Neuroimaging, Responsibility, and Law

How Neuroscientific Explanations Challenge the Fundaments of Legal Responsibility

Nolina Doud

14/07/2016

Master Thesis
MSc Philosophy of Science, Technology, and Society - PSTS

Supervisor: Prof. dr. Ciano Aydin
Examiner: dr. Saskia Nagel

University of Twente
Faculty of Behavioural, Management, and Social Sciences

Abstract:

This colloquium gives an overview of my master's thesis, in which I explore how neuroscience, law, and philosophy overlap at the site of neuroimaging. I focus on the arguments proffered by four main stakeholders: cognitive and neuro scientists, legalists, philosophers, and laypeople. I introduce the controversy between neuroimaging's enormous potential in the courtroom and the concern that the neuroscientific modes of explanation which accompany them undermine the traditional legal notion of responsibility. I contend that in order to reconcile this conflict, it must be understood on the conceptual level. I introduce and critically evaluate two conceptual approaches which offer ways to integrate neuroscience and law: "neuro-reductionists" and "distributed causalists." The research question which I seek to answer is: how can a conceptual analysis of the neuro-reductive and distributed causalist stances on the relationship between the mind, brain, body, and world offer a way to reconcile novel neuroimaging applications with the traditional legal notion of responsibility? I respond to this question by conducting an extensive literature review. First, I enumerate what arguments, both pro and contra neuroimages, are articulated inside the courtroom, and how these conflicts are appropriated into broader issues by the public. Second, I explore the philosophical concepts which co-constitute the legal notion of responsibility: causality, agency, and mental states. Third, I compare this traditional legal approach to the framework of neuro-reductionism, demonstrating how neuro-reductionism challenges each presupposition of legal responsibility, and therefore, the concept of legal responsibility as a whole. Fourth, I compare the legal approach to an alternative conceptual framework, distributed causalists. I demonstrate how this conceptual approach allows stakeholders to take neuroimages into account without engendering major conceptual conflicts between this framework and the legal concepts. I conclude that it is possible to take neuroimages and neuroscientific explanations into account in the courtroom without provoking a major conceptual conflict between different approaches to responsibility.

Table of Contents

CHAPTER 1: IDENTIFYING THE PROBLEM

1: Introduction

This introductory section will proceed through several steps. 1.1 will review the public conceptions of neuroimaging, with a particular focus on fMRI. 1.2 will connect these emerging neurotechnological practices with the embedded practice of law. 1.3 will introduce a tension produced by the introduction of neuroimages into the courtroom. 1.4 will introduce my research question and how I intend to address this tension. 1.5 will introduce a taxonomy of analytic angles and terms that I use throughout the thesis. 1.6 connects the neuroimaging enterprise with attempts to "read minds." 1.7 demonstrates how the introduction of "mind reading" through brain-scanning into the courtroom provokes stakeholders to reconsider and/or reify their stance on the mind, brain, body, and world.

1.1 Public Conceptions of Neuroimaging

Since around the 1990's, there has been an enormous increase in neuroscientific funding, research, publications, public interest, and emerging applications. Recent years have witnessed such an immense rise in popularity that various scholars and policy-makers have termed the current era as "the decade of the mind" (Kutas & Federmeier, 1998; Moriarty, 2008:29). Today, the neurosciences play an increasingly influential role in the practices and conceptual frameworks not only in the laboratory and medical field, but also in pubic-policy making, popular media, social sciences, private enterprises, and beyond. The technological practice of neuroimaging has commanded particular attention from all these spheres.As Pardo & Patterson put it, "Neuroscience has been aided significantly by technological advances over the past couple of decades. The most significant development has been 'neuroimaging' in general and functional magnetic resonance imaging (fMRI) in particular" (Pardo & Patterson, 2013:xxiii). The ability to visualize brain activity is fairly new and inchoate, yet it is precisely this visual quality of these productions makes them particularly appealing, not only in the lab setting, but also in popular media and public imagination (Beaulieu, 2002; Dumit, 2004; Eastman & Campbell, 2006; Weisberg et al., 2008). Researchers, policymakers, philosophers, and lay-people alike can "read" these images in some way or another, arriving at diverse interpretations. As Roskies (2007) articulates, "The seeming accessibility of these images, and the grip they have on the scientific and public imagination, makes them important conduits of information about the progress of neuroscience" (863-864). The perceived accessibility of neuroimages is key to the dissemination of what Choudhury et al. (2009) terms "brain facts." These are productions like "the neural basis of gender, criminality, morality or culture, from the point of its entry into the lab to its treatment in the lab through various technical

and knowledge practices to interaction with the media and policy and its reception by the public" (64.) These "brain facts" move from the lab and spread into other disciplines, changing meaning as they change context and viewer. Neuroimages have a mass-appeal that helps "brain facts" move beyond the relatively stable lab, into new disciplinary domains.

Of all the neuroimaging technologies, Functional Magnetic Resonance Imaging (fMRI)[1] draws a particular amount of commentary and interest, often ascribed to the fact that it is non-invasive and can therefore conduct studies on larger sample sizes and in more diverse contexts than its invasive counterparts. fMRI neuroimages have an appeal and perceived accessibility that makes them a particularly popular and exemplary conduit for neuroscientific modes of explanation. A keyword search of "fMRI" in the University of Twente's research database yields 37,594 results, indicating the immense interest in this particular technology. As one news article put it, "Hardly a week goes by without the media showcasing beautiful three-dimensional images of the brain in action, which supposedly explain how and why humans do the things we do" (Oullier, 2012:7). This interest has also been noted in more formal analyses. The preeminent neuroscientists Nikos Logothetis also performed a keyword search (yielding 19,000 results), commenting that "Its popular fascination is reflected in countless articles in the press speculating on potential applications, and seeming to indicate that with fMRI we can read minds..." (Logothetis, 2008:869). In summary, across peer-reviewed journals, news, forums, and social media, fMRI is increasingly a topic of interest, not just for what it can tell us about the brain, but primarily for what it can tell us about the mind.

1.2. A Turn To "Neurolaw"

One of the most fertile and growing spheres of emerging fMRI applications is in the field of law. The field of "neurolaw" is "a growing list of legal domains in which neuroscience may prove relevant" (Wolf, 2008:21), and it is undoubtedly on the rise, particularly in the United States. Neuroimages are a key technological practice in disseminating neurosciences into new domains, and often recognized as neurolaw's "most powerful technical adjunct" (Snead, 2007:1269). Proposed applications range from detecting jury biases, to measuring pain or damage for proper mitigation, to tamper-proof lie detection, to new insights into addiction and violent behavior. This is a particularly stakeholder-diverse field, and for the purposes of this thesis I focus on four primary stakeholder groups: neuro and cognitive scientists, legalists (lawyers, judges, legal scholars), philosophers, and lay-people. Since the legal applications of fMRI are primarily taking place in the U.S., the role of the jurors must

[1] For a brief technological description, see Appendix A

be taken into account. This broadens the stakeholders to include all citizens, because they could be expected to use neuroimages in considering their verdicts, or, in the event that they became a defendant, he/she could be tried with the use of such technology.

Law frequently requires demonstrations that the defendant has or had a particular state of mind, and this is often where neuroimaging technologies are deployed. Wolf (2008) connects these endeavors when she writes, "law has everything to do with human intentions, states of mind, competence, culpability, and responsibility, the points of possible connection between law and neuroscience are legion" (Wolf, 2008:21). Criminal law, in particular, is integrally concerned with understanding the mind of the defendant. In order to prosecute, the lawyers are required to "prove beyond a reasonable doubt" that a particular mental state pertained, such as the intent to kill before committing the murder (this is a defining feature for first degree murder), the intent to deceive (a defining feature of fraud charges), or the capacity to control his/her actions (a defining feature of the insanity defense). To any crime, there are two parts, the *actus reus*, the actual crime committed, and the *mens reas*, the requisite state of mine for that criminal act. So while the *acteus reus* of a homicide would be the act of killing, the *mens reas* would be the intent to kill. Jones and Shen (2012) clarify, "Culpability of the accused thus depends, in part, on a determination of his/her mental state at the time of the offense. The phrase 'mens rea' ('guilty mind') derives from the Latin phrase 'Actus non facit reum nisi rea sit,' which means 'An act is not guilty unless the mind is guilty'" (361). All issues in criminal law depend, to some extent, on inquiring into the content of "the mind" of the defendant, and this is notably the most difficult aspect of a trial (Sallet, 1985;1549). This challenge makes neuroimages extremely alluring because they may enable us to shed some light on the opaque, complex, controversial issues regarding the *mens rea* of a crime.

1.3 Orienting to the Problem: Hopes and Fears in Neurolaw

The question of whether neuroimages are ready for the courts is often provoked when court cases seeking to introduce neuroimages in novel domains. Within the court, experts and lawyers articulate specific practical and empirical arguments regarding what a particular neuroimage can show about a particular defendant's state of mind. They must respond to the general question: is it appropriate to use a neuroimage in this proposed role? However, outside of the court, these specific instances become exemplars of much broader, conceptual issues. Section 2 introduces several court cases which have increased the visibility of neuroimaging in these fields and provoked stakeholders to reflect on its role. I focus on the particularly controversial topic of "insanity" or "diminished capacity" defenses in

which the defendant claims that he did not have the capacity to be aware of and/or control his wrongdoings. The question on everyone's mind outside of the court is: how do these neuroimages and neuroscientific explanations provoke our intuitions about legal responsibility?This potential partnership, characterized in the term "neurolaw," is generally agreed to be a fascinating, but controversial, relationship; fraught not only with the empirical and practical difficulties articulated within the court, but also the ethical and conceptual difficulties articulated outside of it.

Whether skeptical or optimistic about the productions of neuroimaging, stakeholders generally agree that neuroscience can yield very serious implications for law. Approximations of its potential impact range from sweeping claims of the potential for neuroscience to "revolutionize the law" (Rosen, 2007:1; Green & Cohen, 2006, Churchland, 2004), more modest proposals that neuroimaging can help the law but not fundamentally change it (i.e. Feigenson, 2006; Sinnott-Armstrong et al., 2008), to claims that neuroimaging does not offer anything new to law (i.e. Morse, 2011). One popular media article promoting a symposium on neurolaw begins, "Discussions about the 'promise of neuroscience' are often tinged with a mixture of hope and fear. Nowhere is this ambivalence more evident than in the courts, as conjecture runs rampant about the legal impact of this research, stoked by claims that neuroscience may soon detect liars, objectively determine criminal responsibility, quantify suffering, and predict violence"(Faigman, 2015). Another popular media article introduces the categories "Promise and Terror" (Davis, 2012), and a commentator on Reddit observes, "Someday brain scans will be a paradigm-shifter in the legal system - which is exciting in some ways and scary in others" (Reddit, 2013), indicating that there is a recurring trope that neuroimaging has significant potential and dangers.

So, for the purposes of this thesis, how do I conceive of these "hopes and fears" regarding neuroimaging in the courtroom? I focus on a specific conceptual tension that recurs in insanity or diminished capacity cases where the issue of responsibility takes center stage. On one side, there is the recurring promise of neuroimaging, particularly noninvasive neuroimaging like fMRI: these technologies can offer insight into the most challenging problem of law, namely, the *mens reas* aspect of a crime. Pardo and Patterson (2013) write in a more general vein, "If psychological states can be reduced to brain states, then fMRI can quite literally 'read another person's mind.' In the legal context, such a powerful technology could have limitless potential" (26). The idea that fMRI can "read minds" is one of its most powerful draws (explored further in 1.6), connecting the mental inquiries of law with the neurological inquiries of neuroscience. Furthermore, there is a prevailing notion that this technique is more "objective" than the law's reliance on psychological analysis (Lamparello, 2012). This claim to

objectivity is bolstered by the perception that neuroimaging enables explanations human behavior in terms of determined, physiological processes occurring in the brain – providing a new variation on biological determinism. This is often termed "neuro-reductionism," which is defined by the assumption that anything that could be explained in terms of the mind could (and should) be explained at the level of the brain. Some authors go so far as to posit that neuroimages actively "produce" or "mediate" neuro-reductionism (Pirruccello, 2012:454; Aydin, 2016). Section 4 commits to exploring these approaches at the conceptual level, and explicating how this stance is relevant to legal concepts. Some scholars claim that only be assuming this stance can neuroscience truly revolutionize the law, for better (Churchland, 2004, Greene, & Cohen, 2006, Lamparello, 2012), or worse (Morse, 2011; Pardo & Patterson, 2013). This neuro-reductionist approach underlies the more ambitious "hopes" for neuroimaging in the courtroom, but what underlies the fear Faigman mentioned?

On the other side, "the fears" Faigmen refers can be understood as the concern that neuro-reductionism could revolutionize the law to such a degree that it threatens traditional notions of free will and responsibility. Roskies (2006), who has a background in both cognitive neuroscience and law, observes that, "Advances in neuroscience provide us with an increasingly mechanistic view of how the brain generates complex thought and behavior. This trend has led some to worry that future advances will lead people to abandon their belief that we are free agents and, consequently, that our views of moral responsibility will be undermined" (419). Roskies identifies the pervasive view that viewing ourselves in mechanistic, determined, and/or reductionistic terms has the potential to undermine the legal model of thought and behavior, and therefore undermine our notions of free will and responsibility. Although the implications for law which neuro-reductionists propose themselves are more diverse than this characterization (as explored in section 4.2.2, 4.3.2, and 4.4.2), there is a prevailing notion that this approach results in a challenge to traditional legal notions. This approach is therefore viewed as engaged in a paradoxical relationship where neuroimages are introduced to support the legal procedures, but their accompanying neuroscientific modes of explanation threaten to debunk its most fundamental notions (Morse 2005; Snead, 2007; Roskies, 2006).

Many stakeholders argue that a "brain based" view of thought and behavior is essentially reductionistic, deterministic, and/or dualistic (further explored section 4.4). However, they still proffer that neuroimages, and their accompanying focus on the brain, can be of some utility to law. I contrast the "neuro-reductive" approach with a group I call "distributed causalists" – a term which encapsulate the notion that causal efficacy is not solely a property of the mind, brain, body, or world, but emerges in an interplay of all of these elements. These authors offer a way for law to take the brain into account

without precluding the legal notion of responsibility. These authors argue that there are more causal elements for which to take into account than just the brain, and the foundations and implications of their ideas constitute Section 5. Other factors which scholars propose as important in an account of human thought and behavior, in addition to the brain, include the mind (5.1.1), the body (5.1.2), social-cultural-and-environmental context (5.1.3), and technology (5.1.4). Section 5.2 explores what these various alternatives framework could imply for law.

1.4 Research Question

There are extreme skeptics and extreme optimists– and even more stakeholders whose views fall in-between – speculating on the potentialities of fMRI in the legal domain. A rigorous scholarship on practical and empirical issues is emerging. However, less scholarship explores the conceptual side of the issue; for instance, how neuroimages frame the causal relationship between mind, brain, body, and world and whether this challenges or coheres with the traditional legal conceptualization of this relationship. Yet, there is a prevailing concern that the neuroscientific modes of explanation ushered into the courts via neuroimaging may ultimately come to conflict with the traditional legal notion of responsibility. Several scholarly works have been devoted to exploring this issue (i.e. Aharoni et al., 2008; Federspiel, 2007; Gazzaniga, 2011, Glannon, 2005; Greene & Cohen, 2006;) and also popular media articles (i.e. Eagleman, 2011; Doherty, 2012; Faigman, 2015). This thesis is partly devoted to mapping the different arguments which have already been articulated and relating them back to conceptual notions, and partly devoted to evaluating them in terms of how they cohere with traditional legal concepts. It will proceed through three main steps, each with a dedicated chapter: *Identifying* the problem, *Clarifying* the problem, and *Resolving* the problem. The central research question to which I will devote the rest of this thesis is: how can a conceptual analysis of the neuro-reductive and distributed causalist stances on the relationship between the mind, brain, body, and world offer a way to reconcile novel neuroimaging applications in the court with traditional legal notions of responsibility?

The subquestions to this main question guide the structure of this thesis. Before each subquestion, I have included the number of the section which is intended to respond to the question:

Chapter One: How do I *identify* the conflict between neuroimaging and law?

Section 2: What emerging conflicts between neuroimaging and law have already been identified?

2.1: What conflicts have been sparked by fMRI in brain anomaly detection?

2.3: How do these conflicts relate to conceptual frameworks?

Chapter Two: How do I *clarify* why this is a conflict?

Section 3: How does law traditionally conceive of the causal relationship between the mind, brain, body, and world?

3.1.1: How does law traditionally conceive of causality?

3.1.2: How does law traditionally conceive of agency?

3.1.3: How does law traditionally conceive of mental states?

3.3.1: How do these traditional notions relate to legal notions of responsibility and free will?

Section 4: Why is neuro-reductionism perceived as a threat to the traditional legal notion of responsibility?

4.1-4.3: How do various neuro-reductionists conceive of the relationship between mind, brain, body, and world and why are these perceived as a challenge to legal responsibility?

4.2: How do these notions influence the legal notions of responsibility and free will?

4.3: How are these neuro-reductionist claims appropriated by other stakeholders?

4.4: What critiques have been articulated to challenge these approaches?

Chapter Three: How do I *resolve* this conflict?

Section 5: How is distributed causality an alternative to neuro-reductionism?

5.1: What are the different approaches to distributed causality?

5.2: How does distributed causality influence the legal notions of responsibility and free will?

1.5 Characterizing a "Conceptual" Analysis

This thesis is organized around two main applications of fMRI in the legal context, the two most talked-about throughout fMRI discourse: the detection of brain anomalies and the detection of lies and/or deception. By focusing on these applications, I narrow my scope to criminal law, although there are potential applications in civil[2] and constitutional law[3]. Within the borders of criminal cases of brain-anomaly and lie detection, there is still a staggering diversity of different considerations presented by the many stakeholders. I borrow a helpful taxonomy from Pardo & Patterson (2013) through which to consider the different (interrelated) types of issues which present in neurolaw. They introduce four categories of argumentation one can encounter in neurolaw. These are: empirical, practical, ethical, and conceptual.

- *Empirical issues* relate to how strong the correlations are between the experimental task the neural activation, how the experiment was set-up, the reliability of the experimental set-up and

[2] I.e. de Kogel et al., 2014.

[3] I.e. Federspiel, 2007; Krauss, 2010.

statistical models employed, etc.

- *Practical issues* relate to "challenges regarding the integration of neuroscience into issues of law and public policy... for example, determining when and how such evidence should be introduced into legal proceedings, as well as determining what legal standards and instructions will govern the inferences that may or not be drawn from that evidence" (Pardo & Patterson, 2013:xv). The arguments which are articulated within the court primarily cite practical and empirical arguments, but they often implicate ethical and conceptual issues. To find dedicated inquiry into the ethical and conceptual side of neurolaw, though, one usually has to look to academia.
- *Ethical issues* are those which concern values like "privacy, safety, dignity, autonomy, and other values implicated by proposed uses of neuroscience in law" (Pardo & Patterson, 2013:xv). The very notion of legal responsibility is most often considered by ethicists because of its integral relationship with moral responsibility. Additionally, much of the literature I use is from the emerging field of neuroethics (i.e. Churchland, 2004; Gazzaniga, 2006; Wolf, 2008; Choudhury et al., 2009; Pirruccello; 2012; Vincent, 2011).
- *Conceptual inquiries* "concern the application of concepts involving the mind and the diverse array of psychological powers, capacities, and abilities that we associate with having a mind. The conceptual issues focus on the scope and contours of the concepts being employed in claims involving law and neuroscience" (Pardo & Patterson, 2013:xvi).

Conceptual inquiries constitute the primary focus of my thesis, Pardo & Patterson's work, as well as several of the other authors upon which I draw throughout this thesis (i.e. Dumit, 2004; Bennet & Hacker, 2009; Moore, 2011; Aydin, 2016). Pardo and Patterson clarify their definition using the example of the lie detector when they write:

> It is an empirical question whether a particular person is lying on a particular occasion, and it is an empirical question whether particular brain activity is correlated with lying behavior. But what constitutes a 'lie' is a conceptual question.... Notice also that any answer to the two examples of empirical questions noted above (i.e., whether a person is lying and whether brain activity is correlated with lying) will presuppose some conception of what constitutes 'lying.' (Pardo & Patteson, 2013:xviii).

Therefore, broad conceptual notions underly the specific statements made about the lie detection test. Similarly, conceptual notions, such as what "the brain" and "the mind" are and how they relate, underly

how any neuroimage is "read" and interpreted. Other authors share Pardo & Patterson's commitment to inquiring into the more conceptual issues of neurolaw. Dumit delineates some fundamental conceptual notions when he writes, "Each piece of experimental design, data generation, and data analysis, however, necessarily builds in assumptions about human nature, about how the brain works, and how person and brain are related" (Dumit, 2004:15). Much like a lie detection test presupposes what a "lie" is, inquiries into "the mind" of presupposes what it is to be a mind, which in turn relates to more general notions about how the brain, mind, body, and world relate. While more authors are calling for these kinds of conceptual inquiries, they generally acknowledge that such inquiry "has garnered considerably less attention in the scholarly literature" (Pardo & Patterson, 2013: xvii). This is where philosophers can offer the greatest insight, in cooperation with neuro and cognitive scientists and legal scholars, who have presented many rich insights into the ethical[4], empirical[5], and practical issues[6].

1.6 Neuroimaging as "Mind-Reading"

FMRI seems to be particularly of interest when it is understood to unlock or expose some secret content of the brain or mind. Gilbert et al. suggest that it is most often the supposed function of "brain reading" or "mind reading" which makes various neuroimages appealing to a diversity of stakeholders. Neuro and cognitive scientists adopt this language within the laboratory (i.e. deCharms, 2008). Talk of the "brain/mind reading" capacities of neuroimaging also occurs prodigiously outside of the lab, among diverse stakeholders like public policy makers (i.e. Littlefield, 2009; Snead, 2007), private enterprises (i.e. NoLie MRI, Cephos), education (Mobbs et al., 2007), law (i.e. Sinnott-Armstrong et al., 2008), and popular media (i.e. Frank, 2009; Smith, 2013). Gilbert et al. conducted a dedicated study on the occurrence of phrases like "brain-reading" and "mind reading" and found such phrases endemic to the neuroimaging discourse. They initially report,

> Indeed, a cursory internet search... reveals several academic articles utilizing the 'brain-reading' metaphor. For instance... 'the possibility of reading out a person's thoughts [from their brain] does exist' (Anonymous 2009).... 'reading the private intentions of a person' from the brain (Amodio and Frith 2006), or the ability to 'decod[e] mental states from brain activity in humans' (Haynes and Rees 2006), or even claims to 'mind-reading with a brain scan' (Smith 2008) are not uncommon uses of the 'brain-reading' metaphor. (Gilbert et al., 2011:229-230)

Their study demonstrates convincingly that the potential of explaining mental phenomenon in

[4] I.e. Choudhury et al., 2009; Vincent, 2011
[5] I.e. Klein, 2010; Roskies, 2007
[6] I.e. Moriarty, 2008; Aharoni et al., 2008

neurological terms, through the conduit of neuroimages, is particularly captivating to all the various stakeholders. The interest in these "brain/mind reading" applications spurs the dissemination of "brain facts." "Brain facts" like fMRI neuroimages can be "read" in many different ways, especially as different stakeholders from beyond the (relatively stabilized) laboratory add their perspectives. This becomes particularly apparent in a stakeholder-diverse field like neurolaw. While this does not mean that an individual can read *anything* from a neuroimage, it does mean that what he/she interprets can differ significantly from what another stakeholder reads.

Due to the mass appeal of the "mind reading" applications for fMRI, particularly in the field of neurolaw, these applications constitute the focus of this thesis. This means that I focus on a specific branch of neuroscience, that of of *cognitive* neuroscience, which Michael Gazzaniga, the so-called "father of cognitive neuroscience," defines as: "the field of scientific endeavor that is trying to understand how the brain enables the mind" (Gazzaniga et al., 2010:2). This discipline is dedicated to understanding the complicated relationship between "mental states" and "brain states." It includes inquiries into a number of different "mental" phenomenon, including: "cognition (Parker et al., 2002), perception (Mather, 2006; Wolfe et al., 2006), emotion (Dalgleish, 2004; Damasio, 1999; Davidson, 2001), decision-making (Moll et al., 2008; Pfaff, 2007), social understanding (Frith and Frith, 2003) and trust (Krueger et al., 2008)" (Choudhury et al., 2009:63). These are applications in which fMRI is directed at investigating a human capacity typically attributed to "the mind." "The mind," broadly conceived, can be thought of as such: "To have a mind is to possess an array of rational and emotional powers, capacities, and abilities exhibited in thought, feeling, and action... It is simply the mental powers, abilities, and capacities possessed by humans" (Pardo & Patterson, 2013: 45). Since I deploy this broad definition, I include authors referring to related modi such as "psychological predicates," "cognitive capacities," "mental states," "intentional states," or similar terms.

What is important for this inquiry is not what "the mind" *is*, or how I conceive of it, but how inquiries regarding neuroimaging and law relate "the mind" to "the brain." As Seaman (2009) argues, "he human mind and the jury room have often been described as 'black boxes.' As the metaphor suggests, for each of these decision-making organs (the brain and the jury) we can measure the input and we can view the output, but the process that takes place inside the box – the process that transforms the inputted data into the outputted decision – is inscrutable, perhaps even magical" (931). This thesis is intended to illuminate how neuroscientific discourse reifies and renegotiates the complexity of relating the mind, brain, body, and world, provoking us to open the "black boxes" -- not to decide which schema is correct. The law traditionally only invokes "the mind," and disregards "the brain," but

also “black-boxes” the meaning of “the mind,” leaving it open to interpretation. Increasingly, as neuroimages are introduced, so, too, are neuroscientific explanations for thought and behavior, which may go so far as to disregard “the mind” and only invoked “the brain.” Rather than argue for one or the other notion, I intend to demonstrate that *some* notion of the causal relationship between mind, brain, body, and world necessarily underlies how one interprets neuroimages. As Glannon notes,“brain imaging may be a helpful tool in determining whether persons... can be held morally and legally responsible for their behavior. Depending on what imaging techniques show about the brain, and how we interpret these images, they could influence moral and legal judgments about culpability, blame, and excuse” (Glannon, 2005:69). Glannon adds that *what* a neuroimage can offer law, and even *whether* it is perceived to have any influence on law, depends on conceptual commitments. He encourages us to open the “black-box” of presuppositions and popular perceptions in tandem with opening the “black-box” of what the mind is, how it works, and how it relates to neuroimaging.

1.7 Linking Philosophy, Neuroscience, and Law through Concepts Presupposed in “Responsibility”

Using insights from philosophy of technology, this thesis will respond to the call for increased conceptual analysis of the intersections of neuroscience and law, by way of fMRI technology. I will focus on a central conceptual issue, that is: how the perceived causal relationships between the mind, brain, body, and world change (and possibly threaten or bolster) the traditional legal notion of responsibility. I will use a diverse array of literature to demonstrate how different notions of this relationship present in neuroimaging discourse. As I perceive it, these are the central factors which may or may not exercise causal efficacy in the human-lived world. My taxonomy coheres with a similar set of factors introduced by authors like Aydin (2016), who includes self, brain, cognition, and world (5-6); Freeman (2006) who includes “world, brain, and mind” (89); and Noë (2009) who uses “brain, body, and world” (10). I have selected these elements in order to stay as inclusive as possible in exploring the varying conceptual notions which form the basis of legal responsibility and emerging neuroscientific forms of explanations. The relationship between mind, body, brain, and world is relevant to neurolaw insofar as it is used to assert some of the basic preconditions for responsibility. The traditional legal notion of responsibility is contingent on several other concepts which I seek to clarify in this thesis. They are: causality, agency, and mental states. The law conceives of these concepts in specific ways, and together they form a basis upon which legal responsibility can be ascribed. The perceived conflict between law and neuroimaging (or neuroscience at large) can be understood as a conflict between these conceptual presuppositions.

So how is *causality* relevant to law? Throughout the history of philosophy, "'Cause' is an important concept in the explanation of human action. It is central to debates about free will, and no explanation of human action is complete without an account of the role of causes in behavior" (Pardo & Patterson, 2013:35). Particularly in the decade(s) of the mind, the causal role of mentation and brain states is a hotly contested but increasingly popular topic. As Kim (2007) puts it, "Devising an account of mental causation has been, for the past three decades, one of the main preoccupations of philosophers of mind..." (8). The causal linkages between mind, brain, body, and world could be invaluable elements in any robust explanation of human activity. However, these issues are incredibly complex and relate to a network of equally complex concepts.

In the legal field, the causal relationship between the mind, brain, body, and world is of particular interest insofar as it is relevant to understanding *agency*. While "causality" is important to law (as explored further in section 3.1.1), "causal agency" is even more important in the legal context because an animal or machine can cause harm, but it will not be held responsible because it is traditionally not conceived of as "an agent" (Morse, 2011; Moore, 2011; Schlosser, 2015). Only when an entity is conceived of as a causal *agent* can he/she/it be considered legally or morally responsible. "Agency" is traditionally defined in terms of "intentionality," and since machines and animals are traditionally not conceived of as having causally efficacious intentions, they are not agents, and therefore they are not held legally or morally responsible (this is further explored in section 3.1.2). Morse (2011), for example, endorses this view when he writes, "Virtually everything for which agents deserve to be praised, blamed, rewarded, or punished is the product of mental causation.... Machines may cause harm, but they cannot do wrong... Machines do not deserve praise, blame, reward, punishment, concern, or respect... Only people, intentional agents with the potential to act, can do wrong and violate expectations of what they owe each other" (212). So while machines may have causal *efficacy*, i.e. they can cause harm, they do not have causal *agency*, because they are not intentional beings – therefore, they are not held responsible. This relates to mental states because intentions are often understood to be mental states, and/or formed in relation to other mental states like representations and desires (further explored in section 3.1.3).

The legal notion of "responsibility," in the sense of deserving punishment for wrong-doings, is integrally linked to the notion that humans are causal agents whose mental states (like beliefs, desires, intentions) have causally efficacy. As Kim (2007) summarizes, "the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world" (9). The posited causal relationship between the mind, brain, body, and world has ramifications

for how one understands causal agency, free will, and responsibility. It is important to note that although causal agency is a necessary condition for responsibility, it is not a sufficient condition. Prosecuting lawyers must demonstrate that the defendant played a demonstrably significant causal role in bringing about the crime (*actus rea), and* he/she was acting as a free (un-coerced, in-control) agent, *and* he/she intended the crime (*mens rea)*. Only with all three requisite conditions in place, as well as other, more practical and bureaucratic matters, can one be eligible for legal responsibility (or, in other words, liability, culpability, etc). This establishes gradations of legal responsibility: there are those who are conceived as causing a crime, but are not conceived of as causal agents, such as the criminally insane, who are not held responsible in the sense that they are sentenced to rehabilitation but not punishment; there are those who are conceived of as causal agents, but lacking the intent of the crime (i.e. criminal negligence), who are semi-responsible and therefore given mild punishments; and those who are conceived of as causal agents, acting freely and with the intent to commit the crime, who are fully responsible and therefore eligible for maximum sentencing. The meaning and relationship of these concepts to responsibility and free will constitutes the topic of section 3.1.4.

1.8 Methodological and Stylistic Considerations

I selected my four stakeholder groups (neuro and cognitive scientists, philosophers, legalists, and lay-people) to include a broad range of different perspectives. However, I can by no means exhaust this diverse field. Therefore, I had to make some borders regarding the stakeholders for which I can account. I took inspiration from a delineation established by Lee between "state-centered" and "individual-centered". Lee (2014) defines, "The state-centered theory focuses on the proper limits of the state's power to criminalize and punish, while the individual-centered theory focuses on questions of innocence and culpability. This division is, of course, somewhat artificial. Both approaches can and do coexist, often within the same piece of scholarly work" (671). While acknowledging their interrelatedness, I focus primarily on the "individual-centered" issues. Since I do not focus on questions of public policy or governmental applications, I do not include such stakeholders[7]. Future research certainly could approach the issue from such a perspective. Additionally, other differentiations might illuminate other stakeholder groups, but the four aforementioned groups are nonetheless the most frequently represented throughout the existing literature, and therefore throughout this thesis.

Focusing on the aforementioned stakeholders, each stakeholder group has a corresponding type of literature. The works of cognitive neuroscientists and philosophers were provided entirely through

[7] For articles which do take a "state-centered" approach, authors include Littlefield, 2009; Kulynych, 2002; Snead, 2007

academic journals. The works of legalists were partly supplied by academic journals, and also case briefs and official laws and statutes. The lay-people, of course, require the most diverse representation. I refer to popular media articles, primarily in the form of online news. When discussing popular media before online news, I refer to newspapers. Additionally, I make use of the even more informal level of these publications – the comments sections. The reason for this diversity of sources is to do justice to the diversity of viewpoints and stakeholders, all with (different) underlying notions about what the mind and brain are, how they relate, and what fMRI can analyze about this relationship. I also include some authors who analyze other, similar neuroimaging technologies like electocephalography (EEG), Computerized Axial Tomography (CAT), and Positron Emission Topography (PET). Although in technical terms these technologies are quite different, they share similar conceptual fundaments due to their focus on the brain's role in explaining human thought and behavior, so although this thesis focuses on fMRI, it often refers more broadly to neuroimaging or even more broadly to neuroscience at large. Rather than focus solely on court cases with fMRI, which are few and far between, I sketch a more general history of the rise of neuroimaging technologies and, most importantly, the brain-based explanations of thought and behavior that pave the way for neuroimages to play a role.

It is also important to note that these cases take place within the United States. It is generally agreed that neurolaw is particularly on the rise in the U.S., and many of the precedents for the whole Western world are being set there, particularly when it comes to neuroimaging in the courtroom. However, this also does not mean that these insights are irrelevant to other contexts. On the conceptual side, I utilize authors from all over the world, indicating that the practices may be largely confined to the U.S., but considerations of what these practices could entail appear from all over the world. As Pardo and Patterson (2013) assert, "...a number of important recent developments involving cases in the United States and criminal law have dominated the discussions. Although the doctrinal analysis will relate primarily to criminal law within the United States, we believe the examples are of more general interest in illustrating how problematic conceptual issues arise at the level of legal doctrine" (Pardo & Patterson, 2013:xxi). Similarly to Pardo and Patterson, I focus on cases which occurred in the U.S., but draw implications from these which could be relevant internationally.

Another methodological consideration is the particular branch of law upon which I focus. Neuroimaging is relevant in several fields of law, but it bears particularly strongly on the issue of responsibility in the context of criminal law. Both my court cases and my conceptual resources primarily focus on legal issues which arise when harm has been done to another individual. Moore (2009) comments, "it is criminal law and the law of torts that are most directly reflective of an

underlying moral responsibility... [They] seemingly predicate legal liability on causal responsibility, among other things" (3). This clear connection between causation, agency, mental states and legal responsibility makes this a particularly fertile domain to study. This focus on criminal acts also influences the types of court cases I have selected.

I chose the application of brain-anomaly detection for a number of reasons. Firstly, these applications fall within the domain of criminal law, and can therefore directly influence assessments of responsibility. As aforementioned, the notion of responsibility is what requires lawyers and cognitive scientists to employ notions of causal agency in the first place. Secondly, they are a particularly popular topic throughout the types of discourse selected. There are simply more commentary regarding in these applications than, for example, screening jury biases or detecting memories of a specific event. Thirdly, these applications are specifically concerned with *cognitive* neuroscience, and inquire into what could be understood as mental states. Part of the controversy of these cases is due to the fact that they necessarily invoke contentious presuppositions about the causal relationship between mind, brain, body, and world. These applications are especially theory-laden, meaning they are particularly informed by conceptual presuppositions. For example, a civil case which involves the use of fMRI to detect a brain anomaly for the sake of damage mitigation is considerably less controversial than using fMRI to detect a brain anomaly for the sake of reducing the defendant's liability for harm he clearly caused. The former case uses stabilized presuppositions and produces stably-interpreted neuroimages. The latter case relies on presuppositions which are not stabilized, and consequently, the interpretations of the neuroimage are also more varied. When the complicated relationship between the mind, brain, body, and world comes into play, the cases are more controversial, and there is a higher degree of commentary from all involved stakeholders.

It is important to note at the outset that this thesis is not intended to be, nor could it possibly be, an exhaustive taxonomy of different notions of the causal relationship between the mind, brain, body, an world. As the reader can guess, this is an incredibly diverse field, ranging from religion, to philosophy, to neuroscience, to law, to basic human self-understanding. Rather than provide "the whole picture," I have conceptualized this thesis as partly a mapping exercise, indicating the broad range of topics at the nexus of philosophy, technology, science, and law. My goal in this exercise is that any of the four stakeholders I have enumerated (cognitive neuroscientists, philosophers, legalists, and laypeople) could, through this work, become acquainted with some of the main authors, arguments, and conceptual standpoints in the field. The second part of this thesis (Sections 4 and 5) is more evaluative based on which conceptual framework can cohere with the traditional legal notion of responsibility.

This presupposes that the reader *wants* to preserve this notion, however this may not always be the case. Some of the challenges to the legal responsibility are enumerated in this thesis, insofar as they are relevant to neuroimaging and neuroscientific modes of explanation. Some of these challenges are beyond that scope, and it is beyond the scope of this thesis to justify *why* we should strive to preserve the legal notion of responsibility (reasons are supplied by authors like Morse, 2005; Gazzaniga, 2011). This thesis could serve as a foundation upon which to critically inquire into this notion, because I clarify it by way of identifying several concepts which co-constitute this notion. By making clear the conceptual presuppositions which constitute this notion, a critical reader could develop of a critique of legal responsibility at a fundamental level.

This broad approach does engender certain hazards. One criticism is that it may not be broad *enough,* and that certain viewpoints which are highly influential in the field do not get articulated here. This may be the case because, as I have mentioned, this is an incredibly fertile and diverse field, and there are simply too many views, all influencing one another, for me to enumerate them all. Again, I emphasize that this thesis cannot be exhaustive, but if the reader should wish to get more informed on these issues, he/she could easily begin with the authors reviewed here. The other criticism, of course, is that it is *too* broad, and therefore does not do justice to the finer details of each framework. This is also a valid criticism, but the response is much the same – this thesis is about providing exposure, and giving the reader the option to pursue the more detailed, tailored works which cohere with his/her various purposes. Philosophy tends to value precision, and for good reason. Indeed, many of the arguments reviewed here will be founded on the notion that descriptions and definitions should be precise and clear. In this thesis, many descriptions and definitions are left purposely broad. For example, I avoid defining "the mind" in any narrow way in order to include as many different viewpoints which relate to the topic. Philosophers, cognitive neuroscientists, legalists, and laypeople use a variety of terms when trying to understand how humans engage in things like perception, intention, decision-making, emotions, and communication (to name a few). In order to relate them to one another, I maintain open, permeable definitions so that I may include varied stakeholders and conceptual stances. So the broad-strokes approach is both the weakness and the strength of this thesis; I have prioritized exposing the reader to a (manageably) diverse set of viewpoints over extremely faithful replication of the finer details of these viewpoints. It is for the reader to decide where his/her purposes lie and whether such an approach can serve them.

2: Legal Cases: From Court to Public Opinion

This section identifies the problem through practices and discourse in the courtroom. 2.1 explores the relevance of these cases and how neuroimaging plays a particularly key role in insanity cases. 2.2 Explores how the application of detecting brain-anomalies with a neuroimage is deployed in court. 2.2.1-2.2.3 explore the Hinckley, Weinstein, and Dugan cases in some detail. 2.4 summarizes how these cases are relevant to the broader conceptual discussion of the role of neuroscientific explanations in deliberating on responsibility.

2.1. Setting the Stage with Public Discourse

Before delving into my more conceptual arguments, I will set the stage by sketching a brief history of the rise of neurolaw, with a specific focus on neuroimaging practices. I will review several prominent court cases to form an initial articulation of the contentious issues regarding the role of neuroimages in court. These cover the two main prospective uses introduced by Aharoni et al. (2008) "defenses that deny intentions and affirmative defenses, such as insanity" (149). These cases are all "affirmative defenses," which means that they affirm that the crime took place, the *actus reus* element of the crime pertained, but that mitigating circumstances prevent the defendant from being held responsible, namely that the *mens reas* was not present. In each of these cases, neuroimaging played a key role in making these claims, and although the success of these defenses vary, they form an important site upon which expert witnesses, lawyers, and the concerned public articulate their views on these emerging technological applications.

Overall, this more practically-oriented section serves several functions of setting the stage for the second section, in which I elaborate more conceptual arguments. This section justifies the importance of studying this topic by demonstrating the rising popularity of neuroimages in this new context. Most importantly, this section serves to demonstrate some of the influential arguments that are provoked by these cases. These include the arguments that appear in the context of the courts, which can only include practical and empirical considerations, as well as commentary on these proceedings from legalists and popular media, which broaden these discussions to include more ethical and conceptual reflections.These cases will show how legalists have adopted and adapted the neurosciences through the deployment of neuroimages, and the inchoate shift from traditional legal notions of "the mind" to references to "the brain." By going chronologically, I will demonstrate how legal notions are adapted to new technologies, practices, and public response. However, these shifts and adaptations are hardly stabilized, as the following conflicts and negotiations are intended to demonstrate. This section,

ultimately, represents how "brain facts" move from within the courtroom to beyond it, and how they become even more conceptually loaded in the process. It demonstrates how practical and empirical arguments constitute the majority of the discourse inside the courtroom, but outside the courtroom the issues become much broader. In order to make any claim to clarifying or evaluation the argumentation of these stakeholders, it is important to understand what the main arguments are.

2.2. Brain Anomaly Cases

Legal arguments regarding brain anomalies can serve a number of purposes in the court. In criminal court, neuroimages can play a role in several stages, from the initial liability phase, to sentencing, to capital punishment, and parole hearings. Within all these stages, neuroimages can be used in a multitude of ways, both in favor of the defendant and against him/her. Jones and Shen elaborate:

> In the criminal system, brain evidence may be offered during the liability phase, the sentencing phase, or both. For example, during the liability phase, the defense may offer brain evidence to support an insanity defense, or to defeat the prosecution's claim that the defendant had (and was therefore capable of having) the mental state requisite for conviction, or to provide evidence of truthfulness. During the sentencing phase, brain evidence may be offered to support a mitigated penalty. (Jones & Shen, 2012:358).

They emphasize the most common application of neuroimages, which is generally introduced on the defendants behalf. However, in an interview with the prominent neurologist Helen Mayberg, she cautions that there is a possible reading of neuroimages which does not favor the defendant when she describes, "You need to be prepared for: 'This spot is a sign of future dangerousness,' when someone is up for parole. They have a scan, the spot is there, so they don't get out. It's carved in your brain." (Qtd. H.S. Mayberg in Rosen, 2007:7). So although arguments invoking "mental defects" and "brain anomalies" may decrease the severity of a sentencing, i.e. getting put in a mental institution instead of a prison, they can also be used to increase the period of incarceration due to the belief that the thought or behavior is "hardwired" into the brain. In fact, the Supreme Court has already upheld multiple civil cases in which people were incarcerated past their sentencing due to the belief that "involuntary commitment is permissible when limited to 'those who suffer from a volitional impairment rendering them dangerous beyond their control'" (Lamparello, 2012:347). So neuroimages are not deployed unilaterally on behalf of the defendant, although in the cases below they serve that function. The multilaterial deployment of neuroimages emphasizes how interpretations of them can vary widely,

especially in the context of law where the same set of evidence is almost always interpreted to mean very different things.

Another de-stabilizing factor in this application is the varying definitions of what a brain anomaly is, what a neuroimage can analyze about these brain anomalies, and what all that has to do with the legal issues like criminal responsibility, let alone the more conceptual issue of the relationship between the mind, brain, body, and world. At this point, it is important to note out now is that what a "brain anomaly" or "mental defect" *is* is constantly renegotiated within the court. In the process of demonstrating that, in fact, a relevant anomaly exists, lawyers and experts also define what that anomaly is. This means that "even what is accepted as 'mental disorder' varies between different branches of law addressing distinct justice issues... In law, there is no such thing as 'real' mental disorder, only definitions of it that are adopted for purposes that usually have nothing to do with medical constructions of mental disorder *per se"* (Eastman & Campbell, 2006:312). What constitutes a mental disorder or brain anomaly, therefore, changes in context. This issue is not exclusive to the field of neurolaw, and even the most accepted taxonomy of mental disorders, *The Disagnostic and Statistical Manual of Mental Disorders,* is highly controversial (i.e. Wakefield, 2013). Therefore, there is no overarching definition of what a mental defect or brain anomaly is for the court and how it can be used in defense or prosecution. These issues, unless regulated by a specific statute, are negotiated case-by-case, with some interpretations of precedents. Therefore, I explicate cases in which arguments about mental defects and brain anomalies were raised, without presuming what they actually are or whether they were actually present in the person of interest.

2.2.1. *People v. Hinckley,* 1982

The first case in this history is the high-profile case of John Hinckley, charged in 1981 with the attempted assassination of President Ronald Reagan, among several other, smaller charges. This case is significant because it brought the insanity defense sharply into the public eye. The insanity defense is an important topic in the legal context of fMRI because it necessarily invokes "mental defects" as a qualifying criteria. Furthermore, after the Hinckley case, the US government regulated that, "Mental disease or defect does not otherwise constitute a defense'" (*IDRA,* 1984:Stat. 2057). Therefore, in the liability stage of the trial, neuroimages regarding brain-anomalies are almost exclusively deployed for insanity defenses and cannot be used, for example, as evidence of good character. The Hinckley trial played a key role in shaping the contours of the insanity defense, which remains largely unchanged to this day. In order to understand the role fMRI could play in insanity defense cases, we must first

understand what is required for an insanity defense.

The next logical question to ask is, what is the insanity defense? Summarized in vernacular, "The federal insanity defense now requires the defendant to prove, by 'clear and convincing evidence,' that 'at the time of the commission of the acts constituting the offense, the defendant, as a result of a severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts'" ("Insanity Defense," 2016). The wording of this is open to interpretation, and in the Hinckley trial, the word "appreciate" was a central issue. His lawyers, "argued (apparently) successfully that it meant not only cognitive awareness, but also included an emotional understanding of the consequences of his actions" (Fuller, 2000:700). Defining and evaluating whether Hinckley could "appreciate" his wrongfulness hinged on the expert testimony of psychiatrists and psychologists.

What is particularly interesting about this case as it relates to fMRI is that it shows the kinds of situations in which neuroimages become desirable. Neuroimaging for brain anomalies becomes appealing when, "As in *United States v Hinckley*, there is often dispute about whether a defendant claiming legal insanity suffered from a mental disorder, which disorder the defendant suffered from, and how severe the disorder was. At present, these questions must be resolved entirely behaviourally, and there is often room for considerable disagreement... In the future, neuroscience might help resolve such questions" (Morse, 2011:227-228). Hinckley's behavior could be analyzed in multiple ways; spun one way, his obsessive behavior preceding the crime could be explained as showing intent and premeditation (required *mens rea* for his conviction); spun another way, it could be interpreted as showing clinical red-flags of an underlying mental disorder significant enough to qualify for the insanity defense (Sallet, 1985:1548). In this quagmire of competing behavioral theories, brain imaging seemed an effective way to tip the scale to the defense's side.

The defense, therefore, sought to introduce another point of reference besides Hinckley's behavior: his neuroanatomy. Brain imaging played a role in this defense in which lawyers used CAT scans to show abnormal activity in his prefrontal lobes. Although the defense could not show the CAT scans to the jury as they were ruled inadmissible, the expert witness was permitted to acknowledge the role they had in his diagnosis. The defense expert, Dr. Bear, "lectured the court on the role of the CAT-scan in his diagnosis and announced that 'I would like the right to state to the jury that an important test which I use in reaching my conclusions has been barred by the court and I was not able to present it to the jury, though I believe it would influence their decision as it has influenced mine'" (Sallet, 1985:1549). This provoked the ongoing question of whether it is worth admitting scientific evidence that may not satisfy the rigorous evidentiary rules for admissibility but which may nonetheless be

crucial in a medical experts' diagnosis. There are compelling arguments for both. This issue remains unresolved, as the next case (*People v Weinstein*)will also demonstrate. However, although the brain scans were not viewed by the jury, "most observers agree that the neuroimaging evidence was instrumental in securing it [the insanity defense]" (Keursten, 2015:para. 20). Furthermore, these lawyers and experts paved the way for neuroimages and neuroscientific modes of explanation to enter the courtroom – via the invocation of "mental defect." Most importantly, this defense established a linkage between statistical anomalies on a neuroimage, a mental defect, and a diminished sense of responsibility.

Hinckley's acquittal led to enormous public outcry. A 1982 newspaper article closely following the ruling stated, "Three of every four people questioned in a national poll said justice was not done when a jury found John W. Hinckley Jr. innocent by reason of insanity..." (AP, 1982:3-A). Two years after the Hinckley trial, the US government responded by releasing the *Insanity Defense Reform Act* (IDRA). This enacted several restrictions on levying insanity defenses, making it significantly more difficult than it was for the Hinckley trial, hence why this defense is rarely made these days. Perhaps most significantly, the IDRA placed strict limitations on what kind of testimony an expert witness is permitted to make. It reads, "No expert witness testifying with respect to the mental state or condition of a defendant in a criminal case may state an opinion or inference as to whether the defendant did or did not have the mental state or condition constituting an element of the crime charged or of a defense thereto. Such ultimate issues are matters for the trier of fact alone" (*IDRA*, 1984:Stat. 2068). This means, in effect, "the prohibition for any expert to express an opinion on the ultimate issue of legal responsibility" (Fuller, 2000:702). The inference, therefore, from a neuroimage to a brain anomaly may be expressed by an expert, but the connection between the brain anomaly, a requisite mental state, and legal responsibility is left to the trier of fact to decide. As another 1982 newspaper reads, "When the moment finally arrives for a verdict – still weeks away – the jurors will get only a bare-bones formula from the court to help them decide, deliberately leaving them to apply their common-sense notions about criminal responsibility to the framework of information provided by the psychiatric specialists" (Kiernan, 1982:8). This is important to note because it means the conceptual frameworks of the expert witness, the legalist, and the jurors are not stabilized around a common understanding. The juror is required to invoke his/her own conceptual notions when evaluating the defendant's legal responsibility. Despite the challenges these reforms mounted against the insanity defense, the next case will show how lawyers continue to make in-roads for neuroimages using the path established in the Hinckley trial.

Before turning to the next case, it is interesting to note that the Hinckley trial has recently

returned to public scrutiny due to an appeal that the state no longer require him to stay in a mental institution. Although he has already been allowed to leave for extended periods of time, he is in the process of making an appeal to be fully emancipated from his sentencing. He would, as a *Los Angeles Times* article notes, be the first man to attempt to assassinate a president and go free (Phelps, 2015). What is most interesting about this is that now, with the wonder of social media, the concerned public can air their diverse readings of the issues at hand. The comments on the *Los Angeles Times* article demonstrated that the public opinion on this issue was far from monolithic. Several responses were along the lines one commentator articulated: "He played the insanity card. Once insane--always insane. Keep the goon in the asylum" (Qtd. In Phelps, 2015). This indicates that people seemed to think, as Helen Mayberg predicted, that a propensity towards violence is innate and therefore permanent. Several dedicated scholars cohere with this opinion, and even argue that sentences can be extended on the basis of neurobiological indicators for violence or other criminal behaviors (i.e. Sapolsky, 2004; Lamparello, 2012). On the opposite side of the spectrum, some commentators fell more into the camp of: "If Hinckley's doctors feel he can be released, and the judge agrees, then he has paid his debt. The guy is 60 years old. ALL the research and data PROVE that nearly 100% of inmates over age 60 are no longer a threat due to age..." (Qtd. In Phelps, 2015). This perspective seems to imply that these violent propensities can be changed over time, that they are not "hardwired" into the brain, or that they are at least mitigated by the limitations of old age. Several dedicated arguments have been made to challenge making a one-to-one relationship between brain anomalies and violent behaviors (i.e. Mayberg, 1992; Aharoni et al., 2008; Morse, 2005). This demonstrates that these cases continue to rouse and challenge the intuitions of legalists, philosophers, cognitive scientists, and concerned citizens. Furthermore, it shows how these cases can provoke conceptual issues, such as whether criminal behaviors are "hardwired," an approach which could possibly cohere with neurological determinism. On the other side, we see the resistance to this approach, and the notion that such behaviors can change due to factors which are not neurobiology, such the capacities of an aging body.

2.2.2 *People v. Weinstein*, 1992

The second significant case in this history is the 1992 case of *People v. Weinstein*, in which Herbert Weinstein plead guilty by way of insanity to the crime of strangling his wife to death before throwing her over the 12th floor balcony to make it look like a suicide. This case is significant because it deployed an insanity defense relying directly on a PET scan presumed to show a brain anomaly which would reflect on Weinstein's criminal responsibility. This case is often cited as an inaugurator of

neuroimages in court. Legal scholars concur, claiming that, “The prosecutor in the case predicted that, with Weinstein, 'the age of scanning has dawned in our courtrooms. This is not a technological genie we are going to be able to put back in the bottle'” (Jones & Shen, 2012:362). His case established a viable route through which neuroimages could enter the courtroom – as evidence of a brain anomaly that mitigates the defendant's criminal responsibility. From then on, attempts to introduce neuroimages in the courtroom increased dramatically. However, they still face the issue of whether they satisfy the empirical and practical standards established within the courts. The Weinstein case is particularly interesting because it is one of the first dedicated evaluations of neuroimages in the court's own empirical standard: rules of evidence.

In pretrial, Weinstein's lawyers faced the issue of whether evidence from the PET scan would be admissible within the legal rules of evidence. The rules of evidence place a number of limitations on the kinds of evidence that may be used in the trial, as well as the kinds of testimony it can receive. Most often, courts refer to the *Frye* test, “The 'general acceptance' test of *Frye* for the admission of scientific evidence at trial is the standard most often used by courts throughout the United States” (*People v Weinstein,* 1992:37). The *Frye* ruling states:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. (*Frye v. United States,*1923:1014).

The “general acceptance” standard which *Frye* established means that independent experts (often from defense and prosecution) must sufficiently agree that a particular technological practice is “generally accepted” in the scientific community. For example, with fMRI, in order to satisfy the *Frye* test, experts must agree that the experimental conditions, the models and formulas used to process the data, and the interpretations of this data are “generally accepted” in the scientific community. Feigenson (2006), who writes from the perspective of a legal scholar, enumerates three basic levels of inference which must be “generally accepted” in order to enter the court. The first is the inference from BOLD levels to fMRI data, or in other words, the models used to translate the raw data from the fMRI into statistical models of BOLD levels. He cautions that these are “affected by researchers’ decisions regarding and assumptions underlying data processing methods, most of which have not yet converged on the kind of consensus that would allow the basic technology to be ‘black boxed’” (Feigenson, 2006:239). The

second inference is from BOLD level to neuronal activity, which he presumes would likely satisfy the *Frye* test because neuroscientists generally agree that BOLD levels directly correlate with neuronal activity (i.e. increased oxygenation means increased neuronal activity). The third inference is from the neuronal activity to the mental state, or, as he terms it, "psychological function." He cautions, "inference (3), from neuronal activity to psychological function, raises fundamental questions about the theories and concepts relied upon in the design of fMRI studies and the associations drawn between fMRI data and the cognitive or emotional function of interest, and hence about the application of the data to the legal issue at hand" (Feigenson, 2006:239). He concludes that it is unlikely in the current state of the art that neuroimages can satisfy the *Frye* test. However, although the Weinstein court came to a similar conclusion, a legal loophole in the insanity defense allowed a certain amount of the PET evidence to receive testimony. This shows that even if neuroimaging does not satisfy the main evidentiary standard, that does not mean it cannot enter the courts.

Although the judge allowed the PET images in the trial, he also added restrictions. Specifically, "certain theories relating to human behavior that may not be mentioned in testimony at the trial. Evidence concerning these theories is not admissible because they have not been generally accepted as valid in the fields of psychiatry, psychology, and neurology" (*People v Weinstein,* 1992:46). The court included a specific prohibition against the inference that an arachnoid cyst or reduced glucose metabolism in the frontal lobes can be linked to violence. They affirm that such theories are not generally accepted and do not satisfy the *Frye* test. So although the expert witness may form his own diagnosis, he/she ultimately cannot comment on whether the defendant is criminally responsible or had the requisite mental state for committing the crime. In-keeping with the *IDRA*, the trier of fact is left to decipher the extent to which the brain abnormalities abnegate Weinstein of his responsibility.

This defense did not prove entirely successful, and although Weinstein received a reduced plea, he still went on to spend many years in prison. Nonetheless, this case is considered a milestone in the history of neurolaw. One popular media article begins with the introduction, "When historians of the future try to identify the moment that neuroscience began to transform the American legal system, they may point to a little-noticed case from the early 1990s. The case involved Herbert Weinstein..." (Rosen, 2007:1). Ultimately, the mere admission of the PET images drew significant scholarly attention to the issue and increased the visibility of neuroimaging as a legal tool in the public eye (Jones & Shen, 2012:361). The neuroimages are still thought to have played a crucial role in the outcome of this case. As Kulynych (1997) wrote "...the visual impact of such neuroimages is hard to overstate. This impact was apparent in *People v. Weinstein*... This lesion appeared in the neuroimages as a gaping black hole

in the frontal lobes, which contrasted dramatically with the bright red and green hues of more metabolically-active regions. The judge's ruling that such evidence would be admissible... reportedly led the prosecution to accept a man-slaughter plea" (1251). For Kulynych, this case is exemplary of the drama and appeal, sometimes termed the "seductive allure," of neuroimages, particularly in situations where psychological theories are competing. For others, this case was a harbinger of things to come. Indeed, the effects of this case has been very real. For example, the forensic psychologist who testified in the case, "found himself in so much demand to testify as a expert witness that he started a consulting business called Forensic Neuroscience. Hired by defense teams and prosecutors alike, he has testified over the past 15 years in several hundred criminal and civil cases. In those cases, neuroscientific evidence has been admitted to show everything from head trauma to the tendency of violent video games to make children behave aggressively" (Rosen, 2007:1). This case established a specific role and an in-road of argumentation through which neuroimaging could enter the court. Following the restrictions created after the Hinckley trial, this case forged a new way to fit neuroimages into the insanity plea. This case further re-enforced the connection established in the Hinckley case between the law's invocation of mental states (specifically "mental defects" in this case) and the neuroimage's invocation of brain states (specifically an arachnoid cyst in this case). The trier of fact alone is left to decide how convincing this connection is and how that bears on the issue of legal responsibility.

2.2.3 *The Trials of Brian Dugan*, 2015

Unlike the previous cases, which occurred in the guilt/liability phase of court, this introduces the role in capital sentencing through a series of cases regarding Brian Dugan. Of all the applications in the courtroom for neuroimages, capital sentencing is the most common. As Daniel Martell, the expert from the Weinstein case said in an interview, "it's in death-penalty litigation that neuroscience evidence is having its most revolutionary effect. 'Some sort of organic brain defense has become de rigueur in any sort of capital defense'" (D. Martell qtd in Rosen, 2007:1). This is largely due to the fact that in this phase of the trial, the evidentiary standards are significantly "relaxed" compared to the guilt/liability phase (Jones & Shen, 2012:359). The strict evidentiary standards makes it rare for neuroimages to enter at earlier stages, but when the evidentiary rules are "relaxed," the potential roles for neuroimages increase. Neuroimages are actually so popular at this stage that "a Florida court has held that the failure to admit neuroscience evidence during capital sentencing is grounds for a reversal" (Rosen, 2007:1). This means that if the prosecution *cannot* procure some kind of neurological evidence against the defendant, that alone is grounds for the defendant to be taken off death-row.

The Brian Dugan trial attracted a particular amount of media attention for neuroimaging in law. This is not only because it is among the first instances where fMRI was admitted as evidence in such a context, but also because its details are particularly disturbing. It involves the kidnapping and murder, among other charges, of a young girl and the wrongful sentencing of two men who served several years on death row until DNA evidence brought to the court in 2009 acquitted them both and pointed to a new defendant – Brian Dugan. Brian Dugan immediately pled guilty and was also put on death row in July 2009. However, the state in which he was being prosecuted overturned the death penalty in 2011, and for this reason he was not executed. However, the county State Attorney's office is still pressing to have him executed, to this day.

In one of Dugan's many appeals, his defense attorney enlisted the help of a foremost expert on psychopathy, Kent Kiehl of the University of New Mexico. Kiehl has been doing brain scans on presumed psychopaths for several years, and continues to this day. Outside of the courtroom, he has prodigiously produced well-received studies on how the brains of those diagnosed with psychopathy differ from those who are not. He observes, for example, "When a normal person sees a morally objectionable photo, his limbic system lights up. This is what Kiehl calls the 'emotional circuit,' involving the orbital cortex above the eyes and the amygdala deep in the brain. But Kiehl says when psychopaths like Dugan see the... picture, their emotional circuit does not engage in the same way" (Hagerty, 2010: para.15). Dugan was a subject of the typical panel of tests in Kiehl's study. As part of his defense, Kiehl testified that Dugan exhibited the same brain abnormalities he detected in most diagnosed psychopaths. However, "he was careful not to stretch beyond what the data show. He didn't claim, for example, that the brain scans prove that Dugan committed his crimes as a result of a brain abnormality" (Miller, 2009; para. 4). However, he did argue that, "psychopaths are a little like people with very low IQs who are not fully responsible for their actions. The courts treat people with low IQs differently. For example, they can't get the death penalty" (Hagerty, 2010: para. 17). Much like the defense of Hinckley and Weinstein, this defense relies on the idea that Dugan's abnormal brain makes him less responsible.

The people also called an expert witness, Johnathan Brodie, to testify against Kiehl's reading of the neuroimages. Brodie cited three main arguments to challenge Kiehl, and these arguments represent some of the main, recurring empirical arguments which challenge fMRI applications of this kind. The first argument regards the issue of timing, and how a neuroimage taken twenty-six years after the fact could possibly reflect anything about Dugan's cognitive state at the time of committing the murder. This specific argument for Dugan relates to the more general argument that unless you were literally

sitting in the brain scanner as you committed the crime, fMRI could tell little about your brain state (or mental state) at the time. Crawford, for example, comments that "Perhaps the most fundamental limitation of functional imaging, vis- à-vis the claim that it allows us to 'peer inside the mind,' is that there is a basic disconnect of time scale" (Crawford, 2008:71). Crawford argues that fMRI might not be sufficiently fast enough to measure cognitive events as they are happening, let alone whether it can measure a cognitive event several years later. It is a conceptual issue whether you conceive of the mind or brain as static enough to provide insight on events that occurred many years ago. Other scholars have recognized this issue, including Rose (2003) and Mayberg (2010).

The second argument Brodie articulated can be termed: the individual versus the average. Neuroscientists generally agree that brains can vary widely, and since fMRI models with statistical averages among groups of test subjects, it may struggle to account for individual differences. Several scholars have levied this argument, including Kutas & Federmeier (1998), Dumit (2004), Feigenson (2006), Sinnott-Armstrong et al. (2008), Rose (2005), Aharoni et al. (2008), and Morse (2011). The last argument Brodie made was that, "Even if fMRI could reliably diagnose psychopathy, it wouldn't necessarily reduce a defendant's culpability in the eyes of a judge or jury. Ultimately, the law is based on an individual's rational, intentional action, not brain anatomy or blood flow..." (Hughes, 2010:342). Other scholars have supported the claim that ultimately, the criteria of the law is, and should remain, mostly behavioral (i.e. Morse, 2011; Pardo & Patterson, 2013). They counter that if one has the neuro-anatomy for a crime, but does not commit it, it would be absurd to consider him/her responsible. Therefore, having the corresponding neuroanatomy to a criminal behavior is not enough to incriminate someone. These three arguments are some of the main empirical and practical challenges to the introduction of neuroimaging to law. Although they challenge the specific interpretations of neuroimages in specific contexts, they do not address the more general conceptual concern that the introduction of neuroscientific modes of explanation threatens the traditional legal notion of responsibility. They hint at this issue, though, by showing how brain images can be used to explain mental and behavioral states. In all cases where these brain-anomaly neuroimages were introduced, the response was that the defendant's culpability decreased. This shows that these neuroimages do have an impact on how the jury understands the defendant's responsibility.

2.4 Summary

This section has sought to provide a practical introduction to the emerging contours of fMRI neuroimages in the legal domain. These cases show that the courtroom serves as a adequate site to

articulate empirical and practical arguments regarding the robustness of a specific neuroimage in a specific applications. They also show that the juror and the concerned public are left to decide what bearing these emerging applications and explanatory modes have on the broader issues of how we conceive of legal responsibility. As one popular media article commented, "Despite the rarity of the defense, we talk about it a lot. In part that's because it makes for particularly colorful moral and legal drama... But it's also because the defense raises deep and eternally controversial questions about compulsion and free will" (Doherty, 2012:para.11). These cases provoke the public to confront difficult questions regarding to what extent neurobiology can account for thought and behavior; furthermore, to what extent that might mean we are biologically determined; and furthermore, to what extent that might shake the fundamental notion of legal responsibility.

So how do these cases relate to my research question? In order to understand or reconcile the perceived conflict between neuroscientific explanations and legal responsibility, it is important to understand how the conflict began. These cases provoke scholars and the concerned public to confront the question of how a neuroimage could relate to deliberations on responsibility as they consider what they would have done in the jurors' and judge's positions. They also show that the conceptual issue of whether neuroimages threaten or bolster traditional legal notion of responsibility cannot be stabilized within the trial; stakeholders are required to make their own deliberations on these matters. The trial can introduce prominent practical and empirical arguments, but can only hint at conceptual ones. These trials help to identify the perceived conflict, and how a specific argument like "Hinckley's abnormal brain made him unable to perform as a morally and legally responsible agent" can give way to the question: "if Hinckley's brain caused his behavior, do all humans' brains cause all human behaviors?". A commentator on a recent article on the Hinckley trial reflected, "'It's not my fault. My defective brain made me do it.' I don't find this defense even remotely convincing. We cannot apportion culpability among a person's body parts. The whole person committed the crime -- or not, as the case may be. A perp's brain scans are interesting but irrelevant" (Qtd. In Kuersten, 2015). This layperson is ultimately grappling with a conceptual issue: is it logically appropriate to ascribe responsibility to the brain? This question has been the subject of multiple scholarly analyses, including Bennet and Hacker's (2009) *Neuroscience and Philosophy.* They also make the argument that this is a conceptual, and not empirical or practical, issue. They argue, "The question we are confronting is a philosophical question, not a scientific one. It calls for conceptual clarification, not for experimental investigation. One cannot investigate whether brains do or do not think, believe, guess, reason, form hypotheses etc. until one knows what it would be for a brain to do so" (Bennet & Hacker, 2009:19). Before one can answer

whether brains think, he/she must presuppose what it would mean for a brain to think. Similarly, before one can answer whether the neurobiology of the brain has any bearing on responsibility, he/she must presuppose what responsibility is and how it relates to the brain.The aforementioned lay-person seems to intuit that the brain is not a logically appropriate place to investigate when deliberating on legal responsibility, and therefore brain scans are not an appropriate technology for the courtroom. However, not all scholars or lay-people cohere with these conceptual intuitions.

On the other side, many laypeople and scholars are more optimistic, concluding that these neuroimages are a more accurate litmus test of the *mens rea* aspect of the crime and therefore give a more accurate judgement on responsibility. One commentator on the same Hinckley article as above "I still envision a progression whereby we come to apply the law more and more based on biology and scientific knowledge than moral or retributive principles" (Qtd. In Kuersten, 2015). While one commentator seemed to imply that the brain is not an appropriate place to look in deliberations on responsibility, this commentator seems to argue that the brain is a *more* appropriate place to look than in traditional deliberations. A constituent of commentators generally seem to accept that the brain can indeed play such a strong causal role that an individuals is not held responsible. This includes statements like, for example, "With brain injuries there can come impulses that cannot be controlled, misperceptions that lead the accused to many kinds of unacceptable behavior. How can 'intent' be formed by the broken brain?" (Qtd. In Kelkar, 2016). This statement introduces several interesting concepts, such as the relationship between a brain state and a mental state like "intent," and how this relationship changes ascriptions of responsibility.

This section has primarily served the purpose of *identifying* the conflict, but what remains to be done is *clarifying* the concepts at stake, and *evaluating* the conceptual frameworks. What remains to be fleshed-out is how law presupposes the brain-mind-body-world relation insofar as it is relevant to the legal notion of responsibility. It is clear that stakeholders are sensitive to the tension between neuroimages and traditional legal concepts, but that the arguments provided within the court are not sufficient to resolve this tension. As De Vos identifies:

> The key issue lies in discerning how the message carried by the brain image, 'look, this is what you actually are,' once it has permeated popular culture, not only invites us to identify with the icon, but also invites us to adopt the iconography. That is, our self-understanding and self-consciousness is solicited by both the images and the signifiers stemming from discourses on the brain. It is important to grasp that what one is actually being called upon to identify with

> here is not the brain image as such, the paradoxical Gestalt signaling the end of unity and agency, but, rather, the perspective of neuroscience itself. (De Vos, 2014:3).

While much has been said about the images themselves and how they are produced, less has been said about how these images help to disseminate neuroscientific modes of explanations and worldviews into the public. Brain images play an active role in instantiating people to think of themselves in terms of their brains. They not only make a way to visually connect to our brains, making them, in some sense, more present (the icon), but they also make a way to disseminate neuroscientific modes of understanding thought and behavior (the iconography). Aydin (2016) recognizes that, "Although... methodological reflections indicate that brain-images might not be reliable and valid pictures of brain process, they, at the same time, reinforce the view of the brain as a causal agent by framing the brain as an isolated realm" (5). Aydin acknowledges that neuroimages often reinforce a neuro-reductionive approach. The cases also demonstrate that, given a neuroimage, it is far easier to make a statement like: Hinckley, Weinstein, or Dugan was not responsible because his abnormal brain caused his behavior. These statements invite the stakeholders to consider: do all brains cause all behaviors and, if so, is anyone responsible? Your response to these questions depends on conceptual commitments. Those conceptual commitments also influence how you perceive of the conflict between neuroimaging and law, or if you perceive there to be one at all. In order to understand this conflict more thoroughly, and clarify it at the conceptual level which cannot be addressed in the courtroom, the next sections explore the conceptual foundations of the legal notion of responsibility and connect these notions to neuro-reductionism. By understanding this conceptual framework more thoroughly, we can clarify what conflicts and coherences exist.

CHAPTER 2: CLARIFYING THE PROBLEM

3: Conceptual Notions In Neurolaw

This section introduces what I identify as the foundational presuppositions for the traditional legal notion of responsibility. 3.1 introduces how I arrived at this taxonomy. 3.1.1-3.1.4 examines the legal understandings of causation, agency, mental states, and responsibility and free will. 3.2 summarizes the relevance of these concepts and connects them to neuroscientific discourse.

3.1. An Introduction to Relevant Concepts

Now that the reader has been sufficiently versed in some of the main methodological, practical,

and empirical conflicts, we will turn to the conceptual conflicts. In particular, the issue is how the neuroscientific explanations which accompany neuroimages are perceived to be in conflict with the traditional legal notion of responsibility. The previous section demonstrated that there is a clear tension among the stakeholders as to whether neuroimaging is ready to fulfill the roles that are expected of it in court. Deliberations on this question are often done in practical and empirical terms, but also invite consideration on conceptual standpoints which are only tacitly acknowledged. Conceptual questions like how Weinstein's arachnoid cyst defense framed the causal relationship between the mind and brain have yet to be asked, and furthermore, the question of how the framing of this causal relationship might conflict or cohere with traditional legal approaches has yet to be asked.

Outside the courtroom, these issues are allowed to expand into broader arguments, and become linked to broader concepts of causality, agency, free will, responsibility, and so on. So while a court case might only address whether fMRI is empirically robust enough to demonstrate the presence of a specific brain anomaly to support a specific argument about a reduced function, the public media surrounding the case can speculate far and wide as to whether such neurological explanations mean the end of free will and responsibility as we know it (i.e. Rosen, 2007; Davis, 2012; Kelkar, 2016; Griffin, 2016). What was a narrow, distinct issue within the court becomes a broad, multi-faceted issue as it is appropriated and re-appropriated by different stakeholders. Implicated philosophical issues come to the fore, provoked by these cases, but certainly not resolved by them. Morse (2015) argues, "About... the types of issues that mutually concern us [legalists], such as the causation of action (the mind-body problem) and the criteria for responsibility (compatibilism v. incompatibilism), metaphysical assumptions matter. The question is whether one must resolve or even defend one's metaphysical and other philosophical foundations in these fraught areas... I shall suggest, however, that when one's philosophical position is foundational and practically important, it must be acknowledged, but need not be defended or, a fortiori, resolved" (2). The aforementioned court cases provoke intuitions about these philosophical issues, and necessitate that the juror and concerned public reflect on these issues.

Morse notes two issues, the causation of action and the criteria of responsibility, both of which relate to the conceptual relationship between the mind, body, brain, and world. In the Hinckley case, for example, the causation of action comes into play in understanding the extent to which his statistically abnormal brain played a causal role in his behavior. His case also relates to the criteria of responsibility because there is a threshold at which he is no longer understood as responsible – where is this threshold? Is there any threshold? Morse also mentions the issue between compatibilism (the belief that determinism and free will co-exist) and incompatibilism (the belief that these do not coexist). This is

relevant because if one assumes an incompatibility stance, that there is no free will if there is determinism, and one accepts that in the Hinckley case, his behavior was determined by his abnormal brain – does this mean there is no free will? In keeping with Morse, I also do not intend to defend or resolve these issues myself, such a project is a lifetime's work, perhaps multiple. However, what I can offer is conceptual clarification of what relevant philosophical notions are presupposed by law, and how these cohere or conflict with the presuppositions of neuro-reductionists and distributed causalists.

There are several main conceptual issues that reoccur throughout the legal argumentation. For an individual to be held legally responsible, he/she must satisfy a number of preconditions. For example, he/she must be conceived of as causally proximate. The legal conception of causality constitutes the focus of Section 3.1. Furthermore, he/she must be an *agent*, the legal understanding of which I clarify in Section 3.2. The notion of *agency* presupposes the causal efficacy of mental states, like intentions, desires, and beliefs, as explored in Section 3.3. Together, these inform the notion of a freely-willing agent who is an appropriate subject for legal responsibility (Section 3.4). All of these elements are presupposed when deploying the traditional legal notion of responsibility, and they postulate a certain relationship between the mind, brain, body, and world. These issues are problematized when legal cases force us to confront the fundamental questions: what causes behaviors (particularly of the criminal kind) – the brain, the mind, some combination of both, neither, or more? And how does this bear on legal notions of responsibility and its presuppositions of free will, agency, and causally efficacious mental states? The interest in answering these questions cuts across law, neuroscience, philosophy, and just human individuals trying to figure themselves out.

3.1.1 Causation and Law

The concept of "causation" seems so obvious that it is practically self-explanatory, but when performing such an interdisciplinary approach, such meanings are not necessarily stable. All disciplines invoke some notion of "cause," whether it is how physical laws "cause" the movement of the universe, or how the sun "causes" me to wake up in the morning, or how the brain and mind "cause" human experiences. As Freeman (2006) writes, "The concept of causality is fundamental in all aspects of human behavior and understanding, which includes our efforts in laboratory experiments and the analysis of data to comprehend the causal relations of world, brain and mind" (89). "Cause" forms an integral part of any systematic attempt to explain a phenomenon, whether it be philosophical, scientific, or legal. Batts (2009) writes from the point of view of a neuroscientist, "To the neurosciences, it is the causality of behavior that is of the highest interest" (264). Indeed, every reader has some intuitive

notion of what causation is and the importance of invoking it in an explanation. We deploy informal notions of causation every day, and this continues in more formal domains. Rarely do we take it upon ourselves to explain what it is we mean by causation.

Only recently, some legalists have authored systematic philosophical approaches to causality in law. The most influential among them are the seminal works of Hart and Honoré (1959), *Causation and Law,* and Moore (2009), *Causation and Responsibility.* Honoré (2010) observes from the perspective of a legalist, "When rules of law attributing responsibility for harm caused are formulated in statutes, regulations and judicial decisions, the word 'cause' is often used. The notion that causal connection between agency and harm must be established is however often implied even when the word is not used... In all these instances the use of the notion of cause is central to the legal inquiry, since to establish responsibility it must be shown that the harm was done or brought about by the agency..." (para. 8). It has been argued that law might deploy notions of causation which are radically different from other disciplines, but both Honoré and Moore argue that law generally presupposes common-sense notions of causation. Moore (2009) advises, "It is better to think that 'cause' is univocal; it means the same thing in contexts of attributing responsibility as in contexts of explanation: it refers to a natural relation that holds between events or states of affairs" (5). So "cause" is not defined in an incomparably discipline-specific way, yet in practice, the causes one refers to in law may not be the same as in other disciplines. Certainly in the aforementioned cases, the causal linkage between the defendant and the crime was self-evident, for example, Weinstein's strangulation and subsequent defenestration of his wife was the clear cause of her death. However, whether his arachnoid cyst was the "cause" of his violence is much less clear cut.

For law, one must demonstrate that the defendant played an integral causal role before he/she can assume any legal responsibility. Although causality is a precondition for responsibility, it does not equate to responsibility. There is a difference between "allowing" harm and "causing" harm – although both can be understood as denoting that the defendant played some causal role in the outcome. For example, if a doctor does not prescribe contraception to a woman who births a serial killer, he/she may have played a causal role in the resulting state of affairs, but it would be absurd to hold him/her responsible. This makes it clear that there are boundaries on what is relevant causation for legal purposes. Law generally strives towards what Honoré (2010) terms "causal minimalism," attempting to rule out causes which are too remote to warrant legal action. Honoré shows that only factors which had a demonstrably significant causal influence can assume legal responsibility. He writes, "This sometimes requires all the limiting factors to be brought under a single umbrella... A number of

expressions are used to describe the allegedly single limiting factor, in particular 'proximate (adequate, direct, effective, operative, legal, responsible)' cause in contrast with 'remote, indirect or legally inoperative' causes" (Honoré, 2010: para. 16). Some ways to decide whether this is the case are to ask oneself if the outcome would have occurred without the defendant's actions, or whether the defendant's actions constituted a necessary element of a set of conditions which jointly brought harm. So it is clear that not all conceivable causal elements should (or even could) be taken into account in deliberating on legal responsibility. For example, Hinckley's therapist could not be put on trial for failing to ameliorate Hinckley's aggressive and obsessive behavior, although in another context, the therapist could be conceived of as part of a causal web of events leading to Hinckley's attempted assassination of the president. The therapist may have "allowed" the harm to happen by not intervening, but he/she did not "cause" it in the legal sense. However, it is also clear that some of the causal elements which *could* be taken into account, according to this view, are not. Namely, this "common sense" view of causation includes non-human things which play a "proximate" causal role in bringing about an undesirable state of affairs. If Weinstein had merely pushed his wife, but she had fallen off the balcony due to a faulty balcony, the defunct balcony may have played a causal role, but Weinstein would be the only actor eligible for an ascription of responsibility. Although other factors can be conceived of as causal, they are not held responsible, so another limiting factor must be at play. I propose that this further limiting factor is "agency," to which I will turn in the next section.

3.1.2 Agency and Law

Although "proximate" causality is a precondition for legal responsibility, it must coincide with another precondition: agency. In the literature on causation and law, notions of agency are inevitably invoked. Harm (or generally legally unwanted effects) must be caused by some sort of agency to warrant legal action. As Honoré (2010) demonstrates, "the notion of cause is central to the legal inquiry, since to establish responsibility it must be shown that the harm was done or brought about by the agency..." (para. 8), so causal proximity is only relevant insofar as it is a precondition an *agent* must satisfy in order to be responsible. So what is "agency"? Schlosser (2015) defines, "In very general terms, an agent is a being with the capacity to act, and 'agency' denotes the exercise or manifestation of this capacity" (para.1). However, if this were the case, then why would a machine not be considered a causal agent? The "standard conception" of agency, which underlies much of Western philosophy and law, places particular value on the role of "intentions." According to this view, since humans have the capacity to act in accordance to his/her intentions, he/she is understood to be an agent. So unless one conceives of machines as having intentionality, he/she typically does not conceive of them as being

legally or morally liable agents. Schlosser (2015) summarizes, "According to this view, a being has the capacity to exercise agency just in case it has the capacity to act intentionally, and the exercise of agency consists in the performance of intentional actions and, in many cases, in the performance of unintentional actions (that derive from the performance of intentional actions...)" (Schlosser, 2015:para. 10). He emphasizes that having the capacity for agency is not the same as exercising agency. This emphasizes the important legal issue one can have all the requisite capacities for causal agency, he/she can be generally conceived of as a causal agent, without having exercised causal agency in a specific state of affairs. For example, in a negligence case, the individual might not have any mitigating factors to why he/she was not an agent in general, but did not have the requisite intentions and actions at that particular moment. So both causality and agency, even in combination, form a necessary, but not sufficient, condition for ascribing responsibility.

The "standard conception" of agency also comes with a corollary "standard theory," which seeks to explain how humans come to act as agents. This relates to how we conceive of action, and the relationship between action and mental states like intending or willing. Schlosser (2015) summarizes, "The standard theory of action provides us with a theory of agency, according to which a being has the capacity to act intentionally just in case it has the right functional organization: just in case the instantiation of certain mental states and events (such as desires, beliefs, and intentions) would cause the right events (such as certain movements) in the right way. According to this standard theory of agency, the exercise of agency consists in the instantiation of the right causal relations between agent-involving states and events" (para. 10). This theory entails making a number of other conceptual commitments, for example, a belief that mental states and events exist and have causal efficacy in producing behaviors. This also entails that when behaviors are caused by intentional, reasoned mental states, they are defined as "actions." Furthermore, this means that notions of rationality and agency are integrally related, as they both presume that humans the defining feature of "action" is that it is a product of reasons, intentions, desires, or beliefs. Davidson (2001) coheres with this "standard theory" when he writes, "In the light of a primary reason, an action is revealed as coherent with certain traits, long- or short-termed, characteristic or not, of the agent, and the agent is shown in his role of Rational Animal" (9). Davidson defines "actions" in terms of agency, or in terms of having the capacity to act according to reasons, beliefs, desires, and intentions. This also demonstrates the strong linkage between how we conceive of agency in relation to conceptions of what it means to act and have rationality. This normative element is apparent in the Hinckley and Weinstein cases, where they were both understood to be acting in accordance with their intentions. There was no mistake about it that Hinckley and

Weinstein intended to kill their victims, and that their actions were in accordance to their intentions. However, the breakdown in their agency occurs at the level of their intentions displaying the requisite relationship with the world.

3.1.3 Mental States and Law

The "standard conception" and "standard theory" of agency both presuppose the causal efficacy of mental states. Since law typically adopts this stance, it also presupposes the causal efficacy of mental states. As Schlosser (2015) relates, "The view explains agency in terms of the agent's desires, beliefs, and intentions. Usually, it is assumed that this is an explanation in terms of mental representations: in terms of intentional mental states and events that have representational contents (typically, propositional contents)" (para. 17). This conception of agency imbricates the "Representational Theory of Mind" (RTM), a foundational theory in cognitive science which underlies most discussions regarding mentation. Pitt (2013) broadly defines RTM as, "any theory that postulates the existence of semantically evaluable mental objects, including philosophy's stock in trade mentalia — thoughts, concepts, percepts, ideas, impressions, notions, rules, schemas, images, phantasms, etc. — as well as the various sorts of 'subpersonal' representations postulated by cognitive science" (para. 3). Most ideas about mentation presuppose that there are some kind of distinct mental states which are available to analysis and articulation, generally presupposing an RTM stance. Furthermore:

> RTM... takes as its starting point commonsense mental states, such as thoughts, beliefs, desires, perceptions and imagings. Such states are said to have 'intentionality' — they are *about* or *refer to* things, and may be evaluated with respect to properties like consistency, truth, appropriateness and accuracy. (For example, the thought that cousins are not related is inconsistent, the belief that Elvis is dead is true, the desire to eat the moon is inappropriate, a visual experience of a ripe strawberry as red is accurate, an imaging of George W. Bush with dreadlocks is inaccurate.) (Pitt, 2013: para. 4)

This presupposes that mental states are integrally related to events in the world, and represents them with varying degrees of reliability (which are available to evaluation). This means accepting some degree of realism about the world. Hansen (2000) adopts a similar stance, also identifying RTM as the predominant conceptual framework. He summarizes it as such:

> (1)Mental phenomena have certain characteristic properties or features that they don't share with non-mental phenomena: the distinctive feature of a sensation, for example, is its qualitative aspect or subjective feel; a belief, to take another standard example, has intentional content.

(2) Mental phenomena are firmly embedded in the causal nexus of the world: they enter into all sorts of causal relations with each other and with non-mental states. (Hansen, 2000:453).

His characterization shares much in common with Pitt's. Mental states are presumed to exist, related to physical states and things but not reducible to them, and engaged in an interplay of causal relationships. Schlosser (2015) adds, "The question of what the possession of representational mental states consists in is one of the most controversial questions in the philosophy of mind and cognitive science, and it is clearly beyond the scope of this entry" (Schlosser, 2015: para. 18). Similarly, it is beyond the scope of my entry. What is important to know is that RTM enables a form of realism which presumes that our representational content (and subsequent mental and intentional content) relates, in some way or another, to things in the world and has a causal effect on human thought and behavior.

Just as law assumes traditionally assumes a "standard" notion of agency, it also subscribes to the complimentary belief in mental representation. If mental states were not representational, they would not be causally efficacious nor would they be available to evaluation in terms of their truthfulness, accuracy, consistency, etc.. Furthermore, if they were not considered representational, the law would likely not invoke them at all and the whole "mind-reading" enterprise would lose its appeal. As is, the *mens rea* part of the crime is equally important as the *actus rea,* and some crimes refer entirely to mental representation, like Hinckley's crime, which was not actual assassination, but intent to assassinate. Pitt (2013) describes the *desire* for Elvis' death as a mental representation of the proposition "Elvis is dead," so similarly, Hinckley's *desire* for Reagan's death was a mental representation of the proposition "Reagan is dead." Many scholars accept that, "What a person believes, doubts, desires, fears, etc. is a highly reliable indicator of what that person will do; and we have no other way of making sense of each other's behavior than by ascribing such states and applying the relevant generalizations. We are thus committed to the basic truth of commonsense psychology and, hence, to the existence of the states its generalizations refer to" (Pitt, 2013:9). Not all scholars are committed, but with law's commitment to traditional notions of causal agency, it is also committed to an RTM in which having the intent to commit a crime is almost as egregious as committing it. Law would likely not place such high value on mental states if it did not assume that they were representational, available to evaluation, and causally efficacious.

References to representational, mental, and intentional states often fringe on discussions regarding consciousness. In general, the mental states of which we are conscious are pervasively representational (i.e. they refer to something), and vice versa, the mental states we deem representational are most often those of which we are conscious. Humans can only come to know,

recognize, or taxonomize mental states through having conscious experiences, so in that sense the two are tightly interrelated. Additionally, "consciousness" might, itself, be a mental state, or array of mental states. As Noë (2009) defines it, "To have a mind is, roughly, in my sense, to be conscious—that is, to have experience and to be capable of thought, feeling, planning, etc" (10). Particularly when authors talk about "the sense" of agency, they refer to consciousness. Daniel Wegner is one of the most prominent social psychologists analyzing the experience of agency. Wegner (2003) suggests, "conscious will is experienced when we draw the inference that our thought has caused our action" (67). Since the conscious experience of willing is foundational to experiencing oneself as an agent, consciousness often becomes an object of inquiry in such discussions. Additionally, Wegner suggests that we experience ourselves as conscious, intentional agents when we presume that our thoughts cause our actions. So inquiries into agency and causation both border on inquiries into consciousness, particularly when talking about the *experience* of being a causal agent.

Since this thesis is not about diverse notions of consciousness, but about diverse notions of causal agency, there is a specific way in which consciousness relates to the issue at hand. Freeman relates these elements when he writes:

> What is consciousness? It is known through experience of the activities of one's own body and observation of the bodies of others. In this respect, the question whether it arises from the soul (Eccles 1994), or from panpsychic properties of matter (Whitehead 1938; Penrose 1994; Chalmers 1996), or as a function of brain operations (Searle 1992; Dennett 1991; Crick 1994) is not relevant. The pertinent questions are—however it arises and is experienced—how and in what senses does it cause the functions of brains and bodies, and how do brain and body functions cause it? How do actions cause perceptions; how do perceptions cause awareness; how do states of awareness cause actions? Analysis of causality is a necessary step toward a comprehension of consciousness. (Freeman, 2006:74).

Notions of what consciousness is, how it comes to be, and what kind of causal role it plays in thought and behavior are perhaps as old as the human species, as everyone experiences his/her own consciousness and continually makes predictions and explanations invoking the consciousness of others. Laypeople, in addition to many scholars, do not talk in terms of "mental representations" and "intentional states," but rather use "consciousness" in a similar way. In order to remain as inclusive as possible, I include literature which inquires into consciousness, but specifically into the causal relationships of consciousness. Wegner (2003), for example, begins an article by asking, "Does consciousness cause action?" (65), and responding that most people would find this question absurdly

obvious. The intuitive notion that consciousness causes action clearly relates to the conceptual framework that intentional states cause, and even define, action. In the context of this thesis, discussions regarding the causal efficacy consciousness, intentional states, mental states, and related mentalia are all relevant. Most individuals assume that his/her conscious thoughts, decisions, and intentions play a causal role in his/her behavior, and by this, cultivate his/her sense of causal agency. Some scholars argue that ascribing causation to conscious, intentional, representational, or otherwise mental states is "common sense," (i.e. Rose, 2005;Morse, 2011), while others term it "folk psychology" (i.e. Churchland, 1995). Nonetheless, all scholars agree that these traditional notions of causal agency underly and inform the law.

3.1.4 Responsibility and Free Will

Recurring throughout these conceptual discussions are mentions of responsibility and free will. These topics bridge the abstract notions of philosophy to the practical dealings of law. There is a larger body of pre-existing scholarship and media relating neuroscience or neurolaw to conceptual notions regarding responsibility (i.e. Morse, 2005; Eastman & Campbell, 2006; Mobbs et al., 2007; Aharoni et al., 2008; Batts, 2009; Mayberg, 2010; Vincent, 2011) and free will (Glannon, 2005; Gazzaniga, 2011; Schleim, 2012; Bear, 2016; Aydin, 2016, Griffin, 2016). What I can add to this literature is supplementary connections between these topics and demonstrate how these notions co-shape the legal notions of responsibility. I have grouped these topics together because they are so closely related, but before getting into how they relate to one another, I will clarify what I mean by "legal responsibility." It is difficult to find a definition of this which does not refer to "moral responsibility" because legal responsibility is, in many ways, constructed as a way to maintain and enforce moral responsibility. "Moral responsibility" can be broadly defined as such: "To regard such agents as worthy of one of these reactions is to regard them as responsible for what they have done or left undone... Thus, to be morally responsible for something, say an action, is to be worthy of a particular kind of reaction—praise, blame, or something akin to these—for having performed it" (Eshleman, 2014:para.1). "Legal responsibility" is based in this notion, but only addresses those acts which warrant blame or punishment. This means there are some acts for which you are morally responsible without being legally responsible, for example, those acts which warrant praise and reward, because law is generally concerned with punishment for harm or undesired outcomes. Also some acts incur legal responsibilities without posing much moral responsibility, such as repercussions for crimes which did not harm or

cause any damage. Nonetheless, these ideas are integrally related, especially in criminal law where the crimes tend to be of a legal and moral nature. Even a victimless crime is considered, on an abstract, societal level, to cause harm or morally reprehensible outcomes (i.e. not paying taxes means that, indirectly, someone is not receiving adequate welfare), but the definitions of legal and moral responsibility overlap the most when a victim is involved. Legal responsibility in criminal law, at its most basic, is only concerned with the punishment and blame aspects of moral responsibility.

There are two primary theories of justice which underly how legal responsibility is allocated and why such a notion exists. They are: consequentialist and retributivst. Consequentialist justifications claim that the benefit to legal punishment is to deter future wrongdoings, while retributivist justifications claim that the benefit to legal punishment is that an individual who breaks societal norms could have done otherwise and therefore deserves punishment (Hart, 2008:1). Retributivism is generally understood to be the dominant ideology in American legal theory, and Greene and Cohen (2004) describe it as such: "Retributivism captures the intuitive idea that we legitimately punish to give people what they deserve based on their past actions-- in proportion to their 'internal wickedness', to use Kant's (2002) phrase" (3). Authors have acknowledged that in practice, judges and juries make judgements based on consequentialist and retributivist grounds but that the prevailing legal notion of responsibility and punishment is grounded in retributivism (Hart, 2008; Greene and Cohen, 2004; Snead, 2007). The current notion of legal responsibility is grounded in this retributivist approach, wherein certain people are presumed to deserve certain punishments because they were free agents who could have done otherwise, and chose to break the law. Therefore, the very grounding of the legal notion of responsibility invokes free will.

The notion of responsibility is, in the retributivist context, grounded by the notion of free will. McKenna and Coates (2015) defines moral responsibility in relation to agency and free will, "A person who is a *morally responsible agent* is not merely a person who is able to *do* moral right or wrong. Beyond this, she is *accountable* for her morally significant conduct. Hence, she is, when fitting, an apt target of moral praise or blame, as well as reward or punishment. And typically, free will is understood as a necessary condition of moral responsibility since it would seem unreasonable to say of a person that she deserves blame and punishment for her conduct if it turned out that she was not at any point in time in control of it" (para. 3).One of the main issues Morse (2015) outlined in his "metaphysical" issues underlying law is "the criteria for responsibility (compatibilism v. incompatibilism)" (2). Compatibilism and incompatibilism are two approaches to free will; as summarized earlier, the former means that free will and determinism can coexist, while the latter denies that they can coexist. Free will

debates are frequently provoked by neuroscientific explanations which explain at least certain behaviors in terms of determined physical/mechanical processes in the brain. If one accepts that one behavior was determined, does that mean all behavior is determined? Is it possible for determined and undetermined processes to coexist?

Although these can be understood as general stances on free will, they also directly relate to an individual's stance on responsibility. The aforementioned insanity defense cases show that when someone's actions are understood to be determined (in this case, by a brain abnormality), he/she is not conceived of as a freely willing agent, and therefore he/she is not held responsible. If all thought and behavior is understood to be determined by physical/mechanical brain processes, can responsibility even exist? This question bears on another relevant framing of "compatibilism" versus "incompatibilism." As Eshleman (2014) summarizes:

> In keeping with this focus on the ramifications of causal determinism for moral responsibility, thinkers may be classified as being one of two types: 1) an *incompatibilist* about causal determinism and moral responsibility—one who maintains that if causal determinism is true, then there is nothing for which one can be morally responsible; or 2) a *compatibilist*—one who holds that a person can be morally responsible for some things, even if both who she is and what she does is causally determined. (para. 10).

Neurolaw is provocative on this front because it introduces increasingly deterministic frameworks for explaining human behavior. Various stakeholders are forced to confront the question of whether free will makes sense when human thought and action is increasingly explained in terms of determined processes occurring in the brain. They can arrive at a number of different conclusions – for example, intuiting that Hinckley, Weinstein, or Dugan's action was determined by their significantly abnormal brain, but within the limits of a normal brain, one has the capacity to exercise free will and therefore responsibility. This compatibilist approach to responsibility has been endorsed by scholars like Churchland (2004) and Glannon (2005). An incompatibilist approach to this issue would either say that all of these defendants were, in fact, not determined by their brain abnormalities and acted as free willing agents – therefore they are responsible. The general public outcry after the Hinckley case shows an instance of that happening. On the other side, you could argue that their actions were determined, and therefore all actions is determined, and therefore the legal notion of responsibility does not make sense. This position has been endorsed by scholars like Greene and Cohen (2004). Free will can be conceived of in a number of ways, but it is generally understood as a unique feature of agency and a

precondition for responsibility, so if we change or abandon the notion of free will, we must change or abandon its imbricated notions of agency and responsibility.

Several authors have already made connections between notions of causal agency, responsibility, and free will. McKenna and Coates (2015) define, "what philosophers working on this issue [free will] have been hunting for is a feature of agency that is necessary for persons to be morally responsible for their conduct... As a theory-neutral point of departure, then, free will can be defined as *the unique ability of persons to exercise control over their conduct in the manner necessary for moral responsibility"* (para. 2). For these authors, free will is a feature of agency, and together these factors enable the ascription of responsibility. Other authors cohere with the notion that agency and free will are integrally related and form a precondition upon which responsibility can be ascribed (i.e. Choudhury and Blakemore, 2006; Rose, 2005). Incompatibilists tend to define free will more strictly, for example, "free will requires the ability to do otherwise" (Glannon, 2005:69) or a choice with "no causal antecedent" (Churchland, 2004:6). The notion of responsibility typically presupposes some degree of free will and agency, and since agency presupposes the causal efficacy of mental states, a stance on responsibility also entails a stance on mentation. The law generally takes a compatibilist stance, acknowledging that some causation is determined, and some causation is the product of agency, and therefore undetermined. Primarily the agent-bound, undetermined causation is relevant for ascriptions of legal responsibility.

Since the "standard conception" of agency includes intentions as a defining feature, so too does the "standard conception" of legal responsibility. The law invokes intentions and mental states as necessary elements in deliberating on legal responsibility precisely because it presupposes that these are causally efficacious. It presupposes the causal efficacious of mental states because that goes hand-in-hand with accepting the "standard conception" of agency. Moore (2011) elaborates two ways in which the notion of "intentions" is central to both moral and legal responsibility. He beings, "The first is as a marker (arguably *the* marker) of serious culpability in the doing of wrongful actions. As the law... recognizes, doing some wrongful action because one intended to do it merits greater blame and more severe sanctions than does doing that same wrongful action recklessly or negligently..." (207). This first issue relates back to the difference between being a causal agent in general and exercising causal agency in a specific moment. In order to have any responsibility, you must be conceived of as a causal agent *in general*, hence why people who successfully plead legal insanity are not (technically) considered responsible (public sentiment may indicate otherwise, as it did with the Hinckley case). In order to have *full* responsibility, thought, you must have been exercising causal agency (acting in a

proximately causal way with the requisite intentional states) in that particular moment. Moore (2011) adds, "The second way in which intention figures into attributions of responsibility has to do with wrongdoing rather than culpability. To do wrong is to *act* in a way that morality or the law prohibits, and intentions are at the root of action and agency... [That] is to say that every action begins with an intention, in the sense that intentions must be the immediate cause of those bodily movements through which persons act, for those movements to be actions at all" (Moore, 2011:207). This coheres with the "standard theory" of agency in which action without intention is better understood as mere behaviors, for which responsibility is out of the question. However, anything that *can* be considered an act must have some underlying intentions guiding it. So any legally or morally reprehensible act, even those of recklessness or negligence, can be understood in terms of intentions. This demonstrates the extent to which causation, agency, mental states, responsibility, and free will are all co-contingent theoretical constructs which shape and inform one another, as well as the practical and empirical arguments that conceptualize them in specific ways.

3.2 Summary

My research questions asks how to reconcile the legal notion of responsibility with emerging neurotechnologies and neurological frameworks, and this section serves to clarify what is at stake in the legal notion of responsibility. This section enumerates the many presuppositions this notion contains which ultimately posit several key conceptual commitments. It demonstrates that in the law's commitment to retributivist logic, it necessarily presupposes some degree of free will. It is also apparently able to take a compatibilist stance on responsibility and free will, in which Hinckley or Weinstein's behavior was understood to be the product of determined processes in the brain, but supposedly, the absence of any major brain abnormality or "mental defect" should mean the full capacity to exercise free will and agency. The law also takes what could be called a compatibilist stance regarding brain states and mental states, positing that both can be understood as causally efficacious, given the right set of evidence and argumentation. It seems, then, that emerging neurotechnologies should not pose any particular threat to these notions. Yet many stakeholders, both in popular media and in academia, seem to put these frameworks at odds, reflected in titles like "Neuroscientific Challenges to Free Will and Responsibility" (Roskies, 2006) and "Brain Overclaim Syndrome and Criminal Responsibility" (Morse, 2005). It is difficult to find anyone from any side of this debate who says outright that human society should jettison the very notion of responsibility – I have yet to

encounter anyone in my research making this argument. Even those perceived to be the most neuro-reductionist still argue that some notion of legal responsibility can be preserved, so it is clear that the challenge to this notion is not an outright one.

I posit, therefore, that the perceived challenge of neuroimaging technologies can be understood in two steps: the first is that with neuroimaging comes neuro-reductionism, and the second is that neuro-reductionism does not undermine the broad notion of responsibility, but it undermines the specific presuppositions of the traditional legal notion of responsibility. The neuro-reductionist challenge, I argue, is at the level of its presuppositions – neuro-reductionism casts doubt on some of the elements which constitute the traditional legal notion of responsibility, so while some notion can be preserved, it is not the same. Specifically, the legal notion of responsibility invokes the causally efficacy and representational realism of mental states. The notion that the mind is a causally efficacious element is presupposed in the "standard conception" of agency and free will which underlies the retributivist approach to responsibility.

Nonetheless, the necessary issue of inquiring into mental content is the most challenging part of criminal law, and therefore it makes sense to call on new practices and technologies which may be able to settle the debates between competing psychological explanations. Technologies like fMRI enter the court to settle these debates, but bring with them a new vocabulary, the vocabulary of brains. The legal vocabulary, as these past sections has demonstrated, is intimately concerned with the mind, and makes no mention of the brain. Cognitive science established a link between the mind and the brain, which law quickly appropriated, and now the concerned public is in the process of deliberating what these new linkages could mean. In their specific contexts, neuroimages do not seem to pose any immediate threat to the legal notion of responsibility due to its compatiblist stance. Neuroimages are understood to "read the mind," presupposing that the mind exists. However, when neuroscientific explanations for thought and behavior are taken more broadly, they are understood to undermine the conceptual commitments of legal theory. Taken to an extreme (as the public often does), they can mean that there is no mind, or that the mind is just the product of determined processes in the brain and does not have any causal efficacy. Pardo and Patterson (2013) go so far as to claim that the neuro-reductive standpoint is pervasive in this field, claiming, "If anything unites the various problems and projects of neurolegalists, it is the belief that the mind and the brain are one. This belief is a pervasive feature of much of the current research in neuroscience and the neurolaw literature as well as more popular writing" (20). As the amount of neuroscience in law increases, so too does the amount of brain-based explanations for thought and behavior. From this standpoint, "the mind" and "mental states" can be

understood as non-existent or rendering no causal agency at all, relegated to the status of "epiphenomenon" or "psychic appendixes that evolution has created but that have no genuine function" (Morse, 2011:219). So while neuroscience enters the courtroom to bolster traditional legal notions regarding agency and responsibility, and therefore investigate "mental" phenomenon, when the "brain facts" leave the courtroom and become appropriated by other scholars and pundits, they are often interpreted to doubt the existence of "mental" phenomenon at all, and therefore mount a significant conceptual challenge to these notions. As I have demonstrated, law frequently invokes (explicitly or implicitly) a notion that mentation exists, is causally efficacious, and co-constitutes our status as freely willing, responsible agents. If the thesis of brain-based causality is accepted, and, for example, all thought and behavior is understood to be determined, the implications for legal theory could be far-reaching and revolutionary. Additionally, brain-based explanations have, thus far, been bolstered by the proliferation of neuroimaging technologies, and the more they enter the courtrooms, the more their corollary brain-based explanations are reified in the public eye.

4: Why is Neuro-Reductionism Perceived as a Threat to Legal Responsibility?

This section is devoted to exploring neuro-reductionism and its perceived conflict with legal responsibility. 4.1 begins by characterizing the standpoint of neuro-reductionism. 4.2 explores Wegner's "weak form" of neuro-reductionism, with 4.2.1 demonstrating how his ideas have been appropriated and 4.2.2 demonstrating how this mounts a challenge to legal responsibility. 4.3 and 4.4. follow the same structure for Churchland's and Greene's frameworks, respectively. 4.5 summaries the relevance of this discourse and its broader impact on the tensions between neuroimaging and law. 4.6 moves to critiques of this framework.

4.1. Characterizing Neuro-reductionism

As the cases demonstrate, the general in-road for fMRI into the courtroom relies on some variation of the argument that "my brain made me do it." In other words, neuroimages are often introduced with arguments that presuppose the brain was the primary, if not only, causal factor behind a criminal act. Recent history demonstrates that this has been a difficult position to successfully deploy in the courts just at the empirical and practical level, but it is also perceived by stakeholders to post conceptual challenges as well. It seems not far to make the conceptual leap from "may brain made me do it" to "you are your brain" (Greene, qtd. In Rosen, 2007:3). Explanations which characterize individuals in terms of their brain are often backed by theoretical frameworks which propose to explain

all the phenomenon of mentation in terms of processes occurring in the brain, hence why I, and other scholars, refer to these perspectives as "neuro-reductionist." Pardo and Patterson (2013) summarize neuro-reductionists as those who "aspire to explain the mind and mental life by 'reducing' them to the brain and states of the brain. We illustrate the neuro-reductionist conception of mind that underlies much neuroscientific research and the proposals for its increased use in law" (23). As the review of cases demonstrated, the claim that the brain is the primary causal factor can serve many legal agendas, especially when paired with compelling neuroimages. Neuro-reductionism also poses a practical benefit for neurotechnologies: if the mind *is* the brain, then brain-imaging technologies like fMRI are more capable of providing authoritative evidence on the mental states posited by law. However, if the mind *is* the brain, does the notion of mind even make sense anymore?

In order to explore this question, I will first enumerate different approaches to neuro-reductionism (4.2-4.4). I use three primary exemplars: Joshua Greene, Patricia Churchland, and Daniel Wegner because they are among the most outspoken and frequently-cited scholars articulating a neuro-reductivist position in the context of law.They demonstrate a spectrum of neuro-reductionist approaches which range from complete negation that the mind even exists to belief that the mind exists but is epiphenomenal (non-causal) to the belief that it exists and is causal, just not as causal as we tend to think it is. These alternative approaches demonstrate that taking a neuro-reductionist stance does not *necessarily* mean negating the legal notion of responsibility, and each author arrives at a different approach to the implications for law. The next sections (4.2.1, 4.3.1, and 4.4.1) moves from what these authors actually say to how they are read, understood, and deployed by others. As in most disciplines, what starts off as a relatively modest claim gets appropriated and re-appropriated into more simplified, and generally more exaggerated, forms of the original ideas. These appropriations have as real an effect as the original works, perhaps even more when they take the form of popular media articles that reach far more readers than the academic work. Additionally, the interpretations are generally more radical than the original work, and make a more striking juxtaposition with traditional legal notions. What is ultimately a challenge at the level of one presupposition (i.e. mental states are not as causally efficacious as we tend to believe) is often perceived to be a challenge on the level of the whole concept of legal responsibility. Before enumerating this conflict, I will first shed some light on the alternative approaches to brain-based explanations of thought and behavior which could be understood as neuro-reductionist. I will go from a more "weak" for of neuro-reductionist in which mental states are just understood to play a *lesser* role in cognition (Wegner), a stronger form in which mental states are understood to be largely epiphenomenal (Churchland), and the strongest form in which mental states

are understood as entirely inappropriate phenomenon into which to inquire (Greene).

4.2 Wegner's Neuro-reductionism

Daniel Wegner is a social psychologist at Harvard university. He is highly influential in the field for his works on the conscious ascription of willing a behavior. Wegner seems to border on the conceptual territory of neuro-reductionism, arguing that the notion that conscious mental states have causality is often illusory – apparent in the title of his article "The Mind's Best Trick: How We Experience Conscious Will" (2003). At first glance, his argument resembles that of Greene and Churchland's – he deems that often the attribution of causal efficacy to a mental state is mistaken, and that more likely, other causal factors determined the resulting thought or behavior. At another glance, they are quite different because Wegner is a social psychologist using the language of the mind while Greene and Churchland are neuroscientists using the language of the brain. However, they have one crucial characteristic in common: their argumentation casts doubt on the causal efficacy of mental states.

All three of these academics invoke the same seminal set of works: the Libet experiments. Certainly the most frequently-cited experiments, among skeptics and optics alike, are Libet's experiment on movement.The *Stanford Encyclopedia of Philosophy* introduces Libet with the remark, "The most influential empirical challenge concerning the role of conscious intentions stems from Libet's seminal neuroscientific work on the initiation of movements" (Schlosser, 2015: para. 50). His work is mentioned by many of the authors cited throughout this thesis, including Wegner (2003), Rose (2005:1002); Gazzaniga, (2006:10); Mobbs et al., (2007:695); Aharoni et al., (2008:148); Pardo & Patterson, (2013:126); Aydin, (2016:3). His experiment has been reproduced several times, and analyzed from various different angles, but all investigating the same phenomenon: do we make decisions consciously or are they the product of unconscious machinations? The typical approach is that subjects are instructed to make some small actions at any time(s) of their choosing during the scan. They were also instructed to push a button when they had made the decision to make this action. The Libet experimenters found what they termed a "readiness potential," or an activation of particular brain activity *before* the subject felt like he/she had made a conscious decision. The interpretations of this result vary, but it is generally taken as such: "This result, which has been reproduced and extended by independent groups (Haggard and Eimer, 1999; Soon et al., 2008; Fried et al., 2011; Rigoni et al., 2013a), seems to indicate that consciousness about a movement decision arises only after the decision has been made by unconscious neural processes" (Guggisberg & Motaz, 2013:1). Libet's experiments

have inspired a number of different experiments which use different methodologies and analytic angles but all basically ask the same thing: does a conscious decision cause action, or is an action the result of determined neural processes?

Wegner's experimentation and argumentation is often placed alongside Libet's. For example, in Pockett et al.'s book *Does Consciousness Cause Behavior?* (2006), they introduce Wegner and Libet as the two most influential scientists provoking us to reconsider the role of conscious intentions in behavior. These experiments both cast doubt on the ascription that a conscious intention or decision *caused* an action. Wegner not only performed experiments to this effect, but also demonstrated that a number of observed phenomenon could also be more readily explained with this framework. He argues that, for example:

> We might understand... Penfield's classic finding on movements induced through electrical stimulation of the motor cortex. Conscious patients were prompted by stimulation of the exposed brain to produce movements that were not simple reflexes and instead appeared to be complex, multi-staged, and voluntary. Yet, their common report of the experience was that they did not 'do' the action, and instead felt that Penfield had 'pulled it out' of them. This observation only makes sense if the experience of will is an addition to voluntary action, not a cause of it. (Wegner, 2003: 65).

Wegner argues that the ascription of conscious will as the cause of actions only makes sense if the actions correspond with the will, but many examples, like that of Penfield's, indicate otherwise. They show a breakdown between the action and the experienced will. He introduces other examples like *alien hand syndrome*, wherein an individual feels that the movements of a part of the body are not the result of his conscious will. He observes, "On the one hand (pun couldn't be helped), the alien hand seems to do some fairly complicated things, acts we might class as willful and voluntary if we were just watching and hadn't learned of the patient's lamentable loss of control... On the other hand (as the pun drags on), however, the patient does not experience these actions as consciously willed. One patient described the experience as a feeling that 'someone from the moon' was controlling her hand" (Wegner, 2002:5-6). This indicates that maybe our ascriptions of the agency behind an action are not particularly accurate. He even cites the eerie phenomenon of oujia boards that seem to move of their own accord as an example of such a mismatch between the ascription of will and the action (Wegner, 2002:7-8). He describes that we can *do* something without feeling like we are the agents of it (Automatism), and we can also not do something and *feel* like we are the agents of it (Illusion of Control) (Wegner, 2002:8).

Wegner has also conducted experiments of his own in order to demonstrate that the ascription of

causation to conscious states is often mistaken. He argues that humans only experience will when the ascribe a conscious thought as the cause of an action, but this is an inference which may or may not be accurate (Wegner, 2003:65). He argues that we infer that our thought was the cause of an action when it satisfies three principles (based on David Hume's general principles on cause and effect), but these principles do not ensure that this inferences is accurate. Pockett et al. (2006) efficiently summarize them as such: "we think that something causes something else if and only if what we think of as the causal event occurs just before what we think of as the effect (the priority principle), is consistent with the putative effect (the consistency principle), and is the only apparent cause of the putative effect (the exclusivity principle)" (2-3). Wegner and Wheately (1999) tested this intuition in an experiment called the *I Spy* study, wherein two subjects moved a cursor over an image depicting fifty small objects. Each subject was instructed via separate sets of headphones to perform different movements and, in specific intervals, stop their cursor. Afterwards, they would rate each stop from a scale between "I allowed the stop to happen" and "I intended to make the stop." Ultimately, as Pockett et al. summarize:

> they [the subjects] proved to be quite bad at telling whether they or the experimenter had caused the cursor to stop. When the subject had really caused all of the stops, the average intentionality rating was only 56 percent... It was 56 percent—the same as if they really had caused the stops themselves—if they heard the name of the object either 5 seconds or 1 second before each forced stop. These results were interpreted as showing that subjects could be fooled into wrongly believing that that they had caused the cursor to stop... if the subject simply heard the name of the object just before the cursor stopped. (Pockett et al., 2006:3).

This seems to support their proposition that causal ascriptions can be misled, particularly when the causation appears to cohere with the three aforementioned principles. This experiment mainly tested the "priority principle," showing that if the subject heard a stimulus right before he/she made an action, he/she was more likely to ascribe conscious causation to the action.

From this, and other experiments, Wegner created his "model of apparent mental causation," in which he proposes that conscious mental causation is much rarer than we tend to believe. Since we ascribe causation based on principles which do not ensure the accuracy of this ascription, we are easily misled. He arrives at a model[8] in which unconscious causes of action are far more pervasive than conscious causes (Wegner & Wheately, 1999:483). By casting doubt on the causal efficacy of conscious mental states, it seems that Wegner also casts doubt on the "standard concepton" of agency, which presupposes that mentalia like intentions, beliefs, and desires play significant causal roles in our

[8] For a visual schema of Wegner's model, refer to Appendix B.

action. However, Wegner's research is about "the sense of agency," and not necessarily "agency" itself. He posits that we *experience* ourselves as agents when we posit that our conscious thoughts were the cause of our action, but that often this ascription is mistaken. He concludes "*the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action*. Conscious will, viewed this way, may be an extraordinary illusion indeed" (Wenger, 2002:2). He does not contend that there is no possibility for agency, or that conscious mental states *cannot* exercise causality. In this sense, his claims are more moderate than Greene and Churchland, hence why he constitutes a "weak" form of neuro-reductionism.

In fact, neuro-reductionism might not even be an accurate characterization because Wegner is still concerned with the content of "the mind," so why have I paired him with Greene and Churchland? His argument does not amount to saying mental states are epiphenomenal or nonexistent, but rather that unconscious mental states are more causally efficacious than conscious mental states, despite deep-seeded beliefs to the contrary. However, the ramifications of this experiment are often interpreted to cast doubt on free will, and as noted already, free will is foundational to the legal notion of responsibility. Although he actually resists determinism in the introduction of his book *The Illusion of Conscious Will,* it is perhaps easy to see how his works get appropriated as such. By saying that most action is caused by unconscious states over which we have no control, he seems to suggest that much of our action is determined by processes which have already occurred. If one takes an incompatiblist approach, the notion of free will is rendered invalid. Even if one takes a compatibilist approach, it seems that his model does reduce the opportunities in which one could be conceived of as exercising free will.

4.2.1 Appropriations of Wegner's Ideas

Serendipitously, an article appeared on my Facebook newsfeed recently boldly declaring "Free Will Could All Be An Illusion, Scientists Suggest After Study Shows Choice May Just Be the Brain Tricking Itself" (Griffin, 2016). The subtitle reads: "Research adds to evidence suggesting 'even our most seemingly ironclad beliefs about our own agency and conscious experience can be dead wrong'" (Griffin, 2016). Cited among this research are the experiments of Wegner, but the article focuses on a Wegner-inspired experiment performed by Bear and Bloom (2016). Since Wegner's experiments are, in turn, inspired by Libet, there is a clear linkage of conceptual frameworks between all of these experiments. They all seek to demonstrate that the feeling that something was consciously willed – in other words, caused by a conscious, intentional mental state and not an unconscious brain state – is an

attribution which occurs after a decision has already been executed. Bear summarizes their research findings in a blog entry for *Scientific American:*

> Suppose, as Wegner and Wheatley propose, that we observe ourselves (unconsciously) perform some action, like picking out a box of cereal in the grocery store, and then only afterwards come to infer that we did this intentionally. If this is the true sequence of events, how could we be deceived into believing that we had intentionally made our choice *before* the consequences of this action were observed? This explanation for how we think of our agency would seem to require supernatural backwards causation, with our experience of conscious will being both a product and an apparent cause of behavior. (Bear, 2016: para. 3).

Bear and Bloom sought a way to resolve this seemingly paradoxical role of conscious intentions. They aim to explicate the machinations of the "illusion" of conscious mental causation which Wegner identified. Bear (2016) summarizes, "Perhaps in the very moments that we experience a choice, our minds are rewriting history, fooling us into thinking that this choice—that was actually completed after its consequences were subconsciously perceived—was a choice that we had made all along" (para 4).

The response to the Griffin article shows how such research, from Libet, to Wegner, to Bear and Bloom, gets appropriated into popular media and public perception. Popular media rarely concerns itself with the fine-grained distinctions maintained in and between formal disciplines, and the result is that Griffin's popular media article ("Free Will Could All Be an Illusion") is only a semi-faithful reproduction of Bear's article, which is itself already a "blogosphere-appropriate" distillation of the study he performed with Bloom. The result is that, despite their crucial differences, Wegner, Bear, and Bloom's psychological conclusions end up resembling the neuroscientific claims of Greene and Churchland when represented to the public. While Bear says "our *minds* are rewriting history" (italics added), Griffin writes, "the *brain* rewrites history when it makes its choices" (Italics added; Griffin, 2016). So, in effect, although the research differs in significant ways, when viewed through the generalized lens of popular media, such differences often fall away from view. Whether a lay-reader should take up the article by Greene or the article by Griffin, he/she will likely read it as an argument that your thought and behavior is determined by events which occurred in your brain. The radicalization of claims is just one of the many transformations "brain facts" encounter as they move through various spheres and stakeholders. These are "not objectively given things-in-themselves but emerge from communities" (Choudhury et al., 2009:65). This shows the extent to which neuroscientific practices, artifacts, and concepts move beyond the lab and become an increasingly multi-faceted "cultural activity" (Choudhury et al., 2009:63).

4.2.2 Wegner-Inspired Challenges to the Traditional Notion of Legal Responsibility

The comments section on Griffin's article shows how interpreting "brain facts" become a "cultural activity," and how a narrow issue about "the experience of agency" becomes a broad issue about the possible end of free will. Interestingly, many of the comments relate to the legal implications of the research, even though legal applications are not even addressed in the article. This goes to show the extent to which legal applications are particularly capturing to public imagination. Many comments were skeptical, such as: "We aren't mindless zombies incapable of decision. The whole we shouldn't arrest criminals thing is to make a point. We must have free will or there is no point to much of anything. None of our actions could be punished from cheating on your spouse to robbing a bank or even murder. 'Why did you do it?' 'I have no free will I couldn't stop from doing it, I must have been destined to do it'. The no free will concept is idiotic" (Anon qtd. in Griffin, 2016). This already shows that what began as a specific experiment on the experience of agency became a broad argument on the existence or possibility of free will. Many commentators also reflected skepticism, positing ad absurdum how no one would be responsible for his/her behavior anymore (i.e. murder, theft, infidelity, etc). However, other commentators fell on the opposite side of the spectrum, already accepting full physical/mechanical determinacy over human behavior. One commentator added nonchalantly, "Old news. Scientists have known for years about the total lack of evidence with regards to free will" (Anon qtd. In Griffin, 2016). Scientists often assume that the public is highly traditional and reticent to new conceptual frameworks[9] (a number of empirical tests have attempted to test these intuitions, such as Weisberg et al., 2008; Demertzi et al., 2009; Hook & Farah, 2013;), but these comments section indicate to me that the public is about as divided as the experts in the field. This demonstrate how common it is to move from the statement: "conscious will is illusory" that "free will is illusory." Furthermore, Griffin changed the implicaton of Wegner and Bear and Bloom's conclusions when he changed the word "mind" to "brain." This change in language allows one to go from the claim that actions are caused by unconscious mental states to saying they are caused by determined brain states.

Although Wegner's conclusions are often understood as a negation of free will, and therefore a challenge to the legal notion of responsibility, they are considerably more moderate than their reproductions. Wegner's argumentation displays a key difference between his approach and that of Greene and Churchland's, evident in the continued use of the word "mind" throughout his works.

9 Several empirical tests have attempted to test these intuitions, assessing to what extent lay-people maintain traditional notions about the mind-brain-body-world relation, such as Weisberg et al., 2008; Demertzi et al., 2009; Hook & Farah, 2013

Although Wegner asserts that our causal attributions are often mistaken, he does not dispense with the notion of mind entirely. He prefaces his book, *The Illusion of Conscious Will*, by writing, "Yes, we feel that we consciously cause what we do; and yes, our actions happen to us. Rather than opposites, conscious will and psychological determinism can be friends. Such friendship comes from realizing that the feeling of conscious will is created by the mind and brain just as human actions themselves are created by the mind and brain" (Wegner, 2002:ix). In all his empirical work on how the brain relates to "the sense of agency," and the "experience of conscious will," he still has not come to dispense with these concepts as "metaphysical fictions." Wegner, ultimately, does not attempt to challenge the fundaments of traditional legal responsibility. He only somewhat troubles the traditional linkage between causally efficacious conscious states and free will, and does not even jettison these concepts entirely. Wegner differs from Greene and Churchland in the sense that he still recognizes "the mind" as a causal element, and "agency" and "free will" as legitimate phenomenon, therefore mounting little to no challenge to the traditional legal notion of responsibility.

4.3 Churchland's Neuro-reductionism

Patricia Churchland is a renowned philosopher of mind and coiner of the term "neurophilosophy." She is dedicated to exploring how neuroscientific studies can change, challenge, support, and otherwise illuminate philosophical questions. She opens her book *Touching a Nerve:The Self as Brain* with reflections inspired by experiments like Wegner's and Libets when she writes, "Unconscious processes have been shown to play a major role in how we make decisions and solve problems... So you may wonder: How can I have control over a domain of brain activity I am not even aware of? Do I have control over brain activity I *am* aware of? And who is *I* here if the self is just one of the things my brain builds, with a lot of help, as it turns out, from the brain's unconscious activities?" (Churchland, 2013:12). Already she makes the linkage between Libet, Wegner, and Bear and Bloom experimental results and the issue of free will and agency. She makes an interpretive leap here that Wegner and his colleagues do not – she goes from mind to brain. She understands the brain as the primary causal factor producing unconscious processes. She frames it as such:

> The brain is a causal machine... By calling it a causal machine, I mean that it goes from state to state as a function of antecedent conditions. If the antecedent conditions had been different, the result would have been different; if the antecedent conditions remained the same, the same result would obtain. Choices and evaluation of options are processes that occur in the physical brain, and they result in behavioral decisions. These processes, just like other processes in the

> brain, are very probably the causal result of a large array of antecedent conditions. Some of the antecedent conditions result from the effects of external stimuli; others arise from internally generated changes, such as changes in hormone levels, glucose levels, body temperature, and so forth. (Churchland, 2004:5)

These are processes which seem to function without our consciousness, control, or intentionality, and in the same way our heart pumps our blood, our brain regulates our unconscious processes. She ascribes all the typical mentalia – thinking, feeling, choosing, remembering, planning – to the brain (Churchland, 2004:5).

Churchland also refers to empirical experimentation to support her argument. For example, she refers to a number of lesion studies as evidence of the immediate causal linkage between the brain and thoughts and behaviors. Lesion studies involve patients who have damaged some part of their brain, and by observing the changes in an individual after the injury, researchers can reverse-engineer the role that part of the brain could play. This practice may have begun with the curious and ignominious case of Phineas Gage in the mid-nineteenth century. He was a railroad worker who impaled his head on a flyaway iron bar, causing extensive damage to the left side of his prefrontal cortex. Remarkably, despite the immense damage to his brain, he continued to have all of his basic functions. However, the community rapidly noticed a marked change in his personality, and while before the injury he was affable, afterwards he became violent and subject to outbursts. As Mobbs summarizes, "Phineas Gage is compelling to both neuroscientists and legal thinkers because it provided the first indication that reasoning and regard for others can be compromised by frontal lobe injury. Harlow's observations have led many experts to speculate that neurological insult may be a prominent factor in recidivistic and violent criminal transgressions" (Mobbs et al., 2007:693). This strange case seemed to demonstrate that cognitive capacities are realized by the brain because damage to the brain led to respective dysfunctions in cognitive capacities.

Churchland cites of number of cases which show that even something which may seem as inherent as an individual's temperament can be radically altered by changes at the physiological level of the brain. Churchland reviews, "Neurologists reported very specific losses of function correlated with damage to particular brain areas. A person who suffers a stroke in a very specific place in the cortex (the fusiform) will likely lose the capacity to recognize a familiar face; a stroke in a somewhat different area will cause the loss of the ability to understand speech. Loss of social inhibition may follow a stroke that destroys the prefrontal cortex just behind the forehead" (Churchland, 2013:49-50). She

refers to a panoply of lesion studies which have yielded odd and provocative observations. These studies have, for example, even demonstrated that the recognition of one's own mother, a seemingly fundamental cognitive capacity, can be compromised by damaged to the fusiform. She summarizes, "All these phenomena seem to point to the nervous system, not to nonphysical, spooky stuff" (Churchland, 2013:50). She also uses more commonplace examples. For example, she demonstrates a more anecdotal applications of the notion that decisions are made before we are conscious of them when she writes, "I find joy in commonplace mental events, such as a many-factored decision that I have mulled over for days coming to consciousness one morning as I stand in a hot shower. My brain has settled into a choice, and I know what to do. Yay brain!" (Churchland, 2013:21). Ultimately, she understands our entire mental lives as reductive to brain processes – by the time thoughts, memories, beliefs, emotions surface to consciousness, they are already fully-formed through underlying mechanical, physical brain processes. She finds that these cases problematize any notion of a causal factor which is independent of the brain because such a notion presupposes that there is a self which is separate from the physical body, whereas these cases show that there is a tight contingency between the physical body and an individual's thought and behavior.

Churchland positions her argument in relation to a number of foils, and one of the main targets is "folk psychology." "Folk psychology," like all "folk" disciplines, is basically an assemblage of our intuitions and experiences, mostly untested and unquestioned, regarding psychology. Churchland (1988) clarifies, "The starting point for theorizing was of course folk psychology, just as the starting point for modern physics was folk physics ... Thus, according to our folk psychological conception, we have a memory, we are conscious, some memories fade with time, rehearsing helps us remember, one recollection sometimes triggers related recollections, and so forth" (149). "Folk psychology" includes those things about which the reader may have had intuitive notions: such as the notion that thinking, feeling, believing, and intending are mental states; or the notion that we have free will and causally efficacious mental states. These are the intuitive ideas about what constitutes our mental lives which inform our actions every day. Churchland argues that these notions are rarely rigorous enough for dedicated philosophical inquiry, and would be more unified and robust if reinterpreted or reframed in terms of neurological findings. She arrives at a position called "eliminative materialism," the stipulations of which she summarizes as such:

1. that folk psychology is a theory

2. that it is a theory whose inadequacies entail that it must eventually be substantially revised or

replaced outright (hence “eliminative”); and

3. that what will ultimately replace folk psychology will be the conceptual framework of a matured neuroscience (hence “materialism”).

(Churchland, 1988:396).

She basically argues that our common-sense intuitions are likely to be misleading as they have hardly been developed in scientifically or philosophically robust ways, and that a prime candidate for replacing these obsolete notions is neuroscientific understandings of thought and behavior.

Eliminative materialism also offers to resolve another, related issue which poses a recurring conceptual challenge to conceptions of a non-reducible mind. As Kim (2007) identifies it, “The problem of mental causation is to answer this question: How can the mind exert its causal powers in a world that is fundamentally physical? The problem of consciousness is to answer the following question: How can there be such a thing as consciousness in a physical world, a world consisting ultimately of nothing but bits of matter distributed over space time behaving in accordance with physical law?” (Kim, 2007:7). In a world where everything else can be understood in material terms, Churchland contends that understanding mental phenomenon as physical phenomenon makes for a more unified theory. She summarizes the very goal of the book *Neurophilosophy* as *“* to paint in broad strokes the outlines of a very general framework suited to the development of a unified theory of the mind-brain” (Churchland, 1988:3), and furthermore argues that a unified framework must also be a reductive theory (Churchland, 1988:59).

4.3.1 Appropriation of Churchland's Ideas

Churchland's rejection of “folk psychology” and endorsement of a neuro-reductionist stance has been a subject of some controversy among academics and lay-people. The accuracy of these appropriations are, again, somewhat suspect. For example, Churchland (2013) writes skeptically, “We are regaled: 'free choice is an illusion,' 'the self is an illusion,' 'love is just a chemical reaction.' ...In my judgment, such startling claims are more sensational than they are good science. They may contain a kernel of genuine evidence, but they stretch out of shape what is actually established—so much so that the kernel of truth is swamped by hype” (19). However, several of these claims have been ascribed to her. For example, in Rose's article he claims, “In each case, free will would seem to be nothing other than a 'user illusion' (Nørretranders, 1998)—an epiphenomenon to be dismissed summarily, as

Churchland does, as 'folk psychology' (Churchland, 1995)" (Rose, 2005:1001). So although she says that the claim "free will is an illusion" is over-hyped, she is also understood to be "summarily dismissive" of the notion of free will. Similarly, De Vos (2014) claims, "With little or no fuss she observes that one's love for one's child is simply a matter of neural chemistry" (5). It is difficult to tell who is inconsistent, Churchland or her readers, and that is not what I am here to evaluate. What it is interesting is how her ideas, her "brain facts," are interpreted and disseminated.

Churchland did an interview in which she had the opportunity to clarify some of the interpretations of her work. It also shows the interplay between the original author and subsequent deployments of her works which are more or less faithful. In her interview, she began, "I feel very comfortable with my brain and with knowing that my perceptions, my consciousness, my beliefs, my desires, they really are a function of the physical brain that resides within my head" (Qtd. Churchland; Tsakiris, 2014). The interviewer took this conclusion to mean "this idea that consciousness is an illusion of a biological robot" (Tsakiris, 2014). He goes straight to the notion that consciousness is epiphenomenal and that all thought and behavior must therefore be determined by brain processes. Rose also reaches the conclusion that Churchland is pushing for a kind of epiphenomenalism towards the mind. Interestingly, Churchland responds to Tsakiris, "No, it's not an epiphenomenon. It is an actual phenomenon in the physical brain. It's one of the things that the physical brain does in just the way that your brain stores memories. Some of those memories change over time as a result of changes in the physical brain... Memory is a real function of the physical brain and so is consciousness. It's not an illusion; it's the real deal" (Qtd. Churchland:Tsakiris, 2014). This indicates that Churchland's argumentation is perhaps not as aggressive as it is sometimes interpreted to be. She reduces consciousness to the brain, but she does not say that it is something that is passively produced or simply emerges through the complexity of brain processes. Rather, she claims that it is a dedicated function of the brain, but rather than being understood in its typical folk-psychological terms, it should be understood in terms of its brain processes.

4.3.2 Churchland-Inspired Challenges to Legal Responsibility

Churchland pays particular attention to how her insights impact law, and moral reasoning more generally. Churchland (2004) finds that, "In this century, neuroscientific advances in understanding higher functions inspire renewed reflections on the fundamentals of responsibility and punishment. At the most basic level reside questions about the relation between free choice, punishment, and

responsibility" (5). She argues that legal notions of free will, punishment, and responsibility does not need to be jettisoned entirely, but rather than they need to be re-framed in neuroscientific understandings. She posits that the current notion of responsibility seems to rely on an understanding of free will which is "contra-causal free will," or the assumption that "a free choice has causal effects, but no causal antecedents" (Churchland, 2004:6). This notion is entirely inconsistent with her position that thought and behavior, such as making choices, is caused by brain processes. She therefore adopts an approach from David Hume, in which he argued that free choices and un-free choices were both caused by antecedent causes, but that the type of antecedent causes differ. The real question for her becomes, "What are the differences between the causes of voluntary behavior and involuntary behavior? What are the differences in causal profile between decisions for which someone is held responsible, and decisions for which someone is excused or granted diminished responsibility?" (Churchland, 2004:6). She does not presume whether voluntary behavior is determined or not, but rather, that it is caused differently than involuntary behavior.

Churchland sees neuroscientific explanations as a way to respond to the tricky question of when someone can be understood as acting voluntarily, and therefore eligible to be held responsible. Rather than someone being held responsible because he/she was understood to be acting freely, he/she would be held responsible because he/she had the neural correlates of being "in-control." She posits:

> developments in neuroscience in the last 50 years have made it possible to begin to probe the neurobiological basis for decision-making and impulse control. Emerging understanding of the role of prefrontal structures in planning, evaluation, and choice, and of the relationship between limbic structures and prefrontal cortex, suggests that eventually we will be able understand, at least in general terms, the neurobiological profile of a brain that is in control, and how it differs from a brain that is not in control... More correctly, we may be able to understand the neurobiological profile for all the degrees and shades of dysfunctional control. (Churchland, 2004:10-11).

She argues that traditional stances on law have been "exposed as untenable," and neuroscientific understandings will finally enable the law to make accurate and consistent distinctions between those who are held responsible for their acts and those who are not (Churchland, 2004:3). So ultimately, Churchland does not jettison the notions of consciousness, free will, or responsibility; rather, she argus that they need to be re-framed in terms of the brain rather than the mind. When they are re-framed, some notions will remain, in their "materialistic" form, and some will be "eliminated," hence

"eliminative materialism."

Her approaches undermines the traditional legal approach to responsibility on a number of levels. Like Wegner, she casts doubt on whether our conscious mental states (i.e. intentions and decisions) have causal efficacy. However, Wegner only goes so far as to say that folk-psychology is sometimes mistaken, not that it is entirely inconsistent and should be re-framed in terms of the brain. By positing that the brain is the "causal machine" behind the mind, she proposes a different relationship between mind, brain, body, and world than that assumed by law. Rather than mind and brain in a symmetrical causal relationship with one another, the brain is the only causal factor for which she needs to account in her framework. Interestingly, she seems to presuppose that "folk-psychology" underlies legal deliberations and is inherently incompatibilist, positing that a freely-willed choice must have no antecedent causes. I posit that law takes a more comptabilist approach, and that it can take into account both determined processes in the brain and undetermined processes in the mind as causal factors. There is a recurring idea that "folk psychology" is inherently incompatibilist, reified in Churchland's work, but challenged in works like, for example, Nahmias (2006). In his study, Nahmias surveyed a number of lay-people using scenarious and questionaires, and he concluded that "Using three different scenarios with hundreds of participants, we consistently found that the majority (2/3 to 4/5) responded that agents in deterministic universes *do* act of their own free will and *are* morally responsible. That is, we found that most ordinary folk do *not* seem to find incompatibilism intuitive or obviously correct" (Nahmias, 2006:215-216). It might be that Churchland's approach is not as at-odds with traditional assumptions as some might believe. The law seems quite capable of taking into account that some brains are more "in-control" than others, as the brain-based defenses of Weinstein and Hinckley showed. Whether it is capable of showing the fine-grained distinctions between these polarities is a matter up for debate. Also, whether neuroscience is robust enough to take *all* the concepts which formerly invoked "the mind" is also a matter up for debate. While authors like Greene (2004), Sapolsky (2004), and Lamparello (2012) surely agree, authors like Noë (2009), Gazzaniga (2011), and Pardo and Patterson (2013) seem to disagree. General critiques will be discussed further in 4.6.

4.4 Greene's Neuro-reductionism

Joshua Greene is a psychologist, neuroscientist, and philosopher at Harvard University and a prolific author on the utility of neuroimaging in the legal domain. Whether as an ally or a foil, he is one of the most numerously cited authors in this field (i.e. Rosen, 2007; Gazzaniga, 2011; Morse, 2011; Schleim, 2012). He and his co-authors have mounted significant argumentation in favor of the

introduction of neuroscience via neuroimages into the legal domain. One article begins, "Thus, for centuries, many legal issues have turned on the question: 'what was he thinking?'...To answer this question, the law has often turned to science. Today, the newest kid on this particular scientific block is cognitive neuroscience, the study of the mind through the brain, which has gained prominence in part as a result of the advent of functional neuroimaging" (Greene & Cohen, 2004:1775). Greene acknowledges the unique role that neuroimaging technology has played in disseminating neuroscientific explanations. The argument Greene makes is not that the turn to neuroimaging is merely a new iteration of an old practice. While it may have begun that way, these authors propose that the introduction of neuroscience via neuroimages in the legal domain has the potential to shake its very fundaments. They claim:

> In our view, neuroscience will challenge and ultimately reshape our intuitive sense(s) of justice. New neuroscience will affect the way we view the law, not by furnishing us with new ideas or arguments about the nature of human action, but by breathing new life into old ones. Cognitive neuroscience, by identifying the specific mechanisms responsible for behaviour, will vividly illustrate what until now could only be appreciated through esoteric theorizing: that there is something fishy about our ordinary conceptions of human action and responsibility, and that, as a result, the legal principles we have devised to reflect these conceptions may be flawed (Greene & Cohen, 2004:1775).

By "ordinary conceptions of human action and responsibility," they refer to the traditional notion that actions are the results of intentionality, agency, and free will, and only when acting as such is an individual eligible to be held responsible. As we have seen, changes in foundational concepts like agency, free will, and the causal efficacy of conscious, mental states also lead to changes in the legal notions of responsibility because they are so closely interrelated.

Greene's argumentation shares much in common with Churchland's. They refer to similar experiments and insights and inhabit the same disciplinary domain. Greene (2011) supposes that their commitment to neuro-reductionism is endemic to the field when he writes, "The modern science of mind proceeds on the assumption that the mind simply is what the brain does" (263). He goes even further than saying that the mind is the brain, and claims that all thought and behavior can be understood in terms of the brain. In an interview, Greene stated, "To a neuroscientist, you are your brain; nothing causes your behavior other than the operations of your brain" (J.D. Greene; qtd. In Rosen, 2007:3). Greene proposes that even when individuals are perceived as acting agents, or

experience themselves as such, that does not mean that they are in the traditional sense of the term. Their intentional, mental states are not causing their behavior, their brain is, and the conscious and intentional states to which we typically ascribe causal efficacy are mere epiphenomenon. Greene and his colleagues reworked the Libet experiment into a thought experiment, and describe thought processes as such:

> Imagine, for example, watching a film of your brain choosing between soup and salad. The analysis software highlights the neurons pushing for soup in red and the neurons pushing for salad in blue.You zoom in and slow down the film, allowing yourself to trace the cause-and-effect relationships between individual neurons—the mind's clockwork revealed in arbitrary detail.You find the tipping-point moment at which the blue neurons in your prefrontal cortex out-fire the red neurons, seizing control of your pre-motor cortex and causing you to say, "I will have the salad, please." (Greene & Cohen, 2006:218).

These experiments are similar because they both provoke the question: did the subject's brain make the decision before the subject was even conscious of it? If so, does that mean that he/she is not the agent freely willing and intending his/her own behaviors? If so, does that mean that the traditional notion of legal responsibility must be changed or abandoned? Greene's responses to these questions are similar to Churchland's, so I will not devote too much time reiterating arguments which have already been mentioned. Ultimately, both authors assume the standpoint that "folk psychology" is largely mistaken and thought and behavior is largely determined by the brain.

4.4.1 Appropriations of Greene

As I have mentioned before, appropriations tend to be more radical than the original work; however, in Greene's case, his argumentation can only be radicalized so much. Greene's claims are themselves so radical that they do not need to be hyperbolized as much in order to sound sensational. In Rosen's interview with Greene, he summarizes Greene's salad-soup case to mean, "In other words, even someone who has the illusion of making a free and rational choice between soup and salad may be deluding himself, since the choice of salad over soup is ultimately predestined by forces hard-wired in his brain" (Rosen, 2007:4). This largely coheres with the interpretations Greene makes himself, although he would likely avoid the language of "predestined" and opt for a language of "antecedent causes." Rosen (2007) summarized after his interview, "Greene insists that this insight means that the criminal-justice system should abandon the idea of retribution — the idea that bad people should be

punished because they have freely chosen to act immorally — which has been the focus of American criminal law since the 1970s... Instead... the law should focus on deterring future harms" (4). In Greene and Cohen's own writings, they are a bit more moderate than how Rosen characterizes their claims. They actually argue that legal *practice* can easily assimilate neuroscientific evidence without significant challenges, but that it is at the level of legal *theory* that the challenge occurs (Greene & Cohen, 2004:1775). However, they do argue that neuroscience has the potential to fundamentally change how we think about punishment and responsibility because it negates notions like free will and the causal efficacy of conscious mental states.

It is often assumed that since Greene challenges the causal efficacy of mental states, agency, and free will, he *must* dispense with the legal notion of responsibility entirely. For example, Aharaoni et al. (2008) ascribe to him the claim that, "Moreover, most agents do not know what is going on in their brains, so they cannot choose certain neural events rather than others with any specificity. In that way, the neural causes of an action are beyond the agent's control. Such considerations, among others, lead some philosophers to deny that agents are responsible for anything they do" (146). This charactrization of Greene's position is only partly faithful. Although Greene does subscribe to a deterministic understanding of thought and behavior, he does not conclude that responsibility ceases to exist. He does deny that *agents* are not responsible for anything they do because he denies the "standard conception" of agency, however, he does still stipulate that some notion of responsibility needs to be preserved. Greene and Cohen conclude, "Free will, as we ordinarily understand it, is an illusion. However, it does not follow from the fact that free will is an illusion that there is no legitimate place for responsibility" (Greene & Cohen, 2004: 1783). This demonstrates that Greene is sometimes read to be more radical than he positions himself, although it seems the only step he does not take is an all-out challenge against the very notion of responsibility.

4.4.2 Greene-Inspired Challenges to Legal Responsibility

This sections demonstrates that various proponents of neuro-reductionism draw different implications for law. Whether neuro-reductionism is an accurate characterization of Wegner or not, he coheres with Greene in that they both think our "folk psychological" ascriptions of causal efficacy to conscious mental states are often mistaken. However, they also differ significantly. Greene himself clarifies, "Daniel Wegner argues that free will, while illusory, is a necessary fiction for the maintenance of our social structure (Wegner 2002, ch. 9). We disagree. There are perfectly good, forward-looking

justifications for punishing criminals that do not depend on metaphysical fiction" (Greene & Cohen, 2004:1783). Wegner does not deny the possibility of causally efficacious mental states, and therefore he does not necessarily deny free will. However, Greene and Churchland both take this a step further, arguing that some notion of responsibility can be preserved even when the notion of free will is significantly altered (in Churchland's case) or dispensed with entirely (in Greene's case). Although they cohere in offering neuro-reductive explanations of thought and behavior, they also share key differences in how they understand the implications for law. Pardo and Patterson (2013) compare, "For Churchland, recognizing 'the mind is the brain' is the basis for delineating a normative distinction between legal responsibility and excusable behavior. For Greene and Cohen, by contrast, recognizing the 'mind is the brain' is the basis for eliminating any coherent notion of legal responsibility" (41). As aforementioned, this claim is slightly more radical than Greene's actual claim. However, he does go a step further than Churchland when he claims succinctly: "when it comes to the issue of free will itself, hard determinism is mostly correct" (1783). Unlike Churchland, Greene does not attempt to re-frame free will in neurological terms, he finds the notion to be a largely obsolete "metaphysical fiction." While Churchland resists her characterization as a determinist or epiphenomenalist, Greene does not share this resistance. Greene and Cohen presume that, "The net effect of this influx of scientific information will be a rejection of free will as it is ordinarily conceived, with important ramifications for the law" (1776). These authors all agree that neuroscientific information certainly *changes* how you conceive of free will, but whether it entails complete rejection seems to vary.

Greene's argumentation is novel in that he mounts a challenge to the whole legal framework of "retributivism," or the prevailing Western idea that criminals "deserve" to be punished in proportion equal to their transgressions. Pardo and Patterson (2011) summarize, "Joshua Greene and Jonathan Cohen challenge retributivism by arguing that neuroscientific data will undermine retributivist intuitions *indirectly* by undermining *directly* the 'free will' intuitions on which, they claim, retributivist theories depend" (3). Although I have previously made the claim that the law is generally compatibilist, both Greene and Churchland resist this claim. Greene and Cohen (2004) begin their article, "We argue that current legal doctrine, although officially compatibilist, is ultimately grounded in intuitions that are incompatibilist... In other words, the law says that it presupposes nothing more than a metaphysically modest notion of free will that is perfectly compatible with determinism. However, we argue that the law's intuitive support is ultimately grounded in a metaphysically overambitious, libertarian notion of free will that is threatened by determinism and, more pointedly, by forthcoming cognitive neuroscience" (1775). So although a negation of free will is not *necessarily* in-conflict with law, if one

conceives of law as incompatibilist, they are certainly more likely to conflict.

While Churchland argues that most of the notions underlying legal responsibility can be preserved but re-framed, Greene argues that all the implicated concepts (agency, mental states, free will) can be jettisoned, leaving only a bare-bones notion of responsibility. Ultimately, Greene recognizes that there is too much social utility in maintaining the concept of responsibility to dispense with it entirely. However, he argues that neuroscientific explanations will change the entire framework within which this concept operates. Greene and Cohen write:

> ...advances in neuroscience are likely to change the way people think about human action and criminal responsibility by vividly illustrating lessons that some people appreciated long ago. Free will as we ordinarily understand it is an illusion generated by our cognitive architecture. Retributivist notions of criminal responsibility ultimately depend on this illusion, and, if we are lucky, they will give way to consequentialist ones, thus radically transforming our approach to criminal justice. At this time, the law deals firmly but mercifully with individuals whose behaviour is obviously the product of forces that are ultimately beyond their control. Some day, the law may treat all convicted criminals this way. That is, humanely. (Greene & Cohen, 2004:1784).

Their conclusion is that human society needs some notion of "responsibility," but casts skepticism each of the presuppositions which underly its traditional legal deployment. Within this framework, legal responsibility would no longer be about "deserving" punishment, it would be about "warranting" punishment for other (presumably consequentialist) reasons. While they do position themselves as anti-retributivists, it is unclear whether they are proponents of a full *Minority Report*-esque scenario in which criminal justice only works on the basis of predicting crimes. However, although they do not directly make these statements , that is the implication that some commentators have read from them (i.e. Rosen, 2007; Snead, 2007).

4.5 Summary

Whether it accurately reflects their statements, neuro-reductionists are often characterized as making much strong claims against the traditional legal notion of responsibility, even the whole theory of retributivism. In this sense, the position *becomes* real, even if it was not directly articulated in the works of these authors. In another sense, these authors effectively paved an inroad to the more radical claims. The "brain fact" that humans do not have free will has been appropriated, and now the interpretations and ramifications of this perception are multiplying. The conception that neuro-

reductionists mean to challenge the fundaments of Western law has become reified throughout the literature in various ways. Increasingly, it might be an accurate characterization. Other neuro-reductionists who were not reviewed here have also increased the perceived tensions between neuroscientific explanations and legal concepts.

For example, Robert Sapolsky is sometimes compared with Greene (Rosen, 2007; Snead, 2007). Rosen (2007) included his interviews with Sapolsky and Greene in the same article, comparing Greene's brain-based explanation of behavior to Sapolsky's statement: "'You can have a horrendously damaged brain where someone... can't control their behavior,' says Robert Sapolsky, a neurobiologist at Stanford. 'At that point, you're dealing with a broken machine, and concepts like punishment and evil and sin become utterly irrelevant. Does that mean the person should be dumped back on the street? Absolutely not. You have a car with the brakes not working, and it shouldn't be allowed to be near anyone it can hurt'" (R. Sapolsky qtd. In Rosen, 2007:4). In likening a human to a machine, Sapolsky seems to imply a physicalistic, deterministic approach to to human thought and behavior. He defends,Legal scholars have objected to this type of thinking for a related reason, as well. In this view, it is desirable for a criminal justice system to operate with a presumption of responsibility because, 'to treat persons otherwise is to treat them as less than human' (Morse 1976). There is a certain appealing purity to this. But although it may seem dehumanizing to medicalize people into being broken cars, it can still be vastly more humane than moralizing them into being sinners" (Sapolsky, 2004:1794).He bolsters the notion that criminality is a determined neurological state, with neurological signatures, for which law should enlist neurosciences help to identify. He correlates the pre-frontal cortex, which is the same area of the brain which was invoked in the Hinckley and Weinstein cases, with the capacity to control one's behavior and delay one's pleasure, which are the key capacities for cohering to social and legal norms. He, like Churchland, argues that there are neural signatures of "in-control" brains and "out of control" brains. He writes, "What the literature about the PFC shows is that there is a reductive, materialistic neurobiology to the containment, resulting in the potential for volitional control to be impaired just as unambiguously as any other aspect of brain function. It is possible to know the difference between right and wrong but, for reasons of organic impairment, to not be able to do the right thing" (Sapolsky, 2004:1793-1794). Sapolsky's comment coheres with some of the deterministic stances on the aforementioned court cases, for example, the comment: "He played the insanity card. Once insane--always insane. Keep the goon in the asylum" (Qtd. In Phelps, 2015). Neurological determinism supported by neuro-reductionism offers a way to substantiate such statements.

Lamparello (2012) endorses a similar position, arguing that inmates should be kept incarcerated

past their sentence if their neurology is consistent with criminal behavior. He argues, "cognitive neuroscience provides an objective basis upon which to predict future dangerousness and provide for the involuntary commitment of violent offenders both during and after their sentence... In assessing future violence, there are two areas that bear directly upon cognitive function, emotion, and behavioral control— the *frontal lobe* (within which lies the pre-frontal cortex) and the *limbic system* (which includes the amygdala)" (Lamparello, 2012:270). He, like Sapolsky, views some criminal behaviors as determined by neurological activity. He argues that studies on damage to the pre-frontal lobe has ultimately demonstrated that, "Ultimately, therefore, individuals with damage to their frontal lobes have shown, *inter alia*, an incapacity 'to develop normal social and emotional responses while retaining other intellectual capacities'" (Lamparello, 2012:270-271). Damage to the amygdala, he argues, also impairs social and emotional responses. He finds that neuroscientific methods, such as neuroimaging, are the most accurate means we have to assess those factors. He concludes that, "Ultimately, because neuroscience can determine—with a reasonable degree of accuracy—whether a criminal defendant remains a threat to himself or others due to an identifiable mental illness, the threshold standard for involuntary commitment of such individuals can be satisfied. As a result, neuroscience can provide a constitutional basis upon which to involuntarily confine criminal defendants either during or *after* their sentences have been completed" (Lamparello, 2012:272). He cites cases in which this practice was actually observed, and individuals convicted of pedophilia were convincingly shown to pose an increased liability for repeat offense, and were therefore incarcerated past their sentence. By positing that the brain determines thought and behavior, and therefore also determines criminality, he challenges law. Is someone responsible for their neurologically determined behavior? Snead (2007), who is characterizes all cognitive scientists as neuro-reductionists, concludes: "in the long term, cognitive neuroscientists aim to draw upon the tools of their discipline to embarrass, discredit, and ultimately overthrow retribution..." (1269). Positing that individual thought and behavior does seem to reify the question: how does one proceed with criminal behaviors? Punishment, rehabilitation, indefinite incarceration?

One response to this question that several authors have noted the connection between neuro-reductive approaches and the movement towards predictive law. Harrop (2013) writes a dissertation dedicated to the claim that, "I will argue that a paradigm shift in the focus of the CJS [Criminal Justice System] from backwards-looking reactionary punishment to forwards-looking crime prevention constitutes not only the probable future of the CJS, but, increasingly, defines its reality too" (8-9). She also dedicates a chapter to exploring the role of neuroimaging technologies (specifically fMRI) play in

the shift towards predictive law (Harrop; 2013:30). While the idea that we would stop punishing for crimes and instead start predicting future criminality (presumably based on neurology, or at least biology) might seem like the stuff of science fiction, this claim has been taken very seriously. Crawford argues that, in general, if one convincingly argues for a determinant causal relationship between a brain state and a criminal behavior (as Greene, Churchland, and co. would at least find conceptually possible), predictive law seems the likely result. He writes, "if human behavior is electrochemically preordained, there remains no discernible ground on which to object to preemptive interventions directed against those identified as 'hard-wired' malfeasants. Such interventions might take the form of surveillance, incarceration, or alteration (through drugs, surgery, or implants)" (Crawford, 2008:76). So whether many neuro-reductionists *actually* argue for the utility of predictive law is less important than the fact that predictive law is almost inconceivable without assuming some kind of reductionist stance, and neuro-reductionism is the "in" reductionism at the moment. Especially when a "brain fact" is appropriated into news and journalism, fine-grained distinctions between "neurological determinism = no responsibility" and "neurological determinism = a revised understanding of responsibility" often fall away. Before a neuroscientist even has the chance to clarify, headlines read boldly "Free Will Could All Be An Illusion" (Griffin, 2016) and "Will Neuroscience Radically Transform the Legal System" (Greely, 2012). So whether the challenging of retributivism issues from neuro-reductionists themselves or from the appropriation of their statements is ultimately irrelevant – the conceptual challenge, for all intensive purposes, has been mounted.

There is also a broad perception that *all* cognitive scientists, neuroscientists, and neurolegalists are neuro-reductionist. Snead (2007), for example, assumes that the introduction of cognitive neuroscience into the court *necessarily* means discrediting the presuppositions of law. Indeed, one prominent neuroscientist estimates that "98 to 99 percent" of neuroscientists subscribe to a neuro-reductive model (M. Gazzaniga qtd. In Pardo & Patterson, 2013:43). Gallagher et al. (2015) proffers that:

> ...cognitive neuroscience aims for reductionist explanations. We explain higher-order cognitive processes ultimately in terms of neuronal processes, and some would argue that the best explanations of cognition are found at the level of molecular neuroscience (Bickle 2003), or even at the quantum level (Penrose 1999). Philosophers, from Dan Dennett (1991) to Patricia Churchland (2011; 2013), argue that explanations in terms of consciousness or conscious intentions (what Dennett [1989] calls the 'intentional stance') are, at best, placeholders for stricter scientific explanations in terms of computational design or neuronal physics. (Gallagher

et al., 2015:9).

These authors claim that within the field of cognitive neuroscience, neuro-reductionism is endemic. This is partly just a symptom of the fact that this discipline focuses, by default, on the brain, partly because "brain-images... reinforce the view of the brain as a causal agent by framing the brain as an isolated realm" (Aydin, 2016:5). Although reductive models for human thought and behavior may have predated neuroimaging technologies, they rarely took the brain as the primary object. The broad appeal of *neuro*-reductive models can, in many ways, be ascribed to neuroimages.

It might seem to the reader that the role of neuroimages has fallen from view in the previous sections. However, it is important to note that these neuro-reductionistic explanations for behavior are made possible by neuroimaging. Although Wegner does not specifically rely on neuroimaging, he does rely on insights from Libet, which did rely on neuroimaging, so he does have some secondary influence. More clearly, though, views like that of Churchland, Greene, or Sapolsky and Lamparello, are heavily mediated by neuroimaging technologies. De Vos (2014) writes in regards to Churchland, "Patricia Churchland, for example... argues that making decisions, going to sleep, getting angry, being fearful... are just functions of the physical brain (Churchland, 2013b).... brain science and its images actually deconstruct... the subject itself: you are not even unified, but, rather, as it were sliced up by the brain image and dispersed in the neural network" (2). De Vos suggests that brain imaging makes it *possible* to conceive of oneself as one's brain. He claims that brain imaging brings with it the icon, the image itself, and the iconography, the neuro-reductive approach. In this sense, the two are interdependent. He writes more generally, "We want to see ourselves, we are fascinated by the made visible brain, that thing that does all that psychological stuff of thinking, wanting and desiring. Perhaps this is why we denounce the idea of rational agency, free will and love altogether. Because when we observe ourselves, via the image of the brain, we take a position outside or beyond cognition, will and desire... the spectacle of the brain engenders the spectator, a paradoxical and emptied out agency outside of itself" (4). By seeing the brain, in action and in causal interplay with our every thought and behavior, it is easy to start identifying those thoughts and behaviors with the brain. If the brain, and not our experienced self, is the agent of our thought and behavior, does the notion of agency even make sense anymore?

Aydin draws our attention to a somewhat paradoxical quality of this reasoning: it frames the brain as the only causal factor for which to account, and yet it seems that the technology itself is a causal factor in the explanation. Aydin (2016) clarifies, "conceiving brain-images as a 'direct representation' of a brain that is unaffected by external influences and, hence, is seen as the locus of our

selves, excludes per definition the conception that our self-identifications are technically mediated" (2). So although these explanatory frameworks rely heavily on the technology, they also tend to render the technology irrelevant in their explanations. As Greene said, "nothing causes your behavior other than the operations of your brain" (J.D. Greene; qtd. In Rosen, 2007:3), but it does seem that the technology has played a causal role in his behavior (i.e. the behavior of ascribing sole causality to the brain). This interesting paradox, and possible frameworks for the causal role of technology, will be explored in greater detail in 5.1.4.

4.6 Critiques of Neuro-Reductionism

Neuro-reductionism has been challenged on a number of levels. Titles like "Our Brains Are Not Us"(Glannon, 2009) and *Out of Our Heads: Why Your Are Not Your Brain* (Noë, 2009) immediately reveal that scholars, especially philosophers of mind, beg to disagree. A popular media article laments, "An intellectual pestilence is upon us. Shop shelves groan with books purporting to explain, through snazzy brain-imaging studies, not only how thoughts and emotions function, but how politics and religion work, and what the correct answers are to age-old philosophical controversies. The dazzling real achievements of brain research are routinely pressed into service for questions they were never designed to answer. This is the plague of neuroscientism – aka neurobabble, neurobollocks, or neurotrash – and it's everywhere" (Poole, 2012). This recurring general critique here is that the advances in neuroscience are being appropriated to the point of nonsense, extending to claims beyond what the laboratory practices can empirically support. Many of the claims made on behalf of emerging neuroscience are, as we have seen, not based in empirical evidence, but in conceptual commitments. Critics have enumerated a number of contentious conceptual commitments which often become "black-boxed," especially when a "brain fact" has been re-appropriated from its original context.

This skepticism is often based in the idea that there is so little which is uncontroversially "known" about the relationship between the mind and brain that the inferential steps it takes to say "the mind *is* the brain" are simply unwarranted. Poole adds, "The human brain, it is said, is the most complex object in the known universe. That a part of it 'lights up' on an fMRI scan does not mean the rest is inactive; nor is it obvious what any such lighting-up indicates; nor is it straightforward to infer general lessons about life from experiments conducted under highly artificial conditions. Nor do we have the faintest clue about the biggest mystery of all – how does a lump of wet grey matter produce the conscious experience you are having right now, reading this paragraph? How come the brain gives rise to the mind? No one knows" (Poole, 2012). Since it is unknown *how* the brain gives rise to the

mind, it is certainly unknown *that* the brain is the mind. Furthermore, since this underlying foundation is, in itself, shaky, further inferences built upon it are also dubitable. For example, if one does not know how the brain relates to the mind, how can he/she make more specific inferences, like that increased blood-oxygenation in a region of the brain revealed on an fMRI scan means an individual has a certain mental state?

4.6.1 Challenges to the Technological Practices of Neuro-Reductionism

Often, empirical arguments about the technical limitations of neuroimaging technologies are used to support this skepticism. Feigenson (2006), Roskies (2007), and Klein (2009) all make detailed, dedicated arguments regarding the many layers of inferring and interfacing at work in the production of a brain scan. Collectively, they combat the notion that a neuroimage is akin to a photograph of the brain, neutrally revealing its mechanisms, and note out the many levels at which abstractions and interpretations occur. Klein (2009), for example, notes that, "That fMRI is an indirect measure is in itself unremarkable, and should not engender skepticism. Neuroimages are not simple pictures of BOLD signal differences, however. Quantitative signal magnitudes are effectively uninterpretable on their own, as there is no general mapping from BOLD signal to functional significance of neural activity. Further, the BOLD differences associated with brain activity are small, noisy, and temporally complex. In lieu of quantitative information, neuroimages instead show maps of regions where there was a *statistically significant* difference in BOLD signal between task conditions" (267). This relates back to the three levels of inference enumerated by Feigenson (2006), from fMRI data to BOLD levels, BOLD levels to neuronal activity, and neuronal activity to a cognitive state. Empirical criticisms tend to focus on the contentiousness of assumptions made on each of these levels. Greely adds in a light-hearted news article, "The fMRI results showing apparently purposeful brain activity in dead salmon are a wonderfully funny example of some of the limits of this technology... Half of what neuroscience is teaching us about human brain function will be shown, in the next 20 years, to be wrong—and we will need each of those 20 years to figure out which half" (Greely, 2012). From the number of empirical challenges, they make a more general, conceptual challenge that neuro-reductionists make unsupported and unsupportable inferences in order to make any claims about causation.

Neuroscientists across the board generally acknowledge that the brain functions through employing networks, and not isolated regions each dedicated to a single cognitive function. This conception is sometimes compared to a 18th century discipline called "phrenology," which posited that there are single, dedicated regions for each cognitive task. "Phrenology" sought to relate specific

cognitive tasks with specific regions of the brain, but is now commonly referred to a mysticism or "pseudo-science." With the rejection of phrenology comes the acceptance that the brain is networked or "modular," tasks are not carried out by single, dedicated part but diffused throughout several parts working together. This poses both a technical challenge and a conceptual challenge. The technical challenge is that fMRI does not scan the whole brain at once, rather, it scans small parts of it at a time. Researchers can make composite images to cover a wider area, but more often than not, they must have an idea of where to look before they start scanning.

The technical challenge of knowing where to look means that they encounter the conceptual challenge of presupposing that a cognitive activity engages that specific part of the brain. For this reason, some scholars have levied the criticism that neuroimaging can sometimes resemble a "new phrenology." This is, for example, the subject of William Uttal's book, *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain* (2001). He levies the argument, which several other authors subsequently deployed, that cognitive neuroscience runs into a massive conceptual roadblock when it comes to the issue of whether cognitive states have corresponding brain states. Hubbard (2003) summarizes his three main questions as:

> 1. Can the mind be subdivided into components, modules or parts?
>
> 2. Does the brain operate as an equipotential mass or is it also divisible into interacting but separable functional units?
>
> 3. Can the components, modules, or parts of the mind, if they exist in some valid psychological sense, be assigned to localized portions of the brain? (Hubbard, 2003:23).

He draws our attention to several key issues which beset the neuroimaging enterprise. This is that researchers must first presuppose what cognitive state they are investigating, but defining a cognitive state is a tricky business. For example, referring back to the legal application, is "lying" a distinct cognitive state with a distinct neural correlate? Or is "lying" just a permutation of the more fundamental cognitive process of "deception"? Or is "deception" just a permutation of an even more fundamental cognitive process? Or are "lying" and "deception" distinct processes with distinct neural correlates? If one subscribes to the notion that "lying" and "deception" are distinct processes, researchers who conflate the two (i.e. use an experiment about lying to detect deception) are conceptually invalidated from the outset. In summary, "Uttal claims that there has been, and can be, no progress on the problem of developing a taxonomy of cognitive processes, and therefore there can be no hope of localizing cognitive processes in the brain" (Hubbard, 2003:24). He substantiates this claim

by providing several different taxonomies of cognitive processes, proposed by various psychologists. Other authors have bolstered this point, agreeing that cognitive neuroscientific research "is a research method the validity of which depends on a premise. That premise is that mental processes can be analyzed into separate and distinct faculties, components, or modules, and further that these modules are instantiated, or realized, in localized brain regions" (Crawford, 2008:66). These authors sometimes term this type of research "the localizationalist enterprise," characterizing it by its attempt to "localize" mental processes to parts of the brain. Kutas and Federmeier (1998) caution, "Cognitive acts do not necessarily have a location and, even if they do, the site of a measure is not necessarily the site of the action" (137). This offers both the conceptual challenge of defending why cognitive acts are "localizable," and the empirical challenge of knowing where that locale might be. Referring back to the legal cases, the reader can see how this "localizationalist enterprise" has emerged thus far in law, in attempts to "localize" violence, control (or lack thereof), lying, and deception to parts of the brain appearing in color on a neuroimage (in particular, the pre-frontal cortex and amygdala).

Due to the variation of underlying conceptual notions, stakeholders with different conceptual commitments can view the same experiment, but interpret it differently. For example, let us refer back to the famous Libet example which emerges in many neuro-reductionists' arguments. This experiment is interpreted by Greene and others to mean that an action or intention has already been executed by the brain before an individual is conscious of it, which means that the brain has primary causality and the conscious state is non-causal. This, in effect, means that humans are not conscious, free agents (as their intuitive "folk psychology" would lead them to believe), but rather that they are simply the outcome of mechanisms in the brain. Yet, Aydin (2016) notes that, "Advocates of free will often draw upon the same brain imaging technologies but claim that they do not display the non-existence of free will" (3). The mere presence of continued debate over his experiments demonstrate that the interpretations of it are controversial.

Aydin adds that Libet himself did not dismiss the notion of free will, especially after conducting a second experiment which had the same conditions (you press a button when you decide to wag your finger), but in which they could "veto" their decision at the last moment, and decide not to wag their finger. In this experiment, the "readiness potential" still presented on the EEG about half a second before the individual pushed the button (indicating that he/she was conscious of making the decision), however, if one chose to "veto" his/her decision, the "readiness potential" would stop immediately before the finger wag. From this experiment, "Libet and some contemporary advocates of free will argue that it might be true that we mistakenly believe that our actions are caused by conscious

intentions, but that does not exclude the ability to intervene in an impulse, and consciously veto over and stop the action that our brain has unconsciously prepared" (Aydin, 2016:3).

Other scholars take it further, claiming that we cannot even go so far as to intuit any degree of neurological determinism from this experiment. For example, relating back to the "localizationalist" critique, "Koechlin and Hyafil (2007) have pointed out that the brain regions that often are being studied in the context of free will debates, namely the (pre-)supplementary motor area and the anterior cingulate motor areas of the brain, may be only involved in the later stages of motor planning. It is possible that other parts of the brain, which in experiments of opponents of free will are not taken into account, are responsible for decision-making and exerting will" (Aydin, 2016:4). Indeed, there are empirical experiments to support the notion that other parts of the brain are also implicated in this process. A group of researchers tried a version of the Libet experiment on a patient who was hysterically paralyzed. They found that the area which Libet specified did not activate, even when he claimed he was genuinely intending to move it. Another area of the brain did activate, the and V.S. Ramachandran (2004) suggests, "It's as if this activity... was inhibiting or vetoing the hysterical patient's attempt to move his leg" (85-86). Whether the area which Libet and his contemporaries focuses on is necessarily the only area of interest is, in itself, a contentious conceptual presupposition. Since the readings of these experiments depend on conceptual presuppositions, and these conceptual presuppositions are generally contentious, the readings of these experiments are also contentious. Many scholars are skeptical that such an experiment has any real bearing on the questions of free will or conscious causation.

It is not only philosophers who claim that the strength of claims to causal linkages is generally exaggerated. As Gazzaniga (2006) cautions, "Being able to see an area of the brain light up in response to certain questions... may reveal some fascinating things about how certain cognitive states may work, but it is dangerous and simply wrong to use such data as irrefutable evidence about such cognitive states. What we know about brain function and brain responses is not always interpretable in a single way and therefore should not be used as infallible evidence" (144). These authors all agree that our knowledge about how the brain and mind relate is not robust enough to assume one-to-one relationships between active regions of the brain and mental states. Just because part of the brain "lit up" before an individual wagged his/her finger does not mean that that part of the brain is the "finger wagging" part. The inferential leaps between the data and the conclusion (i.e. "humans do not have free will" or "mental states are just brain states") are highly contentious to philosophers (i.e. Noë, 2009, Bennett & Hacker, 2010; Pardo & Patterson, 2013), neuroscientists (i.e. Gazzaniga, 2006), legalists

(i.e. Feigenson, 2006; Roskies 2007; Goldberg, 2011), and laypeople (i.e. Poole, 2012) alike.

This argument that these inferential leaps are unsubstantiated also casts doubt on the lesion cases to which many neuro-reductionists refer. No one denies that these cases are fascinating and thought-provoking, but whether they substantiate the claim that "mental causation is a myth" is another matter entirely. V.S. Ramachandran is a preeminent researcher in brain lesion studies, and yet he always precedes his claims with a "might" or a "seem" or "tentatively," indicating that he is hypothesizing more than concluding. He notes at the beginning of his book, *A Brief Tour of Human Consciousness*, that "I make no apology for the fact that it is speculative... Speculation is fine, provided it leads to testable predictions and so long as the author makes it clear when he is merely speculating..." (Ramachandran, 2004:x). He frequently encourages that future research needs to be done on these curious cases before strong, causal claims can be made. He also refers, throughout his book, to "consciousness" and "the mind," not as dualistic fallacies, but as necessary terms which cannot simply be reduced to "the brain." Sinnott-Armstrong et al. (2008) discuss a lesion case in which an individual experienced a sudden onset of sexual desires towards children which was associated with a tumor. When the tumor was removed, the desires abated, when it returned, they also returned. However, they claim unequivocally, "When all the evidence we have is functional brain scans, then we do not have enough evidence to infer causation" (365). These authors show that the line between what constitutes a provocative correlation versus what constitutes a conclusive causation is up for negotiation.

4.6.2 Challenges to the Conceptual Frameworks of Neuro-reductionism

One of the more general critiques regarding neuro-reductionism is that it simply does not make sense to attribute cognition and consciousness to one part of the human being. Bennett and Hacker are two of the most outspoken and frequently-cited authors who take issue with the ascription of "psychological attributes" to the brain, claiming that it conceptually does not make sense to claim that the brain performs cognitive tasks. They refer to a quote by Wittgenstein in which he states, "Only of a human being and what resembles (behaves like) a living human being can one say; it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious" (Qtd. In Bennett et al., 2009:19). They find that there is insufficient empirical data to indicate that brains are capable, in themselves, of performing the same cognitive functions as humans. What kind of empirical experiment could possibly indicate whether brains can or cannot perform all the tasks of mentation without presupposing one way or the other? They summarize, "it is our contention that this application of psychological predicates to the brain *makes no sense*... The brain neither sees *nor is it blind* – just as sticks and stones are not awake,

but they are not asleep either... The brain is not a logically appropriate subject for psychological predicates. Only a human being and what *behaves* like one can intelligibly and literally be said to see or be blind..." (Bennett & Hacker, 2009:21). They term this misattribution "the mereological fallacy," which is "the neuroscientists' mistake of ascribing to constituent *parts* of an animal attributes that logically apply to the *whole* animal" (Bennett & Hacker, 2009:22). Other authors cohere with this critique, for example Noë (2009) argues, "Consciousness is not something the brain achieves on its own. Consciousness requires the joint operation of brain, body, and world. Indeed, consciousness is an achievement of the whole animal in its environmental context. I deny, in short, that you are your brain. But I don't deny that you have a brain. And I certainly don't deny that you have a mind. To have a mind, though, requires more than a brain. Brains don't have minds; people (and other animals) do" (10). These authors generally agree that mental capacities are a feature of people, not brains. Morse (2005) applies this critique in the legal context when he writes, "Brains do not commit crimes; people commit crimes. This conclusion should be self-evident, but, infected and inflamed by stunning advances in our understanding of the brain, advocates all too often make moral and legal claims that the new neuroscience does not entail and cannot sustain" (397). He coheres with Bennett and Hacker in agreeing that it does not make sense to ascribe mental states, let alone the combination of mental and behavioral states that constitutes a crime, to brains.

In order to explicate what Bennett and Hacker mean by "mereological fallacy" take, for example, Churchland's description of memory. She writes, "Memories of childhood, social skills, the knowledge of how to ride a bicycle and drive a car— all exist in the way neurons connect to each other" (Churchland, 2013:12). Does it makes sense to say that memories are neurons? Several films and science fiction novels have provoked this question by supposing that someone had artificial memories implanted in his/her brain. The protagonist and audience often spend the majority of the movie wondering "did it really happen?" If one argues that a memory *is* the neurons in the brain, then the artificially implanted memory is as real as any other. We would not be left asking if it "really" happened, because the neural signature would be the only relevant marker. This seems to tease the intuition that a memory is more than just the connections of neurons, but rather an experience in which the whole person was immersed, the recollection of which is *enabled* by the connection of neurons, but not reducible to it. Churchland (2013) ascribes some very higher-level thinking to the brain, writing, "There are some things that brains do very slowly and that involve deep intelligence or deep shifts in worldview. Human brains do figure things out, sometimes only after years and years of pondering and marinating and seeing what makes sense" (24). Here she asserts that brains have a variety of cognitive

attributes and predicates: their own intelligence and worldviews, they figures things out, they ponder, they see, and they make sense of things. To some scholars, this ascription of the whole panoply of mentation to the brain starts to resemble another version of Descartes "ghost in the machine." Bennett and Hacker argue that one explanation for how the ascription of psychological predicates to the brain went without question is: *"*We suspect that the answer is – as a result of an unthinking adherence to a mutant form of Cartesianism. It was a characteristic feature of Cartesian dualism to ascribe psychological predicates to the mind, and only derivatively to the human being... the predicates which dualists ascribe to the immaterial mind, the third generation of neuroscientists applied unreflectively to the brain instead" (Bennett et al., 2009:21). Their critique regarding the "mereological fallacy" coheres with a more common critique that neuro-reducionists employ a permutation of Cartesian dualism.

Cartesian dualism is a conceptual framework dating back to the seminal philosopher, Renée Descartes. In his famous *Meditations,* he performed his famous thought experiment wherein he attempted to cast doubt on everything he believed in order to see that which could not be doubted. He arrived at his famous conclusion: I think, therefore I am. He finds, "I am, I exist, that is certain. But how often? Just when I think; for it might possibly be the case if I ceased entirely to think, that I should likewise cease altogether to exist. I do not now admit anything which is not necessarily true: to speak accurately I am not more than a thing which thinks, that is to say a mind or a soul, or an understanding, or a reason, which are terms whose significance was formerly unknown to me" (Descartes, 1641:10). Descartes intuited that since only the existence of thought could not be doubted, the essence of a human being is a "thinking thing." Therefore, he posited that the mind or soul is the central causal force of thought and behavior. His position is called dualistic because it forms a split between the mind and body, and he posits that humans must ultimately identify themselves with their minds and not their bodies. He proposes, "it is now manifest to me that even bodies are not properly speaking known by the senses or by the faculty of imagination, but by the understanding only, and since they are not known from the fact that they are seen or touched, but only because they are understood, I see clearly that there is nothing which is easier for me to know than my mind" (Descartes, 1641:12). He posits that the mind is the causal factor for which to account in explaining human thought and behavior, and that the body, including bodily organs like the brain, are only secondarily understood because they are understood *through* the mind.

Cartesian dualism has come to mean, over the years, the position that there is a separate, immaterial soul or mind which operates the physical body from within the head. As Pardo and Patterson (2013) define, "Under this conception, the mind is thought to be some type of non-material

(i.e., nonphysical) entity or thing that is a part of the human being and is somehow in causal interaction with the person's body. The non-material substance that constitutes the mind is the source and location of the person's mental life—her thoughts, beliefs, sensations, and conscious experiences" (43). In the pursuant four-hundred years or so, this stance has fallen considerably out of favor. A popular media article appearing in *Forbes* describes dismissively, "'dualism,' the old–fashioned notion that the mind is something distinct from its mechanism, the brain and the body...This is also known as the 'ghost in the machine' fallacy, the quaint belief that there is a ghostly 'self' somewhere inside the brain that interprets and directs its operations" (Wolfe, 1996).

Ironically, despite the critique that neuro-reductionists rehabilitate a kind of Cartesian dualism, they often position themselves in opposition to such dualism.Greene and Churchland often frame their argument as a trade-off, either the mechanical, causal brain, or some "folk psychological" notion that presupposes a Cartesian "ghost in the machine" or "inner homunculus" – a soul endowed with all human capacities operating the body from within. Greene reifies this dichotomy when he writes, "You *are* your brain...There is no little man, no 'homunculus', in the brain that is the real you behind the mass of neuronal instrumentation" (Greene & Cohen, 2004:17779). Both Greene and Churchland's argumentation is constructed as a refutation of what they understand to be the core tenet of dualism and "folk psychology": a split between the immaterial soul/mind and the physical brain. Greene (2011) begins one of his articles by stating, "Most people are dualists... Intuitively, we think of ourselves not as physical devices but as immaterial minds or souls housed in physical bodies. Most experimental psychologists and neuroscientists disagree, at least officially. The modern science of mind proceeds on the assumption that the mind simply is what the brain does" (263). The assumptions that "folk psychology" presupposes dualism has been reified in various academic works, supported by empirical studies like Demertzi et al. (2009) but challenged in studies like Hook and Farah (2013). Churchland (2004) similarly frames the issue, "Broadly speaking, the evidence from evolutionary biology, molecular biology, physics, chemistry, and the various neurosciences strongly implies that there is only the physical brain and its body; there is no non-physical soul, spooky stuff, or ectoplasmic mind-whiffle" (5). She likens the Cartesian assumption of a "non-physical soul" to non-scientific and non-philosophical terms like "spooky stuff" and "ectoplasmic mind-whiffle." The trade-off is clear, either your position is that "you are your brain" or your position presupposes the supernatural or absurd. Neuro-reductionists posit that they challenge Cartesian dualism because they do not posit a split between the physical body and an immaterial mind/soul. Conceiving of the world in solely physicalistic terms, and conceiving the brain (a bodily organ) as the primary causal factor in thought and behavior,

surely contrasts with Descarte's approach.

So how is it that many critics of neuro-reductionism claim that they rehabilitate Cartesian dualism? Clearly neuro-reductionists do not fall within what is typically construed as dualism because they do not think of the mind and the brain as separate and independent substances. However, Descartes also established another important dualism in his writings, the separation of the interior and exterior world. This dualism is expressed in the notion of the "inner homonculus," a fully operational human mind which was "inside" the skull. Aydin hones in on the issue when he writes:

> …many other advocates of modern (neuro)biological approaches, clearly uphold a distinction between an inside world of cognition and an outside world of material objects but often claim that this 'inside world' is physically realized or constituted. Although Descartes would never have situated cognition in or reduced it to physical processes... the division between 'internal' and external' that breaks along the line of the knowing subject is, as Rouse (1996) has pointed out, a cartesian legacy. (Aydin, 2015:76).

Although to Descartes "the brain" would be part of the body, and therefore fundamentally different from "the mind," modern neuroscientists refer to "the brain," but in such a way that it closely resembles the Cartesian conception of mind. Rather than focus on the dualism between body and mind, these critics focus on the complimentary dualism between the internal world in which cognition occurs and the external world in which action (or behavior) occurs.

These critics point out that often, in a neuro-reductionist framework, the brain is not understood as just another part of the body, but as a separate, inner realm where cognition is mechanically produced. Indeed, the rhetoric of the brain as an "inner realm" is reified in academic and popular literature. For example, deCharms (2008), a neuroscientist writing an academic article, likens the brain to an "inner realm" when he writes, "Real-time functional brain imaging enables us for the first time to look inside our own brains and view the biological underpinnings of our own unique conscious experiences and behaviours, and potentially the causes of psychiatric or neurological diseases" (728). He assumes that consciousness is something "inside" the brain, and that neuroimages can grant access to that "inner realm." This same rhetoric is echoed in popular media articles, for example in an article comparing fMRI images to those produced by the Hubble Telescope: "The latter shows us awe-inspiring vistas of distant nebulae... the former peers the other way, into psychedelic inner space" (Poole, 2012). Although they have overcome the mind/body dualism, these authors reify the internal/external dualism.

So in formal and informal language, there is a recurring theme of thinking of the brain as an *interior* realm where cognition takes place. Noë (2009) summarizes this vision to mean, "According to the now standard view... We really are our brains, and our bodies are at most robotic tools at our brains' disposal... If the truth be told, we are brains in vats on life support. Our skulls are the vats and our bodies the life-support systems that keep us going" (4). By this account, the neuro-reductionists' conception of the brain very much resembles the Cartesian conception of mind. It is an internal world, demarcated by the skull, in which the whole cognition takes place. It is, in a sense, an inner homunculus or a ghost in the machine, but now it is a mechanical ghost, maybe a golem of some kind. As Pardo and Patterson (2013) note in direct response to Greene and Churchland:

> Under this conception, the mind is a material (i.e., physical) part of the human being—the brain —that is distinct from, but is in causal interaction with, the rest of the body. The brain is the *subject* of the person's mental properties (the brain thinks, feels, intends, and knows) and is the *location* of the person's conscious experiences. This conception is deeply problematic. It fails to recognize the proper criteria for the attribution of many of our mental concepts, and it incoherently ascribes psychological attributes to the brain (rather than the person). Notice that this second conception, although repudiating substance dualism, keeps intact the same logical, formal structure. The mind is a kind of entity that interacts with the body (one inner agent is replaced by another: the Cartesian soul by the brain). Conceptual problems arise for neuro-reductionism because the mind-is-the-brain conception is still operating with a formal Cartesian structure. (Pardo & Patterson, 2013:44).

Pardo and Patterson cohere closely with Bennett and Hacker, among others, all of whom argue that the empirical support offered by neuro-reducionists is rendered invalid when the conceptual fundaments upon which their interpretations rest are deemed invalid. When these authors take into question the conceptual fundaments of neuro-reductionism, rather than staying at the practical and empirical level, they claim to find serious inconsistencies, like offering a way to overcome dualism while rehabilitating it in another.

CHAPTER 3: RESOLVING THE PROBLEM

5: Distributed Causalists

This section analyzes an alternative perspective on the brain, mind, body, world relation: distributed causalists. 5.1 begins by characterizing this approach. 5.1.1-5.1.4 explore approaches for

taking into account the causality of the mind, body, socio-cultural-environmental context, and technology, respectively. 5.2 evaluates this approach as an alternative to neuro-reductionism in the legal context.

5.1 Characterizing Distributed Causation

Now that we have reviewed some of the main arguments and critiques articulated for and against neuro-reductionism, we can explore alternative approaches. There are many alternatives to neuro-reductionism, such as other forms of reductionism (i.e. genetic, sociological, behavioral). Since these alternative forms of reductionism also reduce causality to a singular element, I have opted to explore conceptual frameworks which view causality as distributed among several elements as a more high-contrast juxtaposition. Morse asked in his interview with Rosen, "How is this [neuro-reductionism] different than the Chicago school of sociology,' which tried to explain human behavior in terms of environment and social structures? 'How is it different from genetic explanations or psychological explanations? The only thing different about neuroscience is that we have prettier pictures and it appears more scientific'" (Morse Qtd. In Rosen, 2007:4). Since several authors have noted that these forms of reductionism are conceptually very similar, I explore approaches which contrast reductionism by invoking multiple explanatory elements. As Rose (2005), states, "there is no conceptual difference between claiming that we are determined through our genes or that we are determined through our childhood experiences and the socio-economic context in which we are reared. In each case, free will would seem to be nothing other than a 'user illusion' (Nørretranders, 1998)—an epiphenomenon to be dismissed summarily, as Churchland does, as 'folk psychology'" (1001). Since most reductionistic frameworks seem to conflict with free will, and free will is such an integral part of the legal notion of responsibility, any reductionist framework is usually perceived as conflicting with the traditional notion of legal responsibility.

The question that remains, therefore, is: is there a way for law to take neuroimages into account which does not entail a reductionist approach? Since it seems that reductionism underlies fears about the conflict between neuroimages and traditional legal notion, are there alternatives? Some explanatory alternatives are consistent with legal presuppositions, but black-box the brain out of an explanation; for example, behaviorism, which has been a popular approach in law. Although behaviorism can attempt to clarify the "mental defects" stipulated by the law, neuroscience and neuroimaging are outside of its domain. For an alternative approach which can take neuroimages into account, I turn to a group which I have terms "distributed causalists." Although "neuro-reductionism" is already an accepted term in the

field, I have formulated this grouping and terminology of "distributed causalists" myself. What joins these different approaches together is their common commitment to viewing causality as something which cannot be explained in terms of any singular element. For them, the neuro-reductionist account is unsatisfactory for many of the reasons enumerated in the previous section, so what kind of account is satisfactory?

One of the main features they all share in common is that they still take the brain into account as a causal factor, just not the *only* relevant causal factor. However, since this thesis is also focused on neuroscience and neurotechnologies, it makes sense to explore theories which do take the brain into account to some degree. Morse (2011), for example, argues that, "It is clear that, at the least, mental states are dependent upon or supervene on brain states, but neither neuroscience nor any other science has demonstrated that mental states do not play an independent and partial causal role" (219). To these scholars, the language of mentalia still expresses a causal element which cannot be expressed solely in terms of "the brain," but this does not mean that these elements are entirely independent the brain. These scholars generally subscribe to an interactive model, where conscious mental states interact with the physiological mechanisms of the brain, both exercising causality (whether fundamental or nonfundamental) on one another, among other potential factors. Morse, as a legalist, clearly wants to preserve the legal notions of responsibility, and therefore maintain the necessary elements for causal agency.

The perspective that the law should take into account both brain *and* mind is not exclusive to legalists. Gazzaniga, for example, establishes a conceptual framework specifically tailored for cognitive neuroscientists:

> This idea represents a fundamental paradigm shift away from the so-called reductionism perspective in which the strongest explanatory power lies at the lowest level of investigation: that is, system phenomena are explained by breaking or reducing the system down into molecules, atoms, particles and then subparticles. However, biological systems such as the brain are fundamentally nonreducible in the sense that nonfundamental components have significant causal power: causation seems to occur both upwards and downwards between multiple levels (either neighboring or distant) of the system creating a complementarity or mutually constraining environment of mental and physical functions. This nonreducibility of the brain might be predicated on its inherent organization, which is not a simple sum of the parts with which it is organized. Nonfundamental causality, unlike correlation or determinism, allows for

> mutual manipulability over levels and multiple realizability of system function. It also provides a framework in which to place human responsibility [61] and relate neuroscientific advances to ethics and law [59,62], a process that poses significant difficulties in the context of deterministic reductionism. (Bassett & Gazzaniga, 2011:204-205)

He maintains that the brain is organized in a "modular" way such that parts of it can perform different tasks while also capable of interacting and exercising causation on other parts. Both "mental" and "physical" functions have causality. He refers to "nonfundamental causality," meaning that the mind cannot exist without some causation from the brain, making the brain fundamentally causal. While the brain *can* physically exist without causation of the mind, this does not rule out the possibility that mental states also exercise causality, exercising nonfundamental causality, but causality nonetheless. He is unwilling to adopt the stance that the mind, particularly conscious states and intentions, are epiphenomenal and non-causal that, therefore, the very notion of responsibility must be revised. Hence why he insistently maintains that the legal notion of responsibility, which is contingent on the "standard conception" of agency, is not under threat from emerging neuroscientific research. In this, he resists Greene's characterization that, "To a neuroscientist, you are your brain; nothing causes your behavior other than the operations of your brain" (Qtd. In Rosen, 2007:4). Gazzaniga is certainly a neuroscientist who performs technically similar experiments to Greene, but gleans very different interpretations and implications from them. While Greene concludes that these experiments invite a total revision of legal theory, Gazzaniga argues that the justice system's traditional notions of responsibility need not be abandoned. This shows that no discipline is monolithic, even if it is characterized as such. Thus, even in the relatively stable lab, among colleagues from the same discipline, everything can differ, from the the interpretations of neuroimages to the significance of the insights they afford.

This section explores several approaches to distributed causality. I demonstrate that there is a spectrum of positions which could be characterized as such. I begin with what could be called the "weak form" of distributed agency, which posits that both the mind and the brain exist and exercise causality on one another. This is a fairly standard view, and at this point, the closest to the view that law adopts. A slightly stronger form of distributed causation posits that the body must also be considered a causal factor. The "strong form" of this view posits that the mind, brain, body, and world all engage in causal interplay with one another. These conceptual alternatives also engender alternative views of agency, the causal efficacy of mental states, free will, and so on. The structure of this section is as follows: First, in 5.1.1, I will explore the "weak form" of distributed causality through the argumentation of Noë (2009). In 5.1.2, I explore a set of arguments proposing that the whole body

should be taken into account in a causal explanation of mental states, supplied mainly by Kutas & Federmeier (1998) and Gallagher (2005). From there, I move to stronger forms of distributed causality in 5.1.3, where I explore the arguments that socio-cultural-environmental context plays a causal role, exemplified by Varela (1996), Dumit (2004), Glannon (2005), Gallagher (2005), and Choudhury et al. (2009). In 5.1.4, I will introduce a set of arguments dedicated to taking the causal agency of technologies into account, articulated by Verbeek (2005), Clark and Chalmers (1998), and Aydin (2015). In 5.2, I will summarize how these various conceptual frameworks relate to the legal notions of free will and responsibility.

5.1.1 Causal Accounts for the Mind

In some ways, an approach like Wegner's could even be classed as a very weak distributed causalist approach. This shows that neuro-reductionism and distributed causality are not incommensurable opposites, but rather different degrees on a spectrum. There are more radical approaches to distributed causality than that of Wegner's, even when the mind and brain are the only elements being taken into account. Noë, for example, proposes that one way to hold the mind and brain as compatible, causal agents is to think of them in different terms. He writes:

> contemporary neuroscience has been in the thrall of a false dichotomy, as if the only alternative to the idea that the thing inside you that thinks and feels is immaterial and supernatural is the idea that the thing inside you that thinks and feels is a bit of your body. It would be astonishing to be told that we've been thinking about consciousness the wrong way—as something that happens in us, like digestion—when we should be thinking about it as something we do, as a kind of living activity. (Noë, 2009:7).

Noë identifies another critique to the argumentation of Greene and Churchland – that they portray a false trade-off between dualistic or physicalistic explanations. Noë argues that both approaches posit that cognition is something which happens “inside” and which can be understood in terms of its constituent processes – and not as something which is interactive, “in the world,” and should be studied in terms of how it is realized in the world rather than it is realized “inside” the brain.

Several conceptual frameworks entail the the argument that consciousness and mentation should be characterized in terms of activities and behaviors which take place in a socio-cultural-environmental context, rather than understood in terms of underlying physical processes. These perspectives offer ways to overcome what some scholars see as the rehabilitation of the Cartesian inside/outside dualism

and broaden the scope of conceptual understandings of human thought and behavior. Noë (2009) frames the issue as such, "In a way our problem is that we have been looking for consciousness where it isn't. We should look for it where it is. Consciousness is not something that happens inside us. It is something we do or make. Better: it is something we achieve. Consciousness is more like dancing than it is like digestion" (xii). He argues that we should seek to explain the relevant features of "the mind" in terms of how we experience them and how they play out in our daily lives, rather than as a contained, determined process occurring "inside" the head.

Just because Noë takes the mind into account as a causal factor does not necessarily mean his views cohere with legal presuppositions. Moore (2011) writes, "As one common law court put it, the law supposes that 'the state of a man's mind is as much a fact as the state of his digestion'" (Moore, 2011:207). This conflicts with neuro-reductionists because it invoked "the mind," but it also contrasts with Noë's approach because it likens mentation to digestion. This goes to show that just because someone is not a neuro-reductionist does not mean that they are committed to rehabilitating traditional notions either. Noë proposes that the very way we explain the mind should change, and that rather than understanding it as some internal process occurring in the head which can be assessed by various experts with privileged access, it should be understood as a social practice, co-constituted through interaction with the world.

The difference between the approaches of Wegner and Noë demonstrates an important difference even among those who posit that the mind *and* brain are causal. Aydin (2016) delineates the difference between an "internalist" account like Wegner's and an "externalist" account like Noë. He writes, "Advocates and opponents of free will do not agree on the question of whether consciousness can be attributed causal efficacy. However, they often share the view that freedom can only be attributed to an agent (consciousness or brain) that is detached or has the capacity to detach itself from its environment or other external influences. They believe that freedom can only exist in a separate, autonomous realm that is not affected or determined by external factors" (Aydin, 2016:8). This relates back to the notion of free choices as those made with no antecedent causes. If nothing "outside" of the mind has any causal influence, this intuitively leads to the idea that the brain or mind constitutes a separate or independent realm. It seems that both incompatibilist approaches to free will – the presumed "folk psychological" approach that we have free will because conscious decisions are made without antecedent causes, and the neuro-reductionist approach that we do not have free will because our decisions are determined by our brains – both arrive at "internalist" perspectives. Aydin (2016)

contrasts this with the "externalist" positions that "attempt to show that what we consider our 'real, inner self' is greatly dependent on factors in the 'outside' world. An important implication of these challenges is that the criteria for what it means to be an 'independent, inner self' are blurred and, hence, the very distinction between 'internal' and 'external' becomes opaque" (8)."Externalist" accounts often result in the conclusion that the division of internal/external is actually irrelevant or even nonexistent. Because the neuro-reductionist account is characterized by being particularly "internalist," and the following section is intended to offer a contrast to this approach, I focus primarily on "externalist" approaches which do not focus on, for example, how consciousness occurs inside the mind (as a true Cartesian would), but rather on how factors beyond the brain, mind, and demarcations of skull and flesh also play important causal roles in human lives, including their mental lives.

5.1.2 Causal Accounts for The Body

It is unclear whether "the brain" is really thought of as part of the body in neuro-reductionist accounts. Some neuro-reductionists claim that they overcome the dualistic separation of mind and body because they take the brain, a bodily organ, into account. Indeed, few neuroscientists would argue that the body is not an important factor. Although the brain is certainly part of the body, some scholars critique that neuro-reductionists do not treat the brain the same way as they treat the rest of the body. The brain is thought to play a "special role in explaining our powers of mind (e.g., thought, memory, perception, emotion, and the like). Indeed, some scientists and philosophers think that the mind *is* the brain" (Noë, 2009:9-10). This privileging of the brain as the site of cognition often means that other bodily factors are often ignored. Some neuro-reductionist arguments give the impression that the brain is an isolated, inner realm in which the whole of cognition takes place, characterizing humans as "brains in vats on life support" (Noë, 2009:4). Kutas and Federmeier put some perspective on this reductionistic exuberance when they write:

> Throughout human history, people in many cultures have sought to more fully understand the mind by understanding its relationship to the body. In so doing, philosophers and scientists have associated the mind with nearly every major internal organ... Modern science now recognizes the brain and the other structures making up the nervous system as the most direct substrate for sensory, cognitive, affective, and motor processing. In the process of landing the mind in the brain, however, we sometimes appear to have forgotten that the brain is both responsive to and responsible for the body in which it is housed. (Kutas & Federmeier, 1998:135).

These authors argue that the causal role that the body as a whole plays in the human experience is often

not taken into account in purely neurological explanations. Neuroimaging, in particular, has a tendency of literally making the body disappear from view.

The privileging of the brain over the body might seem odd, as the brain is part of the body, but it becomes apparent, for example, in the rhetoric of fMRI lie-detection versus polygraph lie-detection. Littlefield (2009) writes, "Proponents of brain-based detection attempt to rhetorically distance themselves from the body, as it was examined in traditional polygraphy and fingerprinting to more accurately record the secret inferiority and intentionality of individuals; the result is not a shift in ideology, but a repackaged psychophysiological approach to a longstanding construct of the mind-brain, what I term the *biological mind*" (366). So while fMRI lie-detection still relies on the presupposition that psychological predicates are physiologically realized (and therefore the mind is biological/physical/mechanical), it is also rhetorically distanced from the rest of the body. She adds, "When researchers argue that brains are the central hub of deception, their rhetoric restricts consciousness, rationality, and intentionality, not to mention experimental methods, to the small material sphere of the physical brain. Such representations also construct the brain as an obliging organ, more compatible with scientific inquiry than the variable and suspicious body" (Littlefield, 2009:369). She emphasizes the idea that the connection between the brain's causality on thought and behavior seems more transparent than that of the body's because the brain is considered more "central" to cognition. There is a general agreement that brain-based lie detection is more accurate than physiological markers, even though statistically, their error rates are the same (Rosen, 2007:6). This indicates the extent to which the brain is privileged over the body, and how the rest of the body can be disregarded when invoking a neurological explanation of thought and behavior. Littlefield (2009) summarizes, "In their claims to objectivity, centralization, and mechanization... fMRI detection attempt[s] to bypass subjectivity, the noncognitive body, and a variety of cultural ideologies. Instead of measuring physiological phenomena such as blood pressure, heart rate, and respiration, brain-based detection focuses on just 'the organ that produces lies, the brain' (Ganis et al. 2003, 830)" (369). This shows that, in many ways, neuro-reductionism can actually veto the body from analysis, rather than bringing it to the fore, as some proponents claim it does. Although the brain is a bodily organ, it is granted privileged status, and therefore often studied in isolation. This has been a recurring subject of critique throughout the history of philosophy of mind. As Wilson and Foglia (2016) write, "In general, dominant views in the philosophy of mind and cognitive science have considered the body as peripheral to understanding the nature of mind and cognition. Proponents of embodied cognitive science view this as a serious mistake" (para. 2). This neglect or isolation of the body also reinforces

the Cartesian notion that the body is a mute vessel for cognition, which does not exercise causation but only has causation exercised upon it.

Various scholars have proposed ways to bring the rest of the body into consideration. This relates to a growing constituent of scholars who subscribe to a theoretical framework called "embodied cognition," which can be defined to mean, "Cognition is embodied when it is deeply dependent upon features of the physical body of an agent, that is, when aspects of the agent's body beyond the brain play a significant causal or physically constitutive role in cognitive processing" (Wilson & Foglia, 2016:para.1). As Gallagher (2005) adds, "The broad argument about the importance of embodiment for understanding cognition has already been made in numerous ways, and there is a growing consensus across a variety of disciplines that this basic fact is inescapable" (1). Another articles begins boldly, "Embodied cognition is sweeping the planet" (Adams, 2010:619). We know that the brain can exercise causal influence on the body (that seems pretty obvious, as it is the organ which regulates the other systems), and also that the body can influence the brain all the way to the level of the conscious experience (i.e. getting moody when you're hungry). There are causal connections that relate the body to the brain without necessarily invoking conscious mental states, like the regulating of sweat or hormones. However, sometimes the connection between the body and conscious mental states seems more directly linked. The two-way causal linkage between the body and mind is more transparent in cases in which, "A bodily condition, such as hunger or overall physical fitness, can significantly alter how we think and what we think about, what we do and what we can do. For example, elderly individuals who exercise regularly show many gains in a variety of mental, and not simply motor, tasks (e.g., Bashore & Goddard, 1993). Cognitive processing, in turn, certainly can affect the body. The same external circumstances can lead to very different patterns of bodily changes in individuals, depending on whether or not they are seen as stressful" (Kutas & Federmeier, 1998:136). This provokes the intuition that the body, brain, and mind are all integrally related and causally acting on one another. Adams (2010) contrasts, "On a non-embodied approach, the sensory system informs the cognitive system and the motor system does the cognitive system's bidding... " (619). This sounds like a typical "internalist" approach to cognition, where the senses communicate information from the outside world to the brain/mind command center, the brain/mind makes a representation and subsequent psychological predicates like intending and believing, and the body quietly obliges to the imperatives of the mind/brain. However, the "embodied cognition" perspective argues that the brain/mind does not only exercise causality on the body, but also vice versa.

A number of empirical experiments have bolstered this claim. In particular, much scholarship has been penned on the role the body plays in language acquisition and comprehension. Whole books are dedicated to the matter, for example, *Grounding Cognition:The Role of Perception and Action in Memory, Language, and Thinking* (2005). This book includes several studies on embodied cognition, one of which relates to a series of studies performed by Arthur Glenberg, a psychologist at Arizona State University. He has performed several particularly influential experiments which seek to demonstrate the role of the body in understanding language. In one of his more accessible experiments, Glenberg asked his experimental subjects to judge the sensibility of several statements. Some of the sentences were intentionally "pleasant," i.e. "The college president announces your name, and you proudly step onto the stage" and "unpleasant," i.e. "Your supervisor frowns as he hands you the sealed envelope" (Glenberg et al., 2010:122). As they were given these statements, a pencil was positioned in half of the subjects' mouths as to make them resemble a smile, while for the other half, it was positioned to make a frown. The results of this experiment showed that pleasant sentences were sensible to the subjects whose mouths were positioned in a smile in less time than those whose mouths were positioned in a frown. Additionally, unpleasant sentences took longer to become sensible to those subjects whose mouths were positioned in a smile than those whose mouths were positioned in a frown. This is just one of several experiments aimed to illustrate the extent to which our bodily movement, positioning, and sensation plays an integral causative role in our cognition. Zwaan and Madden, authors who also appear in the book *Grounding Cognition,* add, "there are no clear demarcations between perception, action, and cognition. Interactions with the world leave traces of experience in the brain. These traces are (partially) retrieved and used in the mental simulations that make up cognition. Crucially, these traces bear a resemblance to the perceptual/action processes that generated them" (Zwaan & Madden, 2005; qtd. In Adam, 2010:624). Where the sensori-motor-perceptual capacities of the body ends and the perceptual-representational-intentional capacities of the mind/brain begins is an unclear demarcation. These authors suggest that neurological and cognitive explanations are not necessarily exclusive from the body, and that increasing scholarship and experimentation should go to testing how the body shapes cognitive processes like language acquisition and exchange, problem solving, and forming the intentions which underly agency.

As one response to this, rather than simply dismissing neuroimages as misleadingly neuro-reductionist, Kutas and Federmeier proposing using as many different approaches to studying these complex causal links as possible. They understand that these complex issues can be framed in so many different ways that *some* framing is required, and framing will necessarily emphasize some elements

while minimizing others. However, they propose to mitigate these limitations by encouraging mixed-method approaches, claiming that, "Understanding how the brain and body act together as a single system to carry out even routine activities clearly requires the information and constraints provided by multiple techniques, even if combining them in any meaningful way raises a whole new set of problems" (Kutas & Federmeier, 1998:137). Kutas and Federmeier cohere with Gallagher in that they both see inquiry into these causal connections as necessarily interdisciplinary. They also all advocate for more stable ways to exchange and interact between disciplines. Gallagher's whole book, *How the Body Shapes the Mind*, is introduced as such:

> There is still a need to develop a common vocabulary that is capable of integrating discussions of brain mechanisms in neuroscience, behavioral expressions in psychology, design concerns in artificial intelligence and robotics, and debates about embodied experience in the phenomenology and philosophy of mind. This book helps to formulate this common vocabulary by developing a conceptual framework that avoids both the overly reductionistic approaches that explain everything in terms of bottom-up neuronal mechanisms, and the inflationistic approaches that explain everything in terms of Cartesian, top-down cognitive states. (Gallagher, 2005:1-2).

Both authors acknowledge that interdisciplinary approaches come with their own difficulties, namely, the difficulty of bringing together different disciplinary vocabularies, practices, conceptual frameworks, and so on. However, they find that the benefit is that an interdisciplinary approach yields more holistic, robust, and adaptable insights. Gallagher also adds that that bringing the role of the body as a whole to the fore helps to strike a conceptual balance between neuro-reductionism (everything is the brain) and outmoded forms of Cartesian dualism (everything is the mind).

5.1.3 Causal Accounts for Social-Cultural-and-Environmental Context

Another set of scholars, often overlapping with the previous set, argue that socio-cultural-environmental context is also an important factor for which to account in a robust causal explanation of human thought and behavior. While approaches which emphasize the body can still maintain an "internalist" approach, they are often complementarily paired with an "externalist" approach to taking socio-cultural-environmental context into account, recognizing that these are interrelated elements. Gallagher et al. (2015) connects this approaches when they lament, "In contemporary discussions... the only elements that tend to be relevant in regard to understanding experience are either mental states

(e.g., beliefs, thoughts, desires) or brain states or processes. Supposedly, anything else, such as some extra-neural bodily state, or some aspect of the environment, or some force of culture doesn't even enter into the explanation. All of these things have already been either reduced or eliminated, or in any case excluded from contributing to an explanation" (155-156). Gallagher et al. point to the same rehabilitation of the inside/outside distinction that several other authors critique with regards to neuro-reductionism. Cognition, mentation, and consciousness are considered isolated, inner processes which supervene on the outside world without the outside world exercising symmetrical causation.

There are quite a few approaches and analytic angles one could take to explain the causal role of socio-cultural-environmental factors. From these, I have chosen to focus on a "strong form" of externalist approach situated in the context of a broad approach known as "neurophenomenoloy." Again, this choice is made for the sake of juxtaposition, and this approach provides an alternative, or complimentary, way of explaining the mind/brain/body/world relationship. It was even developed by Chilean biologist, neuroscientist, and philosopher Fransisco Varela in direct response to neuro-reductionist approaches. For this reason, this approach makes an effective foil because it takes up many of the same issues, but frames them in a radically different way. In some ways, this approach can be thought of as diametrically opposed to neuro-reductionism because neuro-reductionism attempts to reduce all subjective notions about the mind to objective facts about the brain, while neuro-phenomenology is thoroughly committed to the non-reducibility of subjectivity to objectivity.

To understand the neuro-phenomenological approach, we must understand some basic characteristics of broader field of phenomenology, a long-established school of philosophy, the inauguration of which is often credited to Husserl, and includes philosophical titans like Heidegger and Marleau-Ponty. Historically, this viewpoint emerged as a challenge to scientific positivism (the view that science neutrally discloses reality). In some ways, neuro-reductionism resembles scientific positivism because it reifies the idea that, with proper methods and technologies, we can reduce the quagmire of subjectivity into objectivity of brain processes. Phenomenology attempts to resist scientific reductionism by first recognizing that science does not describe "the thing in itself," but rather a mediated object which has been framed in a specific, often limited, way. Furthermore, "among philosophers the insight grew that the human experience of reality is always mediated" (Verbeek, 2005:105). Perception, the process through which humans come to experience the world, necessarily mediates the world. The most important feature about phenomenology is that it does not take as its primary object of analysis "the thing in itself" (a.k.a. the noumena), but rather "the thing to you" (a.k.a.

the phenomena). Varela (1996) characterizes this approach as "more than anything else, a style of thinking … a special type of reflection or attitude about our capacity for being conscious" (335). Rather than being understood as a detailed plan for what to think about the grand metaphysical questions of philosophy, phenomenology should be understood as a way to take into account the indispensable, irreducible element of subjectivity in human lives.

So what does this position entail for "neurophenomenology"? This is an approach to studying the brain/mind/body/world *through* first-person experience, embracing subjectivity rather than attempting to gain objectivity through reduction. Varela (1996), the inaugurator of the term "neurophenomenology," defines it as such: "Neuro-phenomenology is the name I am using here to designate a quest to marry modern cognitive science and a *disciplined approach* to human experience, thus placing myself in the lineage of the continental tradition of phenomenology" (330; his italics). Rather than attempting to reduce experience to the outcome of preceding processes, and therefore viewing experience as a secondary and/or irrelevant factor, neuro-phenomenologists take experiences as their starting point.

Varela posits that the first-person experienced cannot be eliminated or reduced from explanation. He argues, "any science of cognition and mind must, sooner or later, come to grips with the basic condition that we have no idea what the mental or the cognitive could possibly be apart from our own experience of it" (Varela, 1996:333). Rather than viewing an experience as reducible to determined brain processes, Varela posits that experience cannot merely be reduced out of an explanation because it is the only conduit through which we can even begin considering the role of brain processes. To not account for the *experience* of mentation is to obscure the way by which you reached your conclusions, or even began your inquiry. Gallagher et al. (2015) argue, "Indeed, it would be methodologically impossible to engage in neuroscience without referring in some way to behavior or experience, and since the latter is first-personal, it requires that the account be first-personal, or include the first-personal perspective in some respect" (9). A detailed analysis of the subjective first-person experience can serve to clarify what it is the neurological explanation actually aim to explain. Varela actually positions himself in direct contrast with Churchland's approach, which he categorizes, in coherence with my taxonomy, as neuro-reductionist. He sees this approach as intent on "eliminating the pole of experience in favour of some form of neurobiological account which will do the job of generating it" (Varela, 1996:333). While Churchland wants to eliminate "folk psychological" accounts, which often include first-person experiences, neurophenomenological approaches seek to make it the

central object of inquiry. Neuro-phenomenology offers a way to critically reflect on "folk psychology" without positioning itself in opposition to it.

That does not, however, mean that the approach is just an unsystematic report of subjective experiences. As Varela (1996) defines, neurophenomenology is characterized by a *disciplined approach* to first-person experience (330). Gallagher et al.'s (2015) book, *A Neurophenomenology of Awe and Wonder: Towards a Non-Reductionist Cognitive Science,* demonstrates how such a disciplined approach might be performed. The book focuses on two subjective experiences: awe and wonder. These were systematically selected because they are conceivably universal across humankind, with the vast majority of known peoples having some experience of them. Although everyone can relate to these experiences, they have hardly been the study of systematic investigation. In order to spotlight this experience, Gallagher et al. focused on the experiences of astronauts, who universally report an experience of awe and wonder on their journeys. An interdisciplinary team of psychologists, neuroscientists, philosophers, engineers, and simulation experts came together to systematically inquire into the phenomenon. Using the astronauts' journals, they did a "hermeneutical analysis," in which the focus was on "the experiences themselves, and we were concerned to define in very precise terms, the different forms that such experiences took" (Gallagher et al., 2015:4). They combined this with a syntactical analysis in which they gathered statistical data on specific words and phrases and when/in what context they occur. So although scholars sometimes levy the criticism that phenomenology can result in idle and disordered introspection, Gallagher et al. offer a structured method through which to approach the phenomenology of the mind/brain/body/world.

Gallagher et al. proffer that neuro-phenomenology offers indispensable conceptual tools for analyzing the role society, culture, and environment plays not just on how we understand cognition, but also on cognition itself[10]. Ultimately, these authors summarize in direct resistance to neuro-reductionism:

> Since cognition is embodied and situated in rich social and cultural environments, not all causal, or constitutive factors of experience are to be found simply in the brain. An integrative cognitive science attempts to grasp as many of these non-neural factors as possible, without ignoring the important role of brain processes. Even to understand what the brain is doing, however, we need the broader picture that involves experiential, embodied, socially and

[10] For a more detailed view on the specific role of society, culture, and environment in Gallagher's framework, the reader may also be interested in "The Socially Extended Mind" (2013).

> culturally situated factors that contribute to make each person's experience what it is. (Gallagher et al., 2015:173).

This coheres with several points made by the previous authors Noë (2009) and Kutas and Federmeier (1998), who both argue that the neuroscientific gaze can be broadened to include factors beyond the brain. Increasingly, empirical and neuroscientific studies are also taking this assumptions as a starting point, inquiring into the extent to which the socio-cultural-environmental context actually alters the physiology of the brain. Aydin (2013) mentions, "there are various studies that indicate that there can also be influence in the opposite direction: socio-cultural practices can reshape certain cortical areas of the brain or transform the brain's representational capacities (see Näätänen et al. 1997; Wheeler 2004; Dehaene et al. 1999)" (8). This shows that the insights from neurophenomenology do not necessarily conflict with "third-person" neuroscience. Socio-cultural-environmental influence can be studied on the level of brain physiology, neurophenomenology, and through other methods, like those of behavioral or cognitive psychology. Each of the different framings add new insights, but most of these authors would argue that no one view could suffice to account for the entire, complex interplay of various causal factors.

These authors argue that it is not only important to take socio-cultural-environmental context into account because it plays a causal role in cognition, but also because it plays such an important role in how neuroscientific studies are interpreted, and ultimately, the self-knowledge we glean from them. So not only does our context actively mediate our cognition as it is happening, but it also plays a role in how we *understand* cognition as we reflect on it. Many authors draw attention to the fact that neuroimages or "brain facts" are not deployed in a void, but rather in a specific context. The social, cultural, and environmental context plays a significant causal role in how they are interpreted. Chouldhury et al. (2009) argue, "While psychological distress no doubt has manifestations at the level of the brain, the biological claims free the person from the social and cultural complexities surrounding her.... the reduction of psychiatry to neurobiology tends to neglect phenomenological insights, biographical accounts of the person and the meaning—that is, the social, cultural, moral or spiritual significances—of mental illness or interventions" (Choudhury et al., 2009: 71). They acknowledge that, for one thing, they are not rejecting the brain as a causal player. Like all of these authors, the brain still plays a role, just not the only role. They agree that socio-cultural-and-environmental context exercises causality on cognition, but add that the context also influences how cogniton is understood. Choudhury et al. (2009) coined the phrase "neuroscience as a cultural activity" (63), acknowledging that the

science itself, not just its objects of inquiry, is socio-culturally-and-environmentally contingent. Dumit (2004) adds specifics for a social-cultural-environmental analysis of the *interpretations* of cognition:

> We should try to become as aware as possible of the *people* who interpret, rephrase, and reframe the facts for us (the *mediators*). We should also critically assess the structural constraints of each *form* of representation—peer review, newsworthiness, doctor presentations to patients (the *media*). In the case of the brain, these processes of fact translation are caught up in a social history that includes how the brain came to be an object of study in the first place, and what factors—conceptually, institutionally, and technically—were part of its emergence as a fact. (Dumit, 2004:5).

Both Dumit and Choudhury et al. trace the propagation of "brain facts," insights about the brain/mind/body/world relation that spread from the lab and into new fields and formats, getting interfaced, reframed, and reified at every step. They all demonstrate how not only cognition, but also the science and philosophy of cognition, is a necessarily socio-cultural-environmental affair because it occurs in a context and not in a void.

5.1.4 Causal Accounts of Technology

Another set of distributed causalists provide conceptual tools in order to understand the causal role of technology. Like arguments regarding socio-cultural-and-environmental context, their argument is two-fold: technology exercises causal agency on cognition and it also exercises causal agency on how "facts" and insights about cognition are framed and understood. This first claim entails that that technology plays a role in the production of brain states and mental states. Clark and Chalmers articulate an influential theory of cognition known as the Extended Mind Thesis (EMT). They argue for taking certain technologies as active causal agents in cognition. They begin their article by questioning the idea that cognition happens "inside the head," deploying the critique that "internalist" approaches rehabilitate the inside/outside distinction of Cartesian dualism. They assume a stance they term "active externalism," denoting from the outset that they will be arguing for taking into account elements which are beyond the "demarcations of skin and skull" (Clark & Chalmers, 1998:1). This reasoning should be fairly familiar to the reader at this point, but it serves to set the stage for EMT.

They introduce the EMT with a thought experiment, comparing three difference scenarios in which an individual is asked to rotate an object on a screen to fit into a socket. In the first scenario, the

person must simply mentally rotate the object. In the second scenario, he/she has the option to mentally rotate the object, or click a button that will rotate the object on the screen. They assume that this would decrease the time. In the third scenario, he/she has a neural implant which can perform the rotation as quickly as the computer, giving him/her the option to mentally rotate the object or use the neural implant. They them claim:

> How much *cognition* is present in these cases? We suggest that all three cases are similar. Case (3) with the neural implant seems clearly to be on a par with case (1). And case (2) with the rotation button displays the same sort of computational structure as case (3), although it is distributed across agent and computer instead of internalized within the agent. If the rotation in case (3) is cognitive, by what right do we count case (2) as fundamentally different? We cannot simply point to the skin/skull boundary as justification, since the legitimacy of that boundary is precisely what is at issue. (Clark & Chalmers, 1998:10-11).

If one is troubled by the notion that cognition occurs "inside the head," it becomes difficult to make a distinction between the three cases. This entails that a technologically-mediated cognitive task is largely the same as an un-technologically-mediated cognitive task. They summarize that "If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process" (Clark & Chalmers, 1998:11). This means that when technology performs a task which is analogous to a cognitive task (i.e. making a calculation), it becomes a causal agent in cognition.

The EMT approach places boundaries on the technologies which qualify as active causal factors in cognition. These boundaries are most succinctly summarized in the Parity Principle: "Processes in the external world can only be included as part of an individual's cognitive process, if: (1) the resource is reliably available, (2) the retrieved information is automatically endorsed and (3) the information contained in the resource is easily accessible" (Aydin, 2013:7). A notebook or memorandum which one keeps on him/her at all times, relies on heavily, and is easily accessible would therefore qualify as part of the cognitive system, because it is, for all intensive purposes, an extended memory. However, our old DSL connections which took forty minutes to (maybe) connect, and six hours to download a jpeg, would not qualify as part of our cognitive system. They explicate with regards to the former set of cases, "In these cases, the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in

the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain" (Clark & Chalmers, 1998:11). They presuppose that both the brain and the mind play a causal role, already assuming a non-neuro-reductionist stance. They then add that, when technology executes a task analogous to a cognitive task, it also plays an active role in cognition, as much as any part of the brain plays an active causal role.

Aydin expands on this argument in his articulation of Artifactual Mind Thesis (AMT). Rather than limiting the technologies which play an active causal role to those which resemble cognitive tasks, he argues that the experience of any technology influences cognition. Taking EMT as his starting point, he argues that Clark and Chalmers do not take their critique of the Cartesian inside/outside distinction far enough. Hequestions why technology must resemble or imitate cognition in order to play a causal role in it. He links this to actually rehabilitating the inside/outside distinction when he writes, "Although cognition can be, according to EMT, extended by technical artifacts, an inner-outer dualism is, as I will show, preserved by ascribing to cognition an original starting point in an internal biological core, an inside that utilises the outside world in order to fulfil certain cognitive tasks that it has set for itself" (Aydin, 2015:1). The very vocabulary of "extension" assumes that there is a center from which to extend. Aydin challenges the idea that cognition is "centered" in and "extended" from the head. He questions whether there even exists such as thing as the "unextended mind." Clark and Chalmers still assume that, without the presence of this particular type of technology, cognition happens "inside the head."

Aydin suggests an alternative stance on cognition in which it is *always* in causal interplay with technology; wherein the mind, brain, body, and artifactual context co-shape one another. He summarizes AMT to mean, "Acknowledging that our thinking has an artifactual character means recognizing that external objects and technical artifacts, rather than being utilised by an inside world, have shaped and are continuously shaping the very fabric of our thinking, of what we take to be our 'inside world.' Not only are thoughts exosomatically embodied but the specific physical characteristics of artifacts also activate new modes of thinking" (Aydin, 2015:1). He claims that the very presence of technological artifacts changes the content of cognition. He challenges the foundational view that cognition occurs "inside the head," claiming that mentation necessitates some kind of interaction with the "outside" world. By recognizing that the "internal" world of the mental and the "external" world of the artifactual are in a constant causal interplay, and fundamentally contingent on one another, the very

distinction between "inside" and "outside" is unsustainable. Our "internal" cognitive world would not exist without an "external" artifactual (and socio-cultural-environmental) world with which to interact. For example, children raised in extreme isolation do not demonstrate the same panoply of cognitive capacities as children raised in a socio-cultural-environmental-artifactual context. Quite likely, their brain physiology would differ as well. This demonstrates that causation is multi-directional between the brain, mind, body, and world.

Another way to take account of the causal role of technology is to recognize the influence it has on how we understand and interpret research about cognition. Using the works of authors like Don Ihde, Verbeek supplies a number of conceptual tools to understand the question: "what role do technological artifacts play in the manner in which human being interpret reality" (Verbeek, 2005:121). Verbeek uses an approach he terms "postphenomenology" to focus on how technology plays a causal role in interpretation, experience, and perception. Situated within the phenomenological tradition, his approach shares some key commitments with the phenomenological framework of neurophenomenology. In particular, he shares a commitment to the notion that we experience things as they are to us and not as they are in themselves. Perception is always mediated, and subjectivity cannot be reduced into objectivity. His works contribute less to understanding the actual causal role of technology on mental states, and more to understanding the "cultural activity" side of this issue – the causal role of the technology on how we interpret and act upon insights about cognition.

Verbeek outlines dimensions through which to analyze this mediating role of the technology. One is the hermeneutic perspective, which demonstrates how "things can mediate the ways in which humans being have access to their world by the roles that such things play in human experience. Questions such as the following arise: In what way do telescopes and electron microscopes, automobiles and airplanes shape our access to the world" (Verbeek, 2005:119). This approach also entails asking questions like: how does fMRI mediate our access to the world and to ourselves? This provokes the question as to whether we would even think of ourselves in neurological terms if it was not for the presence of such neurotechnologies. Is it possible to think "you are your brain" without brain-imaging technologies? An fMRI does not merely reproduce an activity in the brain, it mediates how we think about the brain, and, in turn, how we understand ourselves. As Verbeek claims, by viewing something as mediated through a technology, the perception is transformed. He writes, "This transformation of perception has... a definite structure involving amplification and reduction. Mediation always strengthens specific aspects of the reality perceived and weakens others" (131). These authors

draw attention to how technology itself shapes our cognition and self-knowledge. A brain, as viewed through an fMRI, is viewed in a specific way. Our sensitivity to BOLD-levels is amplified (something which we cannot do with the naked eye), but our sensitivity to other neural activity or other causal elements, for example, is reduced. Aydin (2016) also identifies that on a more conceptual level, our identification of ourselves with our brains is amplified, but our identification of ourselves with other causal factors is reduced (1).

Verbeek dedicates some specific writings to the role of imaging technologies in his postphenomenological/ Technical Mediation framework. He posits that, "perhaps the most expansive set of studies in postphenomenology regards the use of imaging technologies in scientific and medical practice" (32) and supplies eleven related works. In specific, neuroimaging is thought to represent an excellent example of hermeneutic human-technology relations at work. He summarizes,

> Broadly speaking, a user's relationship to an image in science can be understood as a hermeneutic relation. That is, a user can be understood to share a reading-style relationship with an image. An imaging device transforms an otherwise imperceptible aspect of the world into a readable form – an image. As the use looks directly at and interprets the image readout, she or he receives a transformed experience of the world. (Rosenberger & Verbeek, 2015:33).

He focuses on an observation which has permeated the neuroimaging literature throughout – neuroimages are not neutral images of the brain, but rather, elegant translations of complex statistics and conceptual presuppositions. Other works have also argued for taking account of the active role fMRI plays in the production of neuroimages and the interpretations (particularly neuro-reductive interpretations) of these images, such as Roskies (2007), Pirruccello (2012), and De Vos (2014).

Verbeek and Rosenberger elaborate two primary insights gleaned from understanding the active role neuroimaging plays when mediating perceptions of the brain. The first is that "scientific images are clearly *not* a simple encounter with the world itself. Images and imaging technologies are better understood... *as* transformative mediators of experience" (Rosenberger & Verbeek, 2015:33). The second insight relates to the role of "human bodily perceptual experience" (Rosenberger & Verbeek, 2015:34). These images demonstrate that an expert familiar with the practice interprets them along "sedimented" lines of reasoning, reflexively contextualizing them in a conceptual framework. The readings of the lay-person may not be so reflexive, and "he or she may need to concentrate and slowly decode the meaning of the colorful brain-shaped display" (Rosenberger & Verbeek, 2015:17). Moreover, he/she may not glean the same meanings as the expert, and even within their demographics,

they may not share the same interpretations.

The observation that different stakeholders may have different meanings and interpretations for the same technology underlies the concept of "multistability." The observation of divergent meanings for the same technology leads one to question: "how should we understand the way that technology at once in part determines our choices and actions, and yet at the same time itself remains open to our manipulations and interpretation?" (Rosenberger & Verbeek, 2015:25). While understanding that a technology mediates phenomenon in specific ways, it is also apparent that this mediation does not amount to determination, otherwise the relationship to the technology would always be the same. The idea of "multistability" "simultaneously points to the fact that the materiality of the device constrains the potential relations to only certain uses and meanings. That is, a technology cannot simply mean anything or do anything; only some relations prove experientially stable" (Rosenberger & Verbeek, 2015:25-26). With regards to neuroimages you cannot just use them any which way, for example, to analyze the heart (except maybe in a metaphorical sense) or lungs. They always mediate taking account of the brain as a causal factor – in this sense they are stable. However, they do not always mediate taking *only* the brain into account – in this sense they are "multistable."

The other dimension Verbeek explicates is the existential dimension. He defines this as such: "things mediate human existence. Here a different set of questions arise: How does the television set affect the way we divide our day?" (Verbeek, 2005:119). This is less about interpretive frameworks, and more about how this relationship actually plays out in the world. Existential questions regarding fMRI might include: how has the introduction of neuroimages changed American legal practices? If, for example, fMRI is operationalized to screen for juror biases (which is a proposed application), it would have a very real effect on "the way in which the material environment of the human being shapes the way in which they realize their existence" (Verbeek, 2005:119). Already, the introduction of neurotechnologies into law has provoked stakeholders into re-examining the legal notions of causality, agency, responsibility, and free will. Real-world practices, indeed the very performances of human existence, are influenced by the presence and use-practices of technologies. This thesis has demonstrated that the real world practices of law are changing due to the introduction of neuroimages, and as these practices change, so too do they enable, provoke, or otherwise mediate changes at the conceptual or perceptual level. In taking account of the causality of technology, technology can be understood as a causal factor on a number of levels, from cognitive states, to interpretive frameworks, to everyday practices and habits.

5.2 Implications for Law

Since the proponents of distributed causality are themselves a diverse group, the implications for law which these perspectives entail are equally diverse. Many of the aforementioned theorists are not making dedicating inquiries into law, so rather than explicate the implications for each sub-section (mind, body, socio-cultural-environmental, and technology), I will sketch some general implications. Future scholars could certainly make dedicated analyses of each perspective and how it could bear on specific issues of law, but that is an enormous task best left aside for the present, where the priority is breadth over depth. Thus far, these authors rarely mount challenges to legal theory, let alone challenging retributivism itself. They often assume what Morse (2011) terms an "internal" mode of critique, which he defines: "An internal contribution or critique accepts the general coherence and legitimacy of a set of legal doctrines, practices or institutions and attempts to explain or alter them. For example, an internal contribution... may suggest the need for doctrinal reform of, say, the insanity defence, but it would not suggest that the notion of criminal responsibility is itself incoherent or illegitimate" (214). He contrasts these with "external" critiques which take the very legitimacy of legal notions of responsibility into question. Sometimes neuro-reductionist arguments are thought to perform such an "external critique," but one rarely finds such claims among proponents of distributed causality.

The more general implications of assuming a distributed stance on causality are, for example, that multiple explanations are more robust than reductionistic ones (whether neuro-reductionistic or otherwise). Gallagher et al. (2015) argue for "a non-reductionist position that, more positively, is close to what Sandra Mitchell (2002) calls 'integrative pluralism'... In the case of integrative pluralism... we have multi-scale explanations involving factors at various scales (neuroscientific, psychological, phenomenological, social, and so on) all contributing to an integrated explanation" (156-157). Rather than just having various disciplines working alongside one another, independently working to solve the same basic set of problems, these authors propose that they should work together, explaining phenomenon on multiple levels. This further emphasizes that these authors generally do not frame their position as a trade-off between studying the brain, the mind, the body, or any other element – the majority of these authors propose that the more elements you can take into account, the more robust the explanation. They offer that the difficulty of integrating different disciplinary vocabularies is outweighed by the difficulty of reducing away the whole vocabulary of the mental. For example, in the case of lie-detection, it might be most accurate if one took a multi-modal approach which included various indicators, such as polygraph, fMRI, brain-fingerprinting, and so on. A question in which all the

various methods cohered would seem particularly authoritative, and a question in which there were discrepancies might provoke crucial considerations about the reliability and variability of these approaches.

This means that neurological explanations can have a place in the courtroom, just not as stand-alone evidence. Even Morse, often considered the diametric opposite of Greene (the two often cite each other as foils against which to argue), an outspoken critic of neuro-reductionism, and coiner of the term "brain overclaim syndrome," concedes that there are legitimate roles which could be strengthened by neuroscience. He writes, "there are four types of situations in which neuroscience may be of assistance: (1) data indicating that the folk-psychological assumption underlying a legal rule is incorrect, (2) data suggesting the need for new or reformed legal doctrine, (3) evidence that helps adjudicate an individual case, and (4) data that help efficient adjudication or administration of criminal justice" (Morse, 2011:227). He advises that neurological explanations should work together with behavioral explanations, which is the more traditional mode of explanations for law (i.e. he is insane because he behaved insanely). He asserts that neuroimages can play a largely supportive role to robust, convincing behavioral explanations (Morse, 2011:223). So proponents of distributed causality do not necessarily seek to discredit neuroscience or neuroimages. However, they recognize that neuroimages are not self-explanatory, and they often direct their "internal" critiques to delineating what experts and legalists should need to make clear to a jury in order for neuroimages to play a properly probative role. Goldberg (2011), for example, advises pragmatically that there is enough "robust evidence suggesting that social and economic conditions...are primary determinants of patterns of violence... While there will always be a need to apply population-based evidence to individual cases in criminal law, the public policy consequences of a system which prioritizes individual assessments of criminality and violence to the virtual exclusion, at least in its formal institutions and procedures, of the socialization of violence, is to be criticized and challenged rather than reified" (2). This shows a specific "internal" critique to law which claims that focus on the brain should not preclude taking into account other factors like socio-cultural (including economic) context.

These authors explicate, as aforementioned, an "internal" critique, which does not seek to challenge the fundaments of law. Some distributed causalists argue that an "external" critique, which seeks to challenge the very fundaments of law's "folk-psychological" notions of agency and responsibility, is not even possible on a neuroscientific basis. Morse (2011), for example, writes, "Neuroscience may help shed light on folk-psychological excusing conditions, such as automatism or

legal insanity, but the truth of determinism is not an excusing condition. The law will be fundamentally challenged only if neuroscience or any other science can conclusively demonstrate that the law's psychology is wrong, and that we are not the type of creatures for whom mental states are causally effective" (216). As this whole thesis has explored, neuroscientists have not satisfactorily "proven" that mental/conscious/intentional states do not exist or do not have causal efficacy. Even the famous experiments of Libet and Wegner, which border on this conceptual territory, are far from uncontentious. Gazzaniga, a neuroscientist himself, has also not been swayed by "external" neuro-reductionist critiques. He writes unequivocally, "Even with all of the fantastic comprehension gained about the mechanisms of mind that neuroscientists now have worked out, none of it impacts responsibility—one of the deep core values of human life... with all the knowledge of physics, chemistry, biology, psychology, and all the rest, when the moving parts are viewed as a dynamic system, there is an undeniable reality. We are responsible agents" (Gazzaniga, 2011:15). Any legalist, cognitive scientist, philosopher, or layperson who takes a distributed causalist stance seems to come to the same conclusion – the notion of legal responsibility does not need to change because its presuppositions (agency, free will, causally efficacious mental states) are not necessarily refuted by neuroscientific findings. The empirical, practical, and conceptual challenges issuing from neuro-reductionists are ultimately unconvincing to them. According to proponents of distributed agency, we can still conceive of ourselves as legally responsible, causal agents, even while still taking into account that our neurological make-up plays a role in our thought and behavior.

These authors acknowledge that responsibility and agency is not equivalent to the absence of determinisms. In-keeping with the very core of distributed causality, these authors posit that many elements cause thought and behavior, but mental/intentional/conscious states are causal *enough* that the law can still conceive of individuals as causal, responsible agents. Gazzaniga would surely support the claim that the brain exercises some causal role in our thought and behavior, but does not conclude that therefore we are neurologically determined and must totally re-think our notions of responsibility. Responsibility is not equivalent to the absence of determinisms. If this were the case, few people would be held responsible because he/she could always point to *some* mitigating causal factor. Although some people have made the argument that legal judgements are actually incompatibilist, legal theory continues to presuppose a compatibilist stance. Morse (2011) writes from his legalist perspective, "All behaviour is the product of the necessary and sufficient causal conditions without which the behaviour would not have occurred, including brain causation... If causation were an excusing condition per se, then no one would be responsible for any behaviour. Some people might welcome such a conclusion

and believe that responsibility is impossible, but this is not the legal and moral world we inhabit" (217). The official compatibilist stance of law means that it can take some degree of neurological determinism into account. It demonstrably has in cases like Hinckley and Weinstein.

The compabilist stance of distributed causalists coheres with the compatibilist stance of law. Wegner (2002) characterizes the incompatibilist perspective as integral to the tension between neuro-reductionists and legalists. He writes, "Those who side with free will view members of the opposition as nothing but *robogeeks,* creatures who are somehow disposed to cast away the very essence of their humanity and embrace a personal identity as automatons... Those who opt for the deterministic stance view the opposition as little more than *bad scientists,* a cabal of confused mystics with no ability to understand how humanity fits into the grand scheme of things in the universe... In each other's eyes, everyone comes out a loser" (Wegner, 2002:319). Indeed, we have seen laypeople show aversion to the deterministic stances of neuro-reductionism, and we have seen neuro-reductionists characterize "folk psychology" commitments to free will as wildly inadequate and misguided. Distributed causality allows for the compatibilism that could potentially resolve this tension. Rose (2005) summarizes how a distributed causalist stance bears on the notion of free will/determinism:

> The conceptual confusion that surrounds determinism and free will is deeply embedded in our way of thinking... In truth, we live at the interface of multiple determinisms... For every action we take, it is possible to define causes at many levels, from antecedent neural events to cultural norms and the financial constraints of a market economy. The important scientific question then is to know at which level it is appropriate to seek an over-determining cause. To understand and hopefully to treat Alzheimer's disease, we need to know about the biochemistry of the amyloid precursor protein, but it would be folly to try to explain the causes of the invasion of Iraq in 2002 in terms of fluctuations in transmitter levels in US President Bush's brain. We are, to summarize my argument, free to act and to shape our own future, although not in circumstances of our own choosing. (Rose, 2005:1004).

This shows that taking a multitude of causal factors into account is consistent with maintaing the "standard conception" of agency and free will, and therefore keeps the legal notion of responsibility intact.

Since neuroscientific challenges to the notion of "responsibility" are empirically and conceptually questionable, and the legal notion of responsibility serves such social utility, these scholars often conclude that it simply makes more sense to maintain this model. Laypeople also cohere

with this viewpoint. For example, one commentator on Griffin's article, "Free Will Could be an Illusion," wrote:

> The one argument in favor of free will no one has used yet is utility. Human life is virtually impossible emotionally or physically without the belief in free will. Free will gives life all its meaning. Without it, there is no morality, there is no love, there are no ethics, and so on. What makes us human is not purely physical, but metaphysical. It may be expressed in physical ways and carried by physical means, but its derivation is that basic truth -- that we are responsible free agents. Are some of our decisions conditioned? Of course. Are all of them? Don't be absurd. (Commentator on Griffin, 2016).

Some of the ideas offered by this layperson cohere closely with a distributed causalist perspective. They also cohere in that they recognize that, even if one upholds the thesis that mental/conscious/intentional states have causal efficacy, this does not mean human thought and behavior is entirely undetermined. They argue that you do not need a notion of absolute, libertarian free will in order to maintain the notion of legal responsibility. He recognizes that human agency can coexist with the recognition of other causal factors. That a thought or behavior has multiple causes does not excuse an individual from responsibility because, ultimately, *all* thoughts or behaviors have multiple causes, so if one took causation as a serious argument for exculpation, no one would be responsible for anything.

Another implication for law which proponents of distributed agency form in response to the perceived threat of neuro-reductionism to retributivism is that predictive law has the potential to be more unjust than retributive law. It seems intuitively problematic to convict someone based on a neurological state rather than an act or behavior. The central issue is that what if the neurological state pertains, but the individual never commits a criminal act? Perhaps in a neuro-reductionist model, if the right brain state were identified, there would be no *possibility* for the individual to act otherwise. However, neuroscience has yet to identify a brain state which invariably leads to criminal behavior. One oft-cited argument is that neuro-reductionists omit the variation between different individuals' brains and their behavior. For example, "If adolescent brains caused all adolescent behavior, 'we would expect the rates of homicide to be the same for 16- and 17- year-olds everywhere in the world — their brains are alike — but in fact, the homicide rates of Danish and Finnish youths are very different than American youths'" (S.J. Morse qtd. In Rosen, 2007:4). Sixteen and seventeen-year-olds share similarly undeveloped prefrontal cortexes (a condition associated with poor impulse control and criminal behavior), yet they do not display the same rates of criminality.

Many authors observe that the connection between brain physiology and behavior is not a simple one-to-one causal relationship. Furthermore, neuroimaging technologies are not nearly robust to identify these relationships with enough fealty to be the grounds upon which incarceration is justified. Aharoni et al. note, "neuroscience cannot demonstrate that all acts are determined. One reason is that most neuro-scientific studies reveal only correlations rather than causation. Even studies that find neural causes do not prove that those causes are deterministic, and they clearly do not generalize to all actions of all sorts" (Aharoni et al., 2008:147). Other scholars agree that the causal linkages are not robust enough to support guilt or incarceration without evidence that an actual crime had been committed, and not just that the neural markers for criminality pertained. Crawford elaborates:

> if you want to predict whether someone is going to break the law in the future, a picture of his brain is no better than a record of his past behavior. Indeed it is quite a bit worse, as the correlation of future behavior with brain abnormalities is weaker than it is with past behavior. Neuroscientist Michael Gazzaniga writes... 'most patients who suffer from . . . lesions involving the inferior orbital frontal lobe do not exhibit antisocial behavior of the type that would be noticed by the law.' It is merely that people with such lesions have a higher incidence of such behavior than those without. So for the pragmatic purpose of predicting behavior, the story of neurological *causation* that is told by pointing to an image of a brain merely adds a layer of metaphysics, gratuitously inserted between past behavior and future behavior despite its lack of predictive power. (Crawford, 2008:76).

At this point, the predictive powers of neuroscience are about on-par with that of behavioral psychology. If someone used your past behavior to make a predictive arrest, it would seem absolutely absurd and unconstitutional, so it seems unlikely that neural signatures will become the new marker for criminality anytime soon. Snead (2007), for example, concludes, "the [neuro-reductionist] project as currently conceived is internally inconsistent and would, if implemented, result in ironic and tragic consequences, producing a death penalty regime that is even more draconian and less humane than the deeply flawed framework currently in place" (1265). So while no one is arguing that law is perfect as is, and they all offer some degree of "internal" critique, they also argue that predictive law on the basis of neurology is empirically and conceptually unsupported, and therefore likely to yield more societal difficulty than utility. While neuroimages may be relevant *after* a criminal act, they need to be much more accurate, perhaps impossibly accurate, to warrant arrest *before* a criminal act.

6: Concluding Remarks

This section concludes the thesis. 4.1 reflects on how this thesis has responded to the research question and other important insights to which it gives rise. 4.2 provides avenues for future research.

6.1. Responding to the Research Question

So what is the yield of this exploration into the conceptual presuppositions regarding causal agency? How can this help resolve the perceived tensions between neuroimages and the traditional legal notion of responsibility? This thesis performed three main steps: identifying the problem, clarifying the problem, and resolving the problem. I identified the problem to be a conceptual tension between neuro-reductionist modes of explaining thought and behavior and the traditional legal notion of responsibility. I clarified that these problem emerges because the legal notion of responsibility presupposes that mental states are causally efficacious, therefore we are agents, therefore we have free will, therefore we can be held responsible. Various approaches to neuro-reductionism cast doubt on each of these presuppositions. The "weak form" only goes so far as to cast doubt on our ascriptions of causal efficacy to mental states, but the "strong form" goes so far as to doubt that mental states have any causal efficacy. The stronger the form of neuro-reductionism, the more it seems to prompt the original author and his/her appropriators to re-frame the notion of responsibility. I propose to resolve this problem by explicating the approaches of distributed causalists. While they can still take the brain into account as a causal factor, their arguments do not amount to undermining the fundaments of the legal notion of responsibility. Their compatibilist stance on free will is consistent with law's compatibilist stance. They still leave conceptual space for agency and causally efficacious mental (intentional, propositional, representational) states.

Ultimately, these different conceptual positions are characterized by their presuppositions regarding the causal relationship between mind, brain, body, and world. Neuro-reductionist posit that the brain is the only causally relevant factors, and the mind, body, and world play more passive or irrelevant roles. Distributed causalists posit several approaches, from "weak forms" of understanding the mind and brain as co-constructive, to "strong forms" which understand the mind, brain, body, and world all in causal interplay with one another. Within these disciplines or stakeholder groups, this relationship is often taken as a given, "black-boxed," and subsequently recedes from the field of argumentation. As Seaman (2009) argues, we tend to "black-box" the human mind and our perceptions about it (931). This thesis should demonstrate that these conceptual notions are anything *but* conceptually stabilized, and therefore black-boxing them or taking such notions as a given might lead

to tacit conflicts. It might seem unfeasible to some readers to consider individuals as neuro-biologically determined, but it also might seem equally unfeasible to others that individuals are entirely freely-willing causal agents. Hopefully, reader who perhaps "black-boxed" their own conceptual stances have a more robust vocabulary through which to understand them, in addition to gaining exposure to some prominent intellectual allies and opponents.

Another benefit of taking the distributed causalist approach towards these neuro-legal issues is that it allows the space to acknowledge the active mediating role of the technology. Throughout this thesis, various authors have demonstrated that neuroimages do not merely accompany neuroscientific explanations, they actively mediate them. It might be more appropriate to say that neuroscientific explanation accompany neuroimages, as they seem to actively instantiate people to consider neuro-reductionist approaches. Methodological and empirical argumentation can draw attention to the fact that the neuroimage requires many interpretive leaps which may or may not be adequately justified and start down the road to the observation that these are not neutral means through which we access an object. However, as Aydin (2016) mentions, these methodological/empirical do not indicate how these images actually mediate (in other words, play an active causal role) our interpretive frameworks and self-conceptions (1). Neuroimaging has influenced the way we perceive of our minds, brains, bodies, and worlds in no small way. It has enabled the dissemination of neuro-reductive frameworks, but also challenges to that framework, as well as alternative frameworks. Only recently, with the advent of "the decade of the mind,"or perhaps more properly, the advent of non-invasive brain imaging technologies, have these legal notions been revisited and even re-imagined. Neuroimaging technologies and the modes of explanation they mediate have provoked us to reify and/or revise our assumptions regarding agency, free will, the causal efficacy of mental states, and ultimately, legal and moral responsibility.

Recognizing the active mediating role of neurotechnologies also enables recognizing the "multistability" of its interpretive content. The reading of neuroimages is highly diverse, and although Verbeek and Rosenberger (2015) posit that experts tend to have more stable interpretations, it is also apparent that even the experts are not in agreement about what can be read from these images. The diversity of different interpretations and the frequent "black-boxing" of conceptual presuppositions gives rise to tacit tensions. Not all the conflicts regarding the proper role of neuroimages in court can be resolved empirically and practically, although these are the only resolutions that can be offered within the courtroom. These conflicts also come down to competing conceptual presuppositions, but, as other scholars have noted, there are fewer dedicated sites to articulate this side of the issue. As Hansen

(2000) notes, "As is the rule in philosophy, much of the advance consists rather in a deeper appreciation of the individual issues and their interconnections. Of course, progress of this kind may itself breed despondency, for it brings with it a realization of just how difficult the problems are" (452). The goal of this thesis was increased clarity in a realm which rarely garners dedicated inquiry – the realm of conceptual frameworks and presuppositions which underly neuroimages and law.

Another important implication that yields from this thesis is to resist monolithic characterizations of the stakeholder groups. Throughout the literature, authors would focus primarily on one stakeholder group, and the others would be characterized in somewhat one-dimensional terms. For example, neuroscientists are often depicted to only think of humans in terms of their brains (i.e. J.D. Greene; qtd. In Rosen, 2007:3; Pardo & Patterson, 2013:28). Indeed, neuroscientists like Greene reaffirm this characterization, but neuroscientists like Gazzaniga resist it. Lay-people are often characterized as being primarily traditionally-minded, "folk-psychological" dualists or incompatiblists (Demertzi et al., 2009; Hook & Farah, 2013; Nahmias, 2006). Yet a cursory look at the comments section on a controversial article indicates that lay-people wrestle with these issues as much as scholars. Philosophers are sometimes depicted as "hating the brain," and wishing the disregard it entirely (Churchland, 2013:13). Yet none of the philosophers reviewed here were entirely dismissive of taking the brain into account in a causal explanation. Legalists were sometimes characterized as eager opportunists, appropriating neuroscientific practices and conceptual viewpoints willy-nilly, sometimes leading to inconsistencies (Pardo & Patterson, 2013; Snead, 2007). All of these stakeholder groups contained individuals who leaned more towards a neuro-reductionist stance and individuals who leaned more towards a distributed causalist stance. Furthermore, even *within* these conceptual grouping, the implications for law varied. For example, Greene and Churchland might share conceptual commitments, but their normative positions regarding law differ.

As aforementioned, this thesis presupposes that the traditional legal notion of responsibility *should* be preserved, and I therefore highlight the coherences between the presuppositions of legal responsibility and the allowances of distributed causalist approaches. However, it is entirely possible that the reader either began this thesis with an intuitive doubt for the legal notion of responsibility, in which case he/she may have a more robust insight into *why* he/she doubts this notion (i.e. on what level of presupposition is his/her doubt). Furthermore, the reader might have begun with an intuitive commitment to the legal notion of responsibility, but found that the evidence and argumentation of neuro-reductionists mounted a convincing challenge. I do not wish to preclude the option that I have also made contentious presuppositions, apt for conceptual challenge. One thing upon which I and all

the authors I have cited can all agree is that there is much more research and reflection to be done on the empirical, practical, ethical, and conceptual levels with regards to this pregnant and provocative topic.

6.2 Avenues for Future Research

This thesis fringes on many topics, so I can see a relation to many avenues for future research. One of the most fascinating emerging fields involves empirical studies about how people of various demographics perceive neuroimages, and what factors influence this perception. For example, Demertzi et al. (2009) and Hook and Farah (2013) performed dedicated studies on whether lay-people were intuitively dualistic or not when interpreting neurosciences. Interestingly, their results conflict, so this is clearly an area that needs more research. McCabe et al. (2011) conducted a fascinating survey on potential jurors and the authoritativeness of fMRI lie-detection, the results of which invite replication and expansion. Another fascinating study along this vein studied the influence of political leanings on the perceived authoritativeness of neuroscientific evidence in court cases (Shen & Gromet, 2015). Such studies consistently yield provocative insights into how perceptions of neuroimaging diverse and converge, and what factors influence these varied approaches. Still little is known about "folk psychology" and lay-person perceptions because it is such a diverse field. Even the representation of lay-people which I included, mainly in the form of news and commentary, is but a snapshot of the diverse opinions and approaches in the world. Our perceptions about neuroscience, technology, each other, ourselves – all of these are being tested in new ways, yielding new insights on the diversity of conceptual commitments, but also the startling recurrence of certain lines of thinking. It seems that the interaction between neuroscience, law, laypeople and philosophers is replete with "multistabilities."

Several of the categorical distinctions established here could be further clarified and investigated. Each category presupposed by law – responsibility, agency, the causal efficacy of mental states, and free will – could be the subject of its own dedicated inquiry, by itself or in relation with neurotechnologies. Further literature could go into what notions of mind, brain, body, and world exist and how they are presupposed or implied by certain statements. In particular, the category of "world" is extremely broad, and could be segmented into different factors – i.e. the socio-cultural-environmental context, the economic context, the artifactual context, geographical context, and/or political context, to name a few. Socio-cultural-environmental context is also a very broad term which could be refined or segmented into specific (externalist) points of view. This thesis has established several lines of questioning, ranging from the specific question "are neuroimages appropriate for their legal

applications?" to the broad question "do neuroscientific explanations challenge concepts like the causality of mental states, agency, free will, and responsibility?" Further research could also go into clarifying how to conduct an interdisciplinary, multi-modal approach to these issues. Each of these questions, and indeed any line of questioning opened by this thesis, could and should be the subject of further inquiry as this field grows in size, complexity, and multistability.

7. Bibliography:

Adams, F. (2010). Embodied cognition. *Phenomenology and the Cognitive Sciences, 9*(4), 619-628.

Aharoni, E., Funk, C., Sinnott-Armstrong, W., & 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(1), 145-160.

Associated Press. (1982, Jun 23). 76 Percent in Poll Against Hinckley Verdict. *The Evening Independent*, pp. 3-A.

Aydin, C. (2015). "The artifactual mind." *Phenomenology and the cognitive sciences*, 73-94.

Aydin, C. (2016, forthcoming). From camera obscura to fMRI: How brain imaging technologies mediate 'free will.'

Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in cognitive sciences, 15*(5), 200-209.

Batts, S. (2009). Brain lesions and their implications in criminal responsibility. *Behavioral sciences & the law, 27*(2), 261-272.

Bear, A. (2016). What Neuroscience Says About Free Will. *Scientific American*. Retrieved from: http://blogs.scientificamerican.com/mind-guest-blog/what-neuroscience-says-about-free-will/

Bear, A. & Bloom, P. (2016). A Simple Task Uncovers a Postdictive Illusion of Choice. *Psychological Science,* 27. pii: 0956797616641943.

Beaulieu, A. (2002). Images are not the (only) truth: Brain mapping, visual knowledge, and iconoclasm. *Science, Technology & Human Values*, *27*(1), 53-86.

Bennett M., Hacker P. (2009). The Argument. In M. Bennett, P. Hacker, D. Dennett, & J. Searle, *Neuroscience and philosophy: Brain, mind, and language* (3-69). Columbia University Press.

Choudhury S. & Blakemore, S.J. (2006). Intentions, Actions, and the Self. In W. Pockett, W.P. Banks, & S. Gallagher (Eds.), *Does Consciousness Cause Behavior?* (39-49)

Choudhury, S., Nagel, S. K., & Slaby, J. (2009). Critical neuroscience: Linking neuroscience and society through critical practice. *BioSocieties*, *4*(1), 61-77.

Churchland, P.S. (1988). *Neurophilosophy: Towards a Unified Theory of Mind and Brain.* Massachusetts, Boston: MIT Press.

Churchland, P.S.. (2004). Moral Decision-Making and the Brain. In Illes, J. (Ed.) *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (3-16). Oxford, UK: Oxford University Press.

Churchland, P. S. (2013). *Touching a nerve: Our brains, our selves*. New York, NY: WW Norton & Company.

Crawford, M. B. (2008). The limits of neuro-talk. *The New Atlantis*, (19), 65-78.

Davidson, D. (2001). *Essays on Actions and Events.* Oxford, UK: Oxford University Press.

Davis, K. (2012). Brain Trials: Neuroscience is Taking a Stand in the Courtroom. *ABA Journal*. Retrieved from: http://www.abajournal.com/magazine/article/brain_trials_neuroscience_is_taking_a_stand_in_the_courtroom/

De Kogel, C. H., Schrama, W. M., & Smit, M. (2014). Civil law and neuroscience. *Psychiatry, Psychology and Law*, *21*(2), 272-285.

De Vos, J. (2014). The Iconographic Brain: A Critical Philosophical Inquiry into (the Resistance of) the Image. *Frontiers in Human Neuroscience,* 8(300), 2-13).

deCharms, C.R. (2008). Applications of real-time fMRI. *Nature Reviews Neuroscience*,*9*(9), 720-729.

Demertzi, A., Liew, C., Ledoux, D., Bruno, M. A., Sharpe, M., Laureys, S., & Zeman, A. (2009). Dualism persists in the science of mind. Annals of the New York Academy of Sciences, 1157(1), 1-9.

Dumit, J. (2004). *Picturing personhood: Brain scans and biomedical identity*. Princeton University Press.

Eagleman, D. (2011). The Brain on Trial. *The Atlantic.* Retrieved from: http://www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520/

Eastman, N., & Campbell, C. (2006). Neuroscience and legal determination of criminal responsibility. *Nature reviews neuroscience*, *7*(4), 311-318.

Eshleman, A. (2014). Moral Responsibility. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* Retrieved from:http://plato.stanford.edu/archives/sum2014/entries/moral-responsibility

Faigman, D. (2015). It's All in the Mind Y'know: Uses and Misuses of Neuroscience in Law. *Exploratorium.* Retrieved from: http://www.exploratorium.edu/visit/calendar/balance-bringing-science-justice-david-faigman-may-21-2015

Federspiel, W. (2007). 1984 Arrives: Thought (crime), Technology, and the Constitution. *William & Mary Bill of Rights Journal*, *16*, 865.

Feigenson, N. (2006). Brain Imaging and Courtroom Evidence: On the Admissibility and Persuasiveness of fMRI. *International Journal of Law in Context,* 2(3), pp. 233-255.

Frank, M. (2009). Reading My Mind. *CBS News.* Retrieved from: http://www.cbsnews.com/news/reading-my-mind/

Freeman, W.J. (2006). Consciousness, Intentionality, and Causality. In W. Pockett, W.P. Banks, & S. Gallagher (Eds.), *Does Consciousness Cause Behavior?* (73-105). Cambridge, MA: MIT Press.

Fuller, V. (2000). United States v. John W. Hinckley Jr. (1982). *Loyola Los Angeles Law Review,* 33, 699-604.

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford, UK: Oxford University Press.

Gallagher, S. (2013). The socially extended mind.*Cognitive Systems Research*, *25*, 4-12.

Gallagher, S., Reinerman-Jones, L., Janz, B., Bockelman, P., Trempler, J. (2015). *A Neurophenomenology of Awe and Wonder: Towards a Non-Reductionist Cognitive Science*

Gazzaniga, M. S. (2005). *The ethical brain*. Washington DC, US: Dana press.

Gazzaniga, M. S. (2006). Facts, fictions and the future of neuroethics. In J. Illes (Ed.), *Neuroethics: Defining the issues in theory, practice, and policy* (pp. 141–148). Oxford, UK: Oxford University Press.

Gazzaniga, M. S. (2011). *Who's In Charge? Free Will and the Science of the Brain*. New York, NY: Harper Collins Press..

Gilbert, F., Burns, L., & Krahn, T. (2011). The Inheritance, Power and Predicaments of the "Brain-Reading" Metaphor. *Medicine Studies*, *2*(4), 229-244.

Glannon, W. (2005). Neurobiology, neuroimaging, and free will. *Midwest Studies in Philosophy*, *29*(1), 68-82.

Glenberg, A., Havas, D., Becker, R., Rinck, M. (2005). In D. Pecher & Zwaan R.A. (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking,* 115-128. Cambridge, UK: Cambridge University Press.

Goldberg, D. S. (2011). Against Reductionism in Law & Neuroscience. *Houston Journal of Health*

Law & Policy, 11, 321-346

Greely, H. (2012). Will Neuroscience Radically Transform the Legal System? *Slate.* Retrieved from: http://www.slate.com/articles/technology/future_tense/2012/10/fmri_in_court_neuroscience_may_ change_the_legal_system.html

Greene, J. & Cohen, J. (2006). For Law, Neuroscience Changes Nothing and Everything, in S. Zeki & O. Goodenough (eds.), 1775-1785. *Law & the Brain*.

Greene, J. (2011). Social Neuroscience and the Soul's Last Stand. In Todorov, A., Fiske, S., & Prentice, D. (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind*. Oxford, UK: Oxford University Press.

Griffin, A. (2016). Free Will Could All Be An Illusion, Scientists Suggest After Study Shows Choice May Just Be the Brain Tricking Itself. *The Independent.* Retrieved from: http://www.independent.co.uk/news/science/free-will-could-all-be-an-illusion-scientists-suggest-after-study-that-shows-choice-could-just-be-a7008181.html

Hagerty, B.B. (2010). Inside A Psychopath's Brain: The Sentencing Debate. *National Public Radio.* Retrieved from: http://www.npr.org/templates/story/story.php?storyId=128116806

Hansen, C. M. (2000). Between a rock and a hard place: mental causation and the mind-body problem. *Inquiry*, *43*(4), 451-491.

Harrop, P.B. (2013) *Minority Report* or Majority Safety? FMRI, Predicting Dangerousness and a Pre-Crime Future (Bachelor's Dissertation). Otago Yearbook of Legal Research. University of Otago: Dunedin, NZ.

Hart, H.L.A. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law.* Oxford, UK: Oxford University Press.

Hart, H.LA., Honoré, A. (1985). *Causation in the Law* (2nd ed). Oxford, UK: Oxford University Press.

Honoré, A. (2010). Causation in the Law. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* Retrieved from: <http://plato.stanford.edu/archives/win2010/entries/causation-law/>.

Hook, C. J., & Farah, M. J. (2013). Look Again: Effects of Brain Images and Mind-Brain Dualism on Lay Evaluations of Research. *Journal of Cognitive Neuroscience*, 25 (9), 1397-1405.

Hubbard, E. (2003). "A discussion and review of Uttal (2001)." *Cognitive Science Online,* 1, 22-33.

Hughes, V. (2010). Science in court: Head case. *Nature* 464, 340-342. Retrieved from: http://www.nature.com/news/2010/100317/full/464340a.html

Insanity Defense. (2016). Legal Information Institute, Cornell University. Retrieved from:

https://www.law.cornell.edu/wex/insanity_defense

Insanity Defense Reform Act (1984). *Public Law,* 98-473. 18 USC 4242.

Jones, O. D., & Shen, F. X. (2012). Law and neuroscience in the United States. In T.M. Spranger (Ed.), *International neurolaw* (pp. 349-380). Berlin, Germany: Springer Science & Business Media.

Kelkar, K. (2016). Can a Brain Scan Uncover Your Morals? *The Guardian.* Retrieved from: https://www.theguardian.com/science/2016/jan/17/can-a-brain-scan-uncover-your-morals

Kiernan, L.A. (1982, May 23). The Mind of John Hinckley. *The Spokesman Review.* pp. 8.

Kim, J. (2007). Chapter 1: Mental Causation and Consciousness. In *Physicalism, or something near enough,* 7-31. Princeton University Press.

Klein, C. (2009). Images are not the evidence in neuroimaging. *The British Journal for the Philosophy of Science*, 61, 265-278.

Krauss, R. (2010). Neuroscience and Institutional Choice in Federal Sentencing Law. *The Yale Law Journal*, *120*(2), 367-378.

Kuersten, A. (2015). Opinion: Brain Scans in the Courtroom. *The Scientists.* Retrieved from: http://www.the-scientist.com/?articles.view/articleNo/44604/title/Opinion--Brain-Scans-in-the-Courtroom/

Kulynych, J. (1997). Psychiatric Neuroimaging Evidence: A High-Tech Crystal Ball? *Stanford Law Review*, 49(5), 1249-1270.

Kulynych, J. (2002). Legal and ethical issues in neuroimaging research: human subjects protection, medical privacy, and the public communication of research results. *Brain and cognition*, *50*(3), 345-357.

Lamparello, A. (2012). Using Cognitive Neuroscience to Provide a Procedure for the Involuntary Commitment of Violent Criminals as a Pat of or Following the Duration of Their Sentence. *Houston Journal of Health Law & Policy*, *11*, 267-302.

Lee, Y. (2014). What is Philosophy of Criminal Law? *Criminal Law and Philosophy*, 8. 671–685.

Littlefield, M. (2009). Constructing the Organ of Deceit The Rhetoric of fMRI and Brain Fingerprinting in Post-9/11 America. *Science, Technology & Human Values*, *34*(3), 365-392.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI.*Nature*, *453*(7197), 869-878.

Mayberg, H. (2010). Does Neuroscience give us New Insight Into Responsibility? In M. Gazzaniga (ed.), *A Judge's Guide to Neuroscience,* (37-41). University of Santa Barbara, CA: University of Santa Barbara Press.

Mayberg, H.S. (1992). Functional Brain Scans as Evidence in Criminal Court: An Argument for Caution. *The Journal of Nuclear Medicine,* 33(6). pp. 18N-25N.

McKenna, M., Coates, D. J. (2015). Compatibilism. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from: http://plato.stanford.edu/archives/sum2015/entries/compatibilism

McCabe, D. P., Castel, A. D., & Rhodes, M. G. (2011). The influence of fMRI lie detection evidence on juror Decision-Making. *Behavioral sciences & the law*, *29*(4), 566-577.

Miller, G. (2009). fMRI Evidence Used in Murder Sentencing. *Science.* Retrieved from: http://www.sciencemag.org/news/2009/11/fmri-evidence-used-murder-sentencing

Mobbs, D., Lau, H. C., Jones, O. D., & Frith, C. D. (2007). Law, responsibility, and the brain. *PLoS Biol*, *5*(4), e103.

Moore, M.S. (2009). *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics.* Oxford, UK: Oxford University Press.

Moore, M.S. (2011). Libet's Challenge(s) to Responsible Agency. In W. Sinnott-Armstrong & L. Nadel (Eds.) *Conscious Will and Responsibility* (207-234). Oxford, UK: Oxford University Press.

Moriarty, J. C. (2008). Flickering admissibility: Neuroimaging evidence in the US courts. *Behavioral sciences & the law*, *26*(1), 29-49.

Morse, S. J. (2005). Brain overclaim syndrome and criminal responsibility: A diagnostic note. *Ohio St. J. Crim. L.*, *3*, 397-412.

Morse, S.J. (2011). Avoiding irrational neurolaw exuberance: a plea for neuromodesty. *Law, Innovation and Technology*, *3*(2), 209-228.

Morse, S.J. (2015). Moore on the Mind. *Public Law and Legal Theory Research Paper Series,* Reseach Paper No. 15-39.

Nahmias, E. (2006). Folk Fears About Freedom and Responsibility: Determinism Versus Reductionism. *Journal of Cognition and Culture*, *6*(1-2), 215-237.

Noë, A. (2009). *Out of Our Heads: Why You Are Not Your Brain and Other Lessons From the Biology of Consciousness.* New York, NY: Hill and Wang Publishers.

Ohikuare, J. [Image of fMRI Scan]. (2014). Retrieved from: http://www.theatlantic.com/health/archive/2014/01/life-as-a-nonviolent-psychopath/282271/

Oullier, O. (2012). Clear up this fuzzy thinking on brain scans. *Nature*, *483*, 7.

Pardo, M. S., & Patterson, D. (2011). Neuroscience, normativity, and retributivism. *The Future of Punishment,* in Nadelhoffer, T. (Ed.). Oxford, UK: Oxford University Press. 1-28.

Pardo M.S. & Patterson, D. (2013). *Minds, Brains, and Law: The Conceptual Foundations of Law and Neuroscience.* Oxford, UK: Oxford University Press.

People v. Weinstein. (591 NYS 2d 715); (Sup. Ct. 1992)

People v. Hinckley, 672 F.2d 115 (D.C. Cir. 1982)

Pirruccello, A. (2012). Reductionism, Brain Imaging, and Social Identity Commentary on "Biological Indeterminacy". *Science and engineering ethics*,*18*(3), 453-456.

Pitt, D. (2013). Mental Representation. In E.N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from: http://plato.stanford.edu/archives/fall2013/entries/mental-representation

Phelps, T.M. (2015). How John Hinckley, Regan's Would-Be Assassin, Could Go Free. *The Los Angeles Times.* Retrieved from: http://www.latimes.com/nation/la-na-hinckley-freedom-20150512-story.html

Pockett, W., Banks, W.P., & Gallagher, S. (2006). Introduction. In W. Pockett, W.P. Banks, & S. Gallagher (Eds.), *Does Consciousness Cause Behavior?* (1-6). Cambridge, MA: MIT Press.

Poole, S. (2012). Your Brain on Pseudoscience: The Rise of Neurobollocks. *The New Statesmen.* Retrieved from: http://www.newstatesman.com/culture/books/2012/09/your-brain-pseudoscience-rise-popular-neurobollocks

Pulse Medical Imaging. [Image of CAT Scan]. (2016). Retrieved from: http://pulsemedicalimaging.com.au/ct-scan/

Raine, A., Satel, S. [Image of PET Scan]. (2013). Retrieved from: https://www.washingtonpost.com/opinions/can-brain-scans-explain-crime/ 2013/06/07/ c88056de-cde8-11e2-8f6b-67f40e176f03_story.html

Reddit. (2013). Can Brain Scans Be Used as Lie Detectors? *Reddit: /r/neuro.* Retrieved from: https://www.reddit.com/r/neuro/comments/z5ulr/can_brain_scans_be_used_as_lie_ detectors_fmri/

Rose, S. P. (2005). Human agency in the neurocentric age. *EMBO reports*, *6*(11), 1001-1005.

Rosen, J. (2007). The brain on the stand. *The New York Times*, 1-9. Retrieved from: http://nyti.ms/1P9m2er

Rosenberger, R., Verbeek, P.P. (2015). A Field Guide to Postphenomenology. In R. Rosenberger & P.P.

Verbeek (Eds.), *Postphenomenological Investigations: Essays on Human–Technology Relations,* 9-42. Lanham, MD: Lexington Books.

Roskies, A. (2006). Neuroscientific Challenges to Free Will and Responsibility. *TRENDS in Cognitive Science,* 10(9), 419-423.

Roskies, A. (2007). Are neuroimages like photographs of the brain?.*Philosophy of Science*, *74*(5), 860-872.

Sallet, J.B. (1985). After Hinckley: The Insanity Defense Reexamined. *The Yale Law Journal,* 94 (6). pp. 1545-1557.

Sapolsky, R. M. (2004). The frontal cortex and the criminal justice system. *Philos Trans R Soc Lond B Biol Sci*, *359*(1451), 1787-1796.

Schleim, S. (2012). Brains in context in the neurolaw debate: the examples of free will and "dangerous" brains. *International journal of law and psychiatry*, *35*(2), 104-111.

Schlosser, M. (2015). Agency. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.* Retrieved from: http://plato.stanford.edu/archives/fall2015/entries/agency/

Seaman, J. (2009). Black boxes: fMRI lie detection and the role of the jury. *Akron L. Rev.*, *42*, 931-939.

Shen, F.X., Gromet, D.M. (2015). Red States, Blue States, and Brain States: Issue Framing, Partisanship, and the Future of Neurolaw in the United States. *The Annals of the American Academy of Political and Social Science*, *658*(1), 86-101.

Sinnott-Armstrong, W., Roskies, A., Brown, T., & Murphy, E. (2008). Brain images as legal evidence. *Episteme*, *5*(03), 359-373.

Smith, K. (2013). Brain Decoding: Mind Reading. *Nature News.* Retrieved from: http://www.nature.com/news/brain-decoding-reading-minds-1.13989

Snead, O. C. (2007). Neuroimaging and the "Complexity" of Capital Punishment. *Scholarly Works*. Paper 542.

Tsakiris, A. (2014). Patricia Churchland Sand-Bagged by Near-Death Experience Questions. *Skeptiko.* Retrieved from: http://www.skeptiko.com/237-patricia-churchland-sandbagged-by-near-death-experience/

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, Massachusetts: The MIT press.

Varela, F. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of*

*Consciousness Studies,*3(4), 330-349.

Verbeek, P.P. (2005). *What Things Do.* University Park, PA: The Pennsylvania State University Press.

Vincent, N. A. (2011). Neuroimaging and responsibility assessments. *Neuroethics*, 4(1), 35-49.

Wakefield (2013). DSM-5: An Overview of Changes and Controversies. *Clinical Social Work Journal*, 41(2), 139-154.

Wegner, D. (2002). *The Illusion of Conscious Will.* Cambridge, MA: Massachusetts Institute of Technology Press.

Wegner, D. (2003).The Mind's Best Trick: How We Experience Conscious Will. *TRENDS in Cognitive Science,* 7(2), 65-69.

Wegner, D., Wheatley, T. (1999). Apparent Mental Causation: Sources of the Experience of Will. *American Psychologist,* 54(7), 480-492.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of cognitive neuroscience*, *20*(3), 470-477.

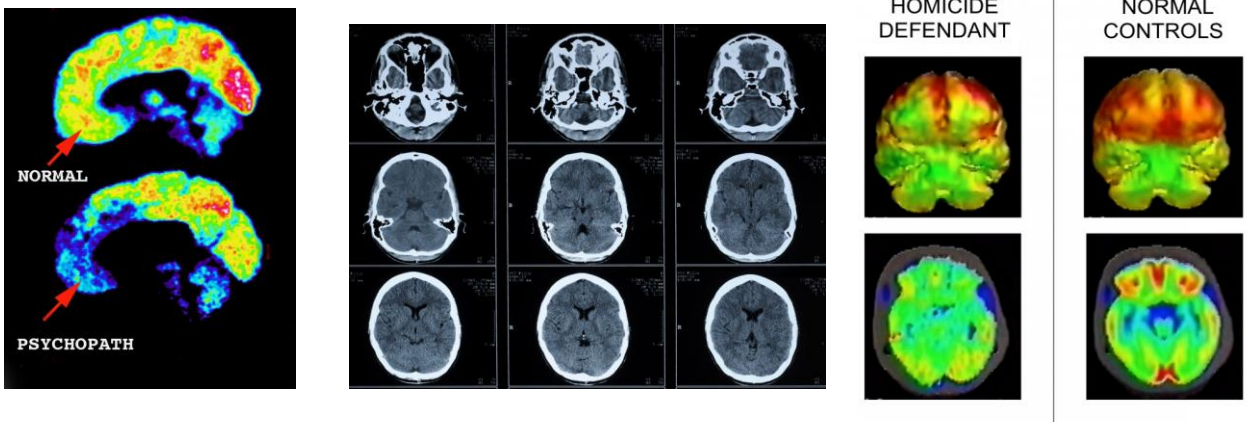
Wilson, R.A., Foglia, L. (2016). Embodied Cognition. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* Retrieved from: http://plato.stanford.edu/archives/sum2016/entries/embodied-cognition

Wolf, S. M. (2008). Neurolaw: the big question. *The American Journal of Bioethics*, *8*(1), 21-22.

Wolfe, T. (1996). Sorry, But Your Soul Just Died. *Forbes.* Retrieved from: http://www.orthodoxytoday.org/articles/Wolfe-Sorry-But-Your-Soul-Just-Died.php

Appendix A:

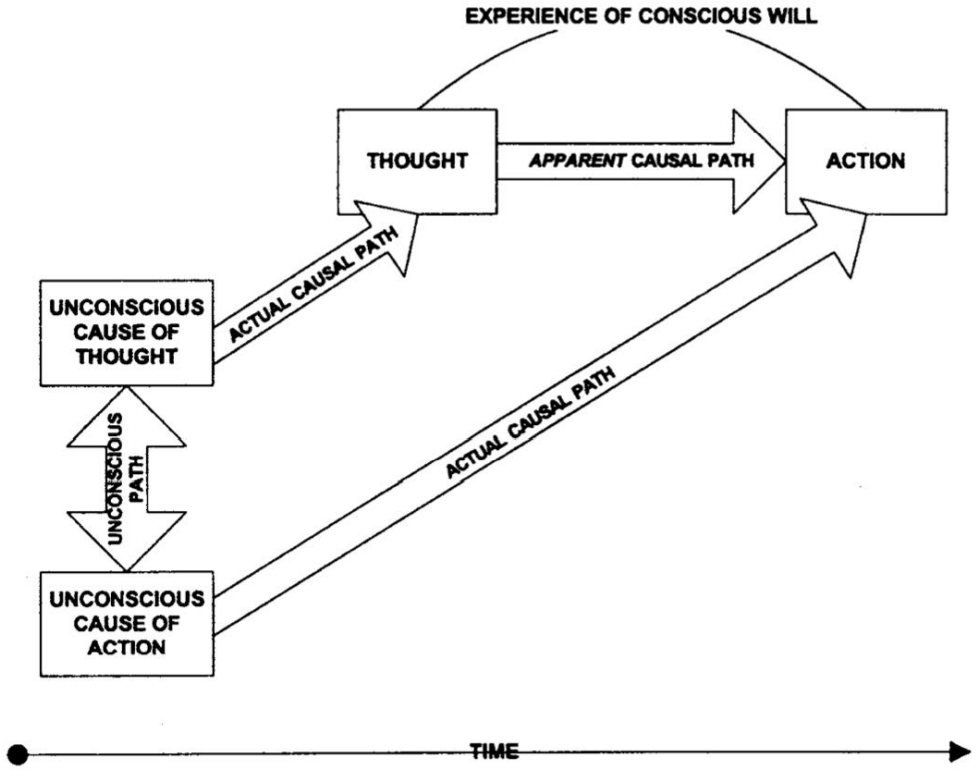
For the sake of this thesis, the reader does not need to be an expert in functional Magnetic Resonance Imaging (fMRI). However, he/she will need to be versed in the basics of the technology. FMRI, at its core, is a modified MRI machine, which is basically a giant magnet. These machines are sensitive to minute changes in magnetization. FMRI proceeds on the crucial insight made by Japanese researcher Seiji Ogawa that oxygenated blood is more highly magnetized than un-oxygenated blood. These changes in blood oxygenation are called Blood-Oxygen-Level Dependent (BOLD). Increased in BOLD levels are thought to correspond with increases in neural activity. So by mapping which areas show statistically significant BOLD signals, fMRI presumes to map the neural activity corresponding to the task at hand. However, it is not as easy as just scanning the brain and pointing to the statistically significant part. A number of factors can complicated the translation of BOLD signals, and researchers must apply sophisticated models in order to mitigate random signal noise, false-positives, and normal differences in BOLD levels, such as those between a large capillary and a small one. There are too many different models to describe, but what it is important is that they also do not image the whole brain at one time. FMRI proceeds in voxels, which are three-dimensional rectangular pixels. These can vary in size, but are not larger than a few millimeters. FMRI research usually combines many images to get a map of the whole brain. While "structural" imaging like an MRI shows the physical shape of the brain (i.e. an MRI used to investigate whether there are tumors), a "functional" imaging took like fMRI shows the brain in-action, performing tasks (i.e. this is what your brain looks like on drugs). This thesis refers to CAT scans, which are structural and employ differently angled x-rays, and also fMRI and PET scans which are functional, which analyze blood-oxygenation and glucose metabolism, respectively.



From Left to Right: An fMRI scan (Ohikuare, 2014); a CAT scan (Pulse Medical Imaging, 2016); a PET scan (Raine & Satel, 2013).

Appendix B:

Figure 1
A Model of Conscious Will



Note. Will is experienced to the degree that an apparent causal path is inferred from thought to action.

Excerpted From: (Wegner & Wheatley, 1999:483)