blakey, R. and Kremsmayer, T.P. (2018) "Unable or Unwilling to Exercise Self-control? The Impact of Neuroscience on Perceptions of Impulsive Offenders," *Frontiers in Psychology*, 8. [online]. Available from: www.frontiersin.org/articles/10.3389/fpsyg.2017.02189/full [Accessed October 10, 2018].

BIB-012 Bobes, M. A., Ostrosky, F., Diaz, K., Romero, C., Borja, K., Santos, Y. and Valdés-Sosa, M. (2013) “Linkage of functional and structural anomalies in the left amygdala of reactive-aggressive men,” *Social Cognitive and Affective Neuroscience,* 8(8), pp.928–936.

BIB-013 Boisgueheneuc, F. D., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Zhang, S. and Dubois, B.(2006). “Functions of the left superior frontal gyrus in humans: a lesion study.” *Brain: A Journal of Neurology*, 129(Pt 12), pp.3315–3328.

BIB-014 Brown, T. and Murphy, E. (2010) “Through a scanner darkly: functional neuroimaging as evidence of a criminal defendant’s past mental states,” *Stanford Law Review,* 62(4), pp.1119–1208.

BIB-015 Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. and Marois, R. (2008) “The neural correlates of third-party punishment,” *Neuron*, 60(5), pp.930–940.

BIB-016 Buckholtz, J. W. and Faigman, D. L.(2014) “Promises, promises for neuroscience and law,” *Current Biology*, 24(18), pp.R861–R867.

BIB-017 Buckholtz, J. W. and Marois, R. (2012) “The roots of modern justice: cognitive and neural foundations of social norms and their enforcement,” *Nature Neuroscience,* 15(5), pp.655–661.

BIB-018 Capestany, B.H. and Harris, L.T. (2014) “Disgust and biological descriptions bias logical reasoning during legal decision-making,” *Social Neuroscience,* 9(3), pp.265–277.

BIB-019 Cecil, K. M., Brubaker, C. J., Adler, C. M., Dietrich, K. N., Altaye, M., Egelhoff, J. C., Wessel, S. and Lanphear, B. P. (2008) “Decreased brain volume in adults with childhood lead exposure,” *PLoS medicine,* 5(5), p.e112.

BIB-020 Chandler, J.A. (2015) “The use of neuroscientific evidence in Canadian criminal proceedings,” *Journal of Law and the Biosciences,* 2(3), pp.550–579.

BIB-021 Choy, O., Raine, A. and Hamilton, R.H. (2018) “Stimulation of the Prefrontal Cortex Reduces Intentions to Commit Aggression: A Randomized, Double-Blind, Placebo-Controlled, Stratified, Parallel-Group Trial,” *Journal of Neuroscience,* pp.3317–17.

BIB-022 Coppola, F. (2018) “Mapping the Brain to Predict Antisocial Behaviour: New Frontiers in Neurocriminology,‘New’ Challenges for Criminal Justice,” *UCL Journal of Law and Jurisprudence – Special Issue*, 1, pp.103–126.

BIB-023 Dambacher, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S. and Sack, A. T. (2015) “Reducing proactive aggression through non-invasive brain stimulation,” *Social Cognitive and Affective Neuroscience*, 10(10), pp.1303–1309.

BIB-024 Darby, R. R., Horn, A., Cushman, F. and Fox, M. D. (2018) “Lesion network localization of criminal behavior,” *Proceedings of the National Academy of Sciences*, 115(3), pp.601–606.

BIB-025 de Kogel, C.H. and Westgeest, E.J.M.C., (2016) “Neuroscientific and behavioral genetic information in criminal cases in the Netherlands,” *Journal of Law and the Biosciences,* 2(3), pp.580–605.

BIB-026 Delfin, C., Krona, H., Andiné, P., Ryding, E., Wallinius, M. and Hofvander, B., (2019) “Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data,” *PLOS ONE*, *14*(5), p.e0217127.

BIB-027 Devinsky, O., Morrell, M.J. and Vogt, B.A. (1995) “Contributions of anterior cingulate cortex to behavior,” *Brain: A Journal of Neurology,* 118 (Pt 1), pp.279–306.

BIB-028 DeWall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V., Pond Jr, R. S., Renzetti, C. M. and Fincham, F. D. (2013) “The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations,” *Aggressive Behavior*, 39(6), pp.419–439.

BIB-029 Economides, M., Guitart-Masip, M., Kurth-Nelson, Z. and Dolan, R. J. (2015) “Arbitration between controlled and impulsive choices,” *Neuroimage*, 109, pp.206–216.

BIB-030 Edersheim, J.G., Brendel, R.W. and Price, B.H. (2012) “Neuroimaging, Diminished Capacity and Mitigation,” In Neuroimaging in Forensic Psychiatry. Wiley-Blackwell, pp. 163–193. [online]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119968900.ch10 [Accessed November 5, 2018].

BIB-031 Ermer, E., Cope, L. M., Nyalakanti, P. K., Calhoun, V. D. and Kiehl, K. A. (2012) “Aberrant paralimbic gray matter in criminal psychopathy,” *Journal of Abnormal Psychology,* 121(3), pp.649–658.

BIB-032 Faigman, D. L., Jones, O. D., Wagner, A. D. and Raichle, M. E. (2013) “Neuroscientists in Court,” *Nature Reviews Neuroscience*, p.730.

Farah, M. J. and Hook, C. J. (2013). “The seductive allure of “seductive allure”,” *Perspectives on Psychological Science*, *8*(1), pp.88–90.

BIB-033 Farahany, N.A. (2015) “Neuroscience and behavioral genetics in US criminal law: an empirical analysis,” *Journal of Law and the Biosciences,* 2(3), pp.485–509.

BIB-034 Fede, S. J., Borg, J. S., Nyalakanti, P. K., Harenski, C. L., Cope, L. M., Sinnott-Armstrong, W., Koenigs, M., Calhoun, V. D. and Kiehl, K. A. (2016) “Distinct neuronal patterns of positive and negative moral processing in psychopathy,” *Cognitive, Affective & Behavioral Neuroscience*, 16(6), pp.1074–1085.

BIB-035 Fink, B. C., Steele, V. R., Maurer, M. J., Fede, S. J., Calhoun, V. D. and Kiehl, K. A. (2016) “Brain potentials predict substance abuse treatment completion in a prison sample,” *Brain and Behavior,* 6(8), p.e00501.

BIB-036 Fuss, J. (2016) “Legal responses to neuroscience,” Journal of Psychiatry & Neuroscience : JPN, 41(6), pp.363–365.

Gilam, G., Abend, R., Gurevitch, G., Erdman, A., Baker, H., Ben-Zion, Z. and Hendler, T. (2018) “Attenuating anger and aggression with neuromodulation of the vmPFC: A simultaneous tDCS-fMRI study,” *Cortex*, *109*, pp.156–170.

BIB-037 Ginther, M. R., Bonnie, R. J., Hoffman, M. B., Shen, F. X., Simons, K. W., Jones, O. D. and Marois, R. (2016) “Parsing the Behavioral and Brain Mechanisms of Third-Party Punishment,” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(36), pp.9420–9434.

BIB-038 Glannon, W. (2011) *What Neuroscience Can (and Cannot) Tell Us about Criminal Responsibility*. Oxford University Press. [online]. Available from: www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199599844.001.0001/acprof-9780199599844-chapter-2 [Accessed November 1, 2018].

BIB-039 Goel, V., Shuren, J., Sheesley, L. and Grafman, J. (2004) “Asymmetrical involvement of frontal lobes in social reasoning,” *Brain: A Journal of Neurology*, 127(Pt 4), pp.783–790.

BIB-040 Greene, E. and Cahill, B.S. (2012) “Effects of neuroimaging evidence on mock juror decision making,” *Behavioral Sciences & the Law*, 30(3), pp.280–296.

BIB-041 Grove, W.M. and Meehl, P.E. (1996) “Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy,” *Psychology, Public Policy, and Law,* 2(2), pp.293–323.

BIB-042 Güroğlu, B., van den Bos, W., Rombouts, S. A. and Crone, E. A. (2010) “Unfair? It depends: Neural correlates of fairness in social context,” *Social Cognitive and Affective Neuroscience,* 5(4), pp.414–423.

BIB-043 Hardcastle, V.G. and Lamb, E. (2018) “What difference do brain images make in US criminal trials?” *Journal of Evaluation in Clinical Practice*, 24(4), pp.909–915.

BIB-044 Heilbrun, K. (2009) *Evaluation for Risk of Violence in Adults. 1st edition.* New York: Oxford University Press.

BIB-045 Hosking, J. G., Kastman, E. K., Dorfman, H. M., Samanez-Larkin, G. R., Baskin-Sommers, A., Kiehl, K. A., Newman, J. P. and Buckholtz, J. W. (2017) "Disrupted Prefrontal Regulation of Striatal Subjective Value Signals in Psychopathy," *Neuron*, 95(1), pp.221–231.e4.

BIB-046 Janes, A.C., Pizzagalli, D.A., Richardt, S., Frederick, B.D., Chuzi, S., Pachas, G., Culhane, M.A., Holmes, A.J., Fava, M., Evins, A.E. and Kaufman, M.J., (2010) "Brain reactivity to smoking cues prior to smoking cessation predicts ability to maintain tobacco abstinence," *Biological Psychiatry*, 67(8), pp.722–729.

BIB-047 Janowski, V., Camerer, C. and Rangel, A. (2013) "Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL," *Social Cognitive and Affective Neuroscience,* 8(2), pp.201–208.

BIB-048 Kiehl, K.A. (2006) "A cognitive neuroscience perspective on psychopathy: evidence for paralimbic system dysfunction," *Psychiatry Research,* 142(2–3), pp.107–128.

BIB-049 Kiehl, K.A., Anderson, N.E., Aharoni, E., Maurer, J.M., Harenski, K.A., Rao, V., Claus, E.D., Harenski, C., Koenigs, M., Decety, J. and Kosson, D. (2018) "Age of gray matters: Neuroprediction of recidivism," *NeuroImage. Clinical*, 19, pp.813–823.

BIB-050 Kiehl, K.A., Smith, A.M., Hare, R.D., Mendrek, A., Forster, B.B., Brink, J. and Liddle, P.F. (2001) "Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging," *Biological Psychiatry*, 50(9), pp.677–684.

BIB-051 Kittilstad, E. (2018) "Reduced Culpability Without Reduced Punishment: A Case for Why Lead Poisoning Should be Considered a Mitigating Factor in Criminal Sentencing," *Journal of Criminal Law and Criminology*, 108, p.29.

BIB-052 Koenigs, M., Barbey, A.K., Postle, B.R. and Grafman, J. (2009) “Superior Parietal Cortex Is Critical for the Manipulation of Information in Working Memory,” *Journal of Neuroscience*, 29(47), pp.14980–14986.

BIB-053 Kraft, C.J. and Giordano, J. (2017) “Integrating Brain Science and Law: Neuroscientific Evidence and Legal Perspectives on Protecting Individual Liberties,” *Frontiers in Neuroscience*, 11, pp.621–621.

BIB-054 LaDuke, C., Locklair, B. and Heilbrun, K. (2018) “Neuroscientific, Neuropsychological, and Psychological Evidence Comparably Impact Legal Decision Making: Implications for Experts and Legal Practitioners,” *Journal of Forensic Psychology Research and Practice*, 18(2), pp.114–142.

BIB-055 Lamm, C., Decety, J. and Singer, T. (2011) “Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain,” *NeuroImage,* 54(3), pp.2492–2502.

BIB-056 Liu, J. (2011) “Early Health Risk Factors for Violence: Conceptualization, Review of the Evidence, and Implications,” *Aggression and Violent Behavior*, 16(1), pp.63–73.

BIB-057 Maddock, R.J., Garrett, A.S. and Buonocore, M.H. (2003) “Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task,” *Human Brain Mapping,* 18(1), pp.30–41.

BIB-058 McCabe, D.P. and Castel, A.D. (2008) “Seeing is believing: the effect of brain images on judgments of scientific reasoning,” *Cognition*, 107(1), pp.343–352.

BIB-059 McCabe, D.P., Castel, A.D. and Rhodes, M.G. (2011) “The influence of FMRI lie detection evidence on juror decision-making,” *Behavioral Sciences & the Law*, 29(4), pp.566–577.

Meixner, J.B. (2016) “The use of neuroscience evidence in criminal proceedings,” *Journal of Law and the Biosciences*, 3(2), pp.330-335.

BIB-060 Meynen, G., (2013) “A neurolaw perspective on psychiatric assessments of criminal responsibility: Decision-making, mental disorder, and the brain,” *International Journal of Law and Psychiatry*, 36(2), pp.93–99.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). “On the (non) persuasive power of a brain image,” *Psychonomic bulletin & review*, *20*(4), pp.720–725.

Monahan, J. (2008) “Structured risk assessment of violence,” *Textbook of violence assessment and management*, 2, pp.17–33.

– – (1981) “Predicting violent behavior: An assessment of clinical techniques,” *Sage Publications*.

BIB-061 Morse, S. (2011) Lost in Translation? An Essay on Law and Neuroscience. Rochester, NY: Social Science Research Network. [online]. Available from: https://papers.ssrn.com/abstract=1904488 [Accessed October 10, 2018].

BIB-062 – – (2016) Neuroethics: Neurolaw. Rochester, NY: Social Science Research Network. [online]. Available from: https://papers.ssrn.com/abstract=2919011 [Accessed October 10, 2018].

BIB-063 Murray, R.J., Brosch, T. and Sander, D. (2014) “The functional profile of the human amygdala in affective processing: Insights from intracranial recordings,” *Cortex*, 60, pp.10–33.

BIB-064 Nicholls, T.L., Ogloff, J.R.P. and Douglas, K.S. (2004) “Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and VSC,” *Behavioral Sciences & the Law*, 22(1), pp.127–158.

BIB-065 Pardini, D.A., Raine, A., Erickson, K. and Loeber, R. (2014) “Lower amygdala volume in men is associated with childhood aggression, early psychopathic traits, and future violence,” *Biological Psychiatry*, 75(1), pp.73–80.

BIB-066 Pardo, M.S. and Patterson, D. (2010) “Philosophical Foundations of Law and Neuroscience,” *University Of Illinois Law Review*, 2010(4), p.40.

BIB-067 Paulus, M.P., Tapert, S.F. and Schuckit, M.A. (2005) “Neural activation patterns of methamphetamine-dependent subjects during decision making predict relapse,” *Archives of General Psychiatry,* 62(7), pp.761–768.

BIB-068 Peer, E. and Gamliel, E. (2013) “Heuristics and Biases in Judicial Decisions,” *Court Review*, 49, p.114.

BIB-069 Phelps, E.A. and LeDoux, J.E. (2005) “Contributions of the amygdala to emotion processing: from animal models to human behavior,” *Neuron,* 48(2), pp.175–187.

BIB-070 Poldrack, R.A., Monahan, J., Imrey, P.B., Reyna, V., Raichle, M.E., Faigman, D. and Buckholtz, J.W. (2018) “Predicting Violent Behavior: What Can Neuroscience Add?” *Trends in Cognitive Sciences*, 22(2), pp.111–123.

BIB-071 Raine, A. and Yang, Y., (2006) “Neural foundations to moral reasoning and antisocial behavior,” *Social cognitive and affective neuroscience,* 1(3), pp.203–213.

BIB-072 Remmel, R.J., Glenn, A.L. and Cox, J. (2018) “Biological Evidence Regarding Psychopathy Does Not Affect Mock Jury Sentencing,” *Journal of Personality Disorders,* pp.1–21.

BIB-073 Riva, P., Romero Lauro, L.J., Vergallito, A., DeWall, C.N. and Bushman, B.J. (2015) “Electrified emotions: Modulatory effects of transcranial direct stimulation on negative emotional reactions to social exclusion,” *Social Neuroscience*, 10(1), pp.46–54.

BIB-074 Saks, M.J., Schweitzer, N.J., Aharoni, E. and Kiehl, K.A. (2014) “The impact of neuroimages in the sentencing phase of capital trials,” *Journal of Empirical Legal Studies*, 11(1), pp.105–131.

BIB-075 Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. and Cohen, J.D. (2003) "The neural basis of economic decision-making in the Ultimatum Game," *Science*, 300(5626), pp.1755–1758.

BIB-076 Schaich Borg, J., Sinnott-Armstrong, W., Calhoun, V.D. and Kiehl, K.A. (2011) "Neural basis of moral verdict and moral deliberation," *Social Neuroscience,* 6(4), pp.398–413.

BIB-077 Schofield, P.W., Malacova, E., Preen, D.B., D'Este, C., Tate, R., Reekie, J., Wand, H. and Butler, T. (2015) "Does Traumatic Brain Injury Lead to Criminality? A Whole-Population Retrospective Cohort Study Using Linked Data," *PLOS ONE*, 10(7), p.e0132558.

BIB-078 Schweitzer, N.J., Saks, M.J., Murphy, E.R., Roskies, A.L., Sinnott-Armstrong, W. and Gaudet, L.M. (2011) "Neuroimages as evidence in a mens rea defense: No impact," *Psychology, Public Policy, and Law*, 17(3), pp.357–393.

BIB-079 Specker, J., Focquaert, F., Sterckx, S. and Schermer, M.H. (2018) "Forensic practitioners' expectations and moral views regarding neurobiological interventions in offenders with mental disorders," *BioSocieties*, 13(1), pp.304–321.

BIB-080 Steele, V.R., Fink, B.C., Maurer, J.M., Arbabshirani, M.R., Wilber, C.H., Jaffe, A.J., Sidz, A., Pearlson, G.D., Calhoun, V.D., Clark, V.P. and Kiehl, K.A. (2014) "Brain potentials measured during a Go/NoGo task predict completion of substance abuse treatment," *Biological Psychiatry*, 76(1), pp.75–83.

BIB-081 Steele, V.R., Maurer, J.M., Arbabshirani, M.R., Claus, E.D., Fink, B.C., Rao, V., Calhoun, V.D. and Kiehl, K.A. (2018) "Machine Learning of Functional Magnetic Resonance Imaging Network Connectivity Predicts Substance Abuse Treatment Completion," Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 3(2), pp.141–149.

BIB-082 Steele, V.R., Claus, E.D., Aharoni, E., Vincent, G.M., Calhoun, V.D. and Kiehl, K.A. (2015) "Multimodal imaging measures predict rearrest," Frontiers in Human Neuroscience, 9. [online]. Available from: www.frontiersin.org/articles/10.3389/fnhum.2015.00425/full [Accessed October 10, 2018].

BIB-083 Tassy, S., Oullier, O., Cermolacce, M. and Wicker, B. (2009) "Do psychopathic patients use their DLPFC when making decisions in moral dilemmas?" *Molecular Psychiatry*, 14(10), pp.908–909.

BIB-084 Tengström, A., Grann, M., Långström, N. and Kullgren, G. (2000) "Psychopathy (PCL-R) as a predictor of violent recidivism among criminal offenders with schizophrenia," *Law and Human Behavior*, 24(1), pp.45–58.

BIB-085 Treadway, M.T., Buckholtz, J.W., Martin, J.W., Jan, K., Asplund, C.L., Ginther, M.R., Jones, O.D. and Marois, R. (2014) "Corticolimbic gating of emotion-driven punishment," *Nature Neuroscience,* 17(9), pp.1270–1275.

BIB-086 Walton, M.E., Devlin, J.T. and Rushworth, M.F.S. (2004) "Interactions between decision making and performance monitoring within prefrontal cortex," *Nature Neuroscience,* 7(11), pp.1259–1265.

BIB-087 Ward, J.T., Hartley, R.D. and Tillyer, R. (2016) "Unpacking gender and racial/ethnic biases in the federal sentencing of drug offenders: A causal mediation approach," *Journal of Criminal Justice,* 46, pp.196–206.

BIB-088 Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E. and Gray, J.R. (2008) "The Seductive Allure of Neuroscience Explanations," *Journal of Cognitive Neuroscience*, 20(3), pp.470–477.

BIB-089 Weisberg, D.S., Taylor, J.C. and Hopkins, E.J. (2015) "Deconstructing the seductive allure of neuroscience explanations," *Judgment and Decision Making*, 10(5), p.13.

BIB-090 Wright, J.P., Dietrich, K.N., Ris, M.D., Hornung, R.W., Wessel, S.D., Lanphear, B.P., Ho, M. and Rae, M.N. (2008) “Association of Prenatal and Childhood Blood Lead Concentrations with Criminal Arrests in Early Adulthood,” in J. Balmes, ed. *PLoS Medicine,* 5(5), p.e101.

BIB-091 Yamada, M., Camerer, C.F., Fujie, S., Kato, M., Matsuda, T., Takano, H., Ito, H., Suhara, T. and Takahashi, H. (2012) “Neural circuits in the brain that are activated when mitigating criminal sentences,” *Nature Communications,* 3, p.759.

Yang, Q., Shao, R., Zhang, Q., Li, C., Li, Y., Li, H., & Lee, T. (2019). “When morality opposes the law: an fMRI investigation into punishment judgments for crimes with good intentions,” *Neuropsychologia*, *127*, pp.195–203.

BIB-092 Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A. and Saxe, R. (2010) “Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments,” *Proceedings of the National Academy of Sciences* of the United States of America, 107(15), pp.6753–6758.

[1] Authorship is equal.

[2] Though this result is correlational in nature, Young and colleagues (2010) has noted the same role for the rTPJ in belief and intent attribution. By utilizing transcranial magnetic stimulation (TMS) the researchers were able to “lesion” the area in order to observe its effects on behavior, and confirm its role with a more causal approach.