Responsible Brains

Responsible Brains

Neuroscience, Law, and Human Culpability

William Hirstein, Katrina L. Sifferd, and Tyler K. Fagan

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in ITC Stone Serif Std by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Hirstein, William, author. | Sifferd, Katrina, author. | Fagan, Tyler, author.
Title: Responsible brains : neuroscience, law, and human culpability / William Hirstein, Katrina L. Sifferd, and Tyler K. Fagan.
Description: Cambridge, MA : MIT Press, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2018008742 | ISBN 9780262038782 (hardcover : alk. paper)
Subjects: LCSH: Responsibility. | Brain.
Classification: LCC BJ1451 .H57 2018 | DDC 153--dc23 LC record available at https://lccn.loc.gov/2018008742

10 9 8 7 6 5 4 3 2 1

Contents

Preface

When we praise, blame, reward, or punish a person for doing something, it is usually because we think that person is *responsible* for that action—and hence that the person is, in some significant sense, the author of his or her actions and thus deserves to be held accountable for them, as well as their consequences. Making these assignments of responsibility is utterly central to human life, yet exactly how to justify them remains an open question.

Common sense tells us that what makes human beings responsible has something to do with their minds, and with the relations between their minds and their actions. And we now have good reason to think that the brain accomplishes the functions of the mind, thanks to an era of rapid progress in neuroscience which began in the 1980s and has continued to gather momentum. This means that the issues surrounding responsibility can now be approached from a relatively secure empirical standpoint, from a branch of biology—neurobiology. Given this, if a defendant, on trial for a horrible murder, were found to have serious brain damage, which brain parts or processes would have to be damaged for this defendant to be considered less responsible, or not responsible at all? Why do certain mental illnesses, with their corresponding profiles of brain dysfunction, seem to justify pleas of legal excuse? What is it about the developing brains of children and adolescents that makes them less culpable than adults, if not excused from responsibility altogether?

If a person's mind is damaged or still developing, or if the person's actions are fundamentally uncoupled from the person's mental processes, we may judge the person less responsible for what he or she does, or perhaps excuse the person from responsibility altogether. Indeed, the idea of a "guilty mind"—mens rea—lies at the core of the criminal law, a stable

and codified system of practices about whom to hold responsible, and for what. It seems obvious that the ability to reason has a great deal to do with being responsible, but exactly what sort of thinking or reasoning abilities are relevant in this way, and which brain processes accomplish them? What other brain functions are required for responsibility? Do all of these functions together constitute a meaningful functional unit in the brain? Could one actually point to a place in the brain (or a network of such places) as the seat of human responsibility?

We believe that the answer to these last two questions is yes. In an attempt to answer the question of where the mechanisms of responsibility reside in the human brain, we have constructed a comprehensive hypothesis about human responsibility couched ultimately in the language of neuroscience. In this book, we describe this theory, then put it to work addressing the above questions, comparing it with competing theories at each point.

This book is written to be accessible to virtually anyone interested in the above questions. Lawyers and other legal professionals should find our commonsense approach to responsibility congenial and practical. Medical professionals who make determinations of competency may also find this book useful. Within the academic community, the book should especially appeal to philosophers working on issues surrounding the concept of responsibility. More broadly, teachers or students in any of the cognitive sciences, including philosophy, psychology, linguistics, neuroscience, and artificial intelligence, who are interested in how their fields connect to issues in ethics may find this book worthwhile.

Acknowledgments

This book was made possible through the support of a grant from the John Templeton Foundation, although the opinions expressed here are our own and do not necessarily reflect the views of the John Templeton Foundation. We owe a great debt to Al Mele, the coordinator of the Philosophy and Science of Self-Control project, and manager of our sub-grant. His project not only funded our research for this book, but also provided venues at which our work could be presented.

We are also grateful to the many philosophers and scientists whose valuable input and feedback helped to shape this book—especially Neil Levy, who provided detailed comments on the chapters that analyze his theory of responsibility. We are deeply thankful to Philip Laughlin at the MIT Press for his guidance and open-mindedness during the book proposal process. We thank Elmhurst College for its support during this project, and we especially thank the students who participated in the *Responsible Brains* research group at Elmhurst College, including Daniel Hayes, Elana Hunt, Sean McKay, Kit Rempala, Margaret Sumney, Brandon Thompson, and Michael White.

Finally, we are grateful to our families for their love, support, and patience—with special thanks to Elsa, Harrison, Joanna, and Sally.

1 Introduction

Case Descriptions

Dominic Ongwen

It is the year 1990. A skinny boy named Dominic, perhaps ten or twelve years old, is abducted on his way to school in the Gulu District of northern Uganda. The men who have taken him are soldiers in the Lord's Resistance Army (LRA), a cultlike guerilla force led by the self-proclaimed prophet Joseph Kony. Dominic's abductors literally carry him to the LRA base; he is too small to keep up with them on foot.[1]

The son of schoolteachers, Dominic is now made to live with a middle-aged LRA commander and to think of this man as a military superior, mentor, and surrogate father. He is told to forget his parents and former life, and taught to empty his mind so it can be filled with passionate devotion to the LRA's holy cause. His indoctrination includes disorienting, exhausting periods of marching and hard labor punctuated by rituals of spiritual purification. In his daily intake of political and religious propaganda, he is told of Kony's supernatural powers, including the ability to detect any sign of disloyalty, or even private misgivings. Punishment is harsh and capricious, perversely fostering blind trust in the group's leaders and paranoia about nearly everyone else. Dominic witnesses murder, enslavement, and rape used as tools of war and discipline and is himself subjected to regular beatings, sometimes for minor infractions and sometimes for no reason at all. He sees other recruits and captives beaten, mutilated, and killed.

Over the next few years, Dominic's part in this milieu of violence shifts gradually from that of a passive witness to an eager accomplice and then to

a primary enforcer. His military training intensifies, and he quickly shows himself to be a natural warrior—brave, intelligent, keen both to please his superiors and to assume his own command. According to charges filed in the International Criminal Court, he begins to plan and execute his own raids and abductions and engages in the same forms of indoctrination used on his younger self. The LRA measures a soldier's loyalty by how many people he has killed or kidnapped, and in this Dominic proves exceptional. He is also lucky: While many of his peers and superior officers are killed, Dominic survives to adulthood. By 18 he is a field commander, and by 25 he has secured a senior rank in the LRA's high command, his promotion allegedly buttressed by an unknown number of brutal attacks, massacres, and abductions.

In 2005, the International Criminal Court (ICC) charged Dominic Ongwen with crimes against humanity, including the murder and enslavement of civilians. He surrendered to authorities in 2015 and is, as of this writing, on trial at the ICC. In the time between the first charges and his surrender, prosecutors have expanded the list of his charges to include over 70 criminal counts. Because he is known to have suffered many of the very crimes he is now accused of committing, Ongwen's case is unusual—he is the first figure of his kind to be subject to international criminal prosecution—but he is not unique in illustrating difficult questions about the culpability of child soldiers, many of whom appear to be both victims and perpetrators.[2]

Reviewing the evidence against Ongwen, it seems likely he has been involved in dozens, perhaps hundreds of abductions and murders. But how many of these crimes is he truly responsible for? The brutalities enumerated in his arrest warrants occurred during his early 20s, although he had surely caused many deaths before that time—and it is hard to see any of those deeds as disconnected from Ongwen's childhood abduction and forcible recruitment into the LRA. Pinning down the precise details of Ongwen's biography has proven difficult, and it is unclear whether it matters exactly how old he was at the time of his kidnapping, or how old he was when he began to commit the crimes he is charged with. We know that many child soldiers, rather than participate actively in atrocities, try to "fade into the background," risk escape, or even choose death; does it matter that Ongwen seems instead to have embraced the goals and tactics of the LRA?

Anders Breivik

It is midday on July 22, 2011, and the final pieces of Anders Breivik's plan are clicking into place. By nightfall he will be in police custody, and 77 innocent persons will be dead by his hand.[3]

At around 2:00 in the afternoon, Breivik sends out a mass e-mail containing a manifesto that sprawls over 1,500 pages. In it, he advocates violence against Muslims and rails against the multiculturalism and "cultural Marxism" he sees as having infected European society. In the manifesto, and in subsequent interrogations, Breivik claims to be acting on behalf of a secret Christian military order, the Knights Templar, with dozens or hundreds of members throughout Western Europe. There is no evidence, however, that any such organization exists.

Shortly after sending the e-mail, the 32-year-old Breivik leaves his mother's flat in west Oslo, driving a white van into the government quarter. The van contains a bomb made of fertilizer and fuel oil, weighing over a ton, which he has spent the past months assembling. He is dressed in a homemade police uniform. He carries a pistol and a semiautomatic rifle, both acquired over years of planned, careful maneuvering through the legal conditions necessary for gun ownership in Norway. Breivik parks the van, lights a seven-minute fuse, and walks away. He gets into a different car, parked nearby the night before as part of his plot, and drives out of the city. At 3:25 p.m., the bomb explodes, killing eight people and injuring many more.

An hour and a half later, some 25 miles northwest of Oslo, Breivik boards a ferry to the island of Utøya, where roughly 600 youths are attending a summer camp organized by Norway's Labor Party. The island is isolated and tiny, only a little larger than Rockefeller Center in New York City. Minutes after arriving on Utøya, and appearing trustworthy in his false uniform, Breivik draws his weapons and begins killing people. Over the next hour he fires hundreds of rounds. He lines up campers against the wall of a cabin and executes them. He shoots teenagers as they attempt to swim away from the island. He discovers young men and women feigning death, by lying on the ground among the slain, and shoots them in the head. Survivors will later report that Breivik persuaded those hidden to come forward by saying he was a policeman, and that he was laughing and shouting as he killed them. After 50 minutes, Breivik calls the police saying: "Yes, hello, my name is Commander Anders Behring Breivik from the Norwegian anti-communist

resistance movement. I'm on Utøya for the moment. I want to give myself up" (Melle 2013). After this call he continues shooting. In his first interrogation after being taken into custody, Breivik again claims to be commander of the Knights Templars Norway and states that "the people on the island were category C traitors" (Melle 2013). What happened on the island was unfortunate, he tells authorities, because he had the right to "kill category A and B traitors, but not a mandate to kill category C traitors" (Melle 2013). Between the Oslo bombing and the Utøya murders, Breivik kills 77 people that day, 55 of them teenagers.

Norway's legal system employs what is sometimes called the "medical model" of legal insanity: If a defendant is judged to have been psychotic at the time he or she committed a criminal act, then that defendant is to be excused from punishment (Moore 2014). Such judgments depend heavily, if not entirely, on psychiatric testimony. Breivik's initial review, conducted by a team of psychiatrists, found that he was psychotic—and thus legally insane—because he was suffering from paranoid schizophrenia. Many of Breivik's outlandish beliefs were construed, through the lens of this diagnosis, as "persistent, systematized, bizarre delusions" (Bortolotti, Broome, and Mameli 2014). But a second review, conducted by a different team of psychiatrists, found that Breivik suffered from antisocial and narcissistic personality disorders but *not* paranoid schizophrenia and was therefore not psychotic. The panel of judges hearing Breivik's case found this second review more persuasive, and Breivik was judged legally sane and sentenced to 21 years in prison, a term that can in principle be renewed until Breivik's death.

From the calculated horror of Breivik's crimes to the bizarre and delusional beliefs he expressed before and after the attacks, to the conflicting psychiatric diagnoses and Norwegian law's unusual deference to medical expertise in determining legal insanity—the facts of Breivik's case raise a number of troubling and thorny questions. We may find ourselves pulled in contrary directions, feeling on the one hand that any person who planned and executed these crimes deserves the harshest punishment possible under the law; and on the other hand that it would be unjust to punish someone who does not understand or cannot control his behavior and is therefore undeserving of blame. What was the state of Anders Breivik's sanity at the time of his crime, and how and why does the answer to that question matter? If Breivik was truly suffering from paranoid schizophrenia

in 2011, does that mean he was not morally responsible for his actions and is therefore undeserving of legal punishment? What is the relevance of mental illness in general, or a specific mental illness in particular, to responsibility? How much evidentiary weight, if any, should a psychiatric diagnosis carry in deciding whether a defendant can properly be said to deserve punishment for his or her crime?

Brian Dugan

It is February 1983, and Brian Dugan is cruising aimlessly around the western Chicago suburbs.[4] Now 26 years old, he has spent the last ten years in and out of prison and the legal system, charged with an array of increasingly violent offenses, from burglary and battery to arson and attempted kidnapping. Jeanine Nicarico is alone at her house in Naperville, Illinois, home from school with the flu. She is ten years old. Dugan has been knocking on doors in the neighborhood, half-randomly, hoping to find an unoccupied house where he can break in and steal some tools. He knocks on Jeanine's door, and although she refuses to let him in, her way of answering Dugan's questions leaves no doubt that she is home alone. Dugan breaks in the front door, abducts Jeanine, rapes her, and beats her to death with a tire iron. Jeanine's body is found two days afterward. The shock of this heinous crime, committed in broad daylight on a quiet suburban street, convulses authorities into a frantic—and arguably incautious—search for the perpetrator. Pretending to have knowledge of the case in a misguided attempt to claim the $10,000 reward, a young man named Rolando Cruz winds up attracting police suspicion and finds himself arrested and charged with Jeanine's murder, along with a man named Alejandro Hernandez. With suspects in custody, Brian Dugan is free to resume his hunting.

In July of 1984, Dugan sideswipes a young woman's car, running her off the road. She is Donna Schnorr, a 27-year-old nurse, and a stranger to Dugan. He forces her into his car, binds her, and drives her to an abandoned quarry a few miles away, where he rapes and drowns her. Less than a year after that, in May of 1985, Dugan—not yet a suspect in the Schnorr killing—commits a string of intermittently successful abductions and rapes, the victims of which are all young women or girls. On June 2, Dugan's spree reaches its final phase; he kidnaps seven-year-old Melissa Ackerman, rapes her, and drowns her. A few weeks later her body is found, and Dugan is

arrested shortly afterward; another young girl had been with Melissa at the time of the abduction, but had narrowly escaped, and she had been able to describe Dugan's car to police.

Under questioning, Dugan confesses to the Ackerman and Schnorr murders, each of which earns him a life sentence. He informally confesses to the Nicarico murder as well, but will only offer a formal confession if he can avoid the death penalty; prosecutors reject his offer. Eventually, DNA evidence is developed linking Dugan to Jeanine Nicarico's death; he is indicted for her murder in 2005 and pleads guilty in 2009, hoping to avoid the death penalty. As the sentencing phase of his trial begins, Dugan's attorneys are casting about for any evidence that might help mitigate their client's guilt in the eyes of the jury and keep him off death row. They learn of a cognitive neuroscientist in New Mexico named Kent Kiehl, who is an expert on psychopaths. Kiehl has spent years collecting data on the neuropsychology of violent offenders. He claims that the brains of psychopaths show distinctive patterns of dysfunction and decreased activity, particularly in areas responsible for integrating emotion with cognition, which explain why psychopaths often appear to be devoid of empathy, remorse, or sensitivity to the harms they cause. Dugan's lawyers reach out to Kiehl, offering him the chance to study, in Kiehl's words, "one of the classic psychopaths in American history" (Hughes 2010, 342).

Kiehl accepts the offer and sets to work, gaining access to Dugan's case files and conducting lengthy interviews with him. He scans Dugan's brain using functional magnetic resonance imaging (fMRI) and tests him on the Hare Psychopathy Checklist, an interview technique meant to detect the presence of psychopathy through 20 distinct personality and behavioral traits. Dugan scores a 38 out of 40 on the test, placing him, in Kiehl's estimation, above the 99th percentile of all prison inmates. Dugan's files reveal a childhood that bears the typical marks of a budding psychopath: cruelty to animals, chronic bed-wetting, and fascination with fire.

In October of 2009, Kiehl participates in a Frye hearing, a legal proceeding meant to assess whether some body of scientific evidence can be admitted at trial. The prosecution argues that allowing Kiehl to testify using the actual brain scans could unduly bias the jury, while the defense argues that the scans are necessary for a complete and accurate picture of Dugan's mental state. The judge finds a middle ground of sorts, allowing Kiehl to share his findings with the jury but not to use the actual brain scans obtained

from Dugan. Kiehl takes the stand in November; his testimony is used to ground the defense's argument that Dugan, as a psychopath, was not fully in control of himself at the time of his crimes and was not capable of suppressing or managing his powerful impulses to harm, violate, and kill. The horrific rape and murder of Jeanine Nicarico was perpetrated not by an evil man, but by a sick man, they argue, and his condition calls for a more lenient sentence, or at least something short of execution. The prosecution counters Kiehl's testimony with their own expert witnesses, who point out the evidential limitations of fMRI and argue that a brain scan taken 26 years after the crime in question can hardly illuminate Dugan's mental state at the time of Nicarico's murder. Even if Kiehl had ironclad evidence of Dugan's psychopathy—not just now but back in 1983—there should be serious doubts about whether psychopathy should eo ipso mitigate a defendant's blameworthiness in the eyes of the law. The jury in Dugan's case returns the unanimous verdict needed for the death penalty, and Brian Dugan is sentenced to die.

In March 2011, however, Illinois Governor Pat Quinn signed a bill abolishing the state's use of the death penalty and commuting the death sentences of 15 Illinois prisoners—Dugan among them. The death penalty had already been under a statewide moratorium since 2000, and Quinn's action was the culmination of concerns about wrongful convictions and executions of the innocent. In a bizarre twist, two of the most notorious wrongful convictions were those of Rolando Cruz and Alejandro Hernandez, who had both been convicted and sentenced to death for the rape and murder of Jeanine Nicarico. Had Brian Dugan not made his informal confession back in 1985, the two men might well have been executed. (Gutowski and Mills 2014).

Today Brian Dugan is 61 years old, serving life without the possibility of parole in the Stateville Correctional Center, less than 20 miles from the scene of his first murder. He appears to have evaded the death penalty for good, but he will die in prison. In his first published interview, in 2014, Dugan told reporters from the *Chicago Tribune* that he still felt he was a danger to society, prone to intense and uncontrollable rage. Describing the murders, he said, "I was driven by some kind of an impulse that kept growing. I could not stop" (Gutowski and Mills 2014).

The case of Brian Dugan presents a number of difficult and unsettling questions: How should we understand the moral responsibility, or lack

thereof, of the psychopath? Are psychopaths "more bad than mad," as Maibom (2008) argues, or is it the other way around? To what extent do psychopaths understand the wrongness or harmfulness of their actions? If they seem unmoved or unmotivated by moral norms, can they really be said to *understand* them? And what about volitional control, often described as a crucial component of moral responsibility—are psychopaths truly in control of their actions? What are the cognitive and neurological signatures of psychopathy? What sorts of evidence from psychiatry and neuroscience would indicate that a defendant was psychopathic, and should that same evidence tend to reduce, or excuse the defendant from, culpability? If there is such evidence to be had, how should it be responsibly introduced into the courtroom?

Responsible Brains

The case studies above involve people who caused serious harm, but to whom attributing responsibility for that harm seems particularly difficult. They are cases where pertinent-seeming facts about the agents' minds muddle the intuitions about agency and intent that underpin attributions of responsibility and, therefore, blame. As a result, courts have disagreed about the level of blame and punishment that ought to be assigned to these sorts of agents—for instance, child soldiers (or children, generally) and the mentally ill, including psychopaths.

In this book we will argue that specific facts about the brains of the agents discussed in these cases, now available due to recent progress in neuroscience, can strongly inform assessment of their culpability. This may not sound like a bold claim: It is not very controversial, after all, to say that facts about people's mental capacities and mental states may be relevant to their responsibility for harmful acts. And it is similarly commonplace to believe that facts about an agent's brain are relevant to facts about that agent's mind. Some scientists and philosophers would even prefer to say that facts about the mind simply are facts about the brain because mental states simply are a type of brain state.[5]

However, one needn't go this far to find persuasive the idea that learning about how the brain works can tell us about how the mind works, or that facts about a person's brain might bear on his or her responsibility. Agents who have a certain level of plaque and tangles in the brain—both

of which are highly correlated with Alzheimer's disease—can be expected to suffer from certain mental incapacities. The link between brain plaque and the symptoms of Alzheimer's is strong enough for facts about levels of plaque and tangles to be relevant in at least certain cases of attributing responsibility. For example, imagine Bob is accused of the theft of a briefcase. Bob claims his Alzheimer's disease led him to forget that the briefcase was not his. However, the victim of the theft claims Bob took the briefcase intentionally and is malingering (is just pretending he has Alzheimer's). If new brain imaging techniques allowed a medical specialist to provide good evidence of plaque buildup or tangles in Bob's brain, this would obviously be relevant to his excuse.

Evidence of traumatic brain injuries or other indications of structural brain damage, severe chemical imbalances, and extremely low activity in certain regions of the brain would all appear to be relevant in at least some cases where we are attempting to attribute responsibility to an agent. To take another example, imagine that Carl, normally a law-abiding family man, starts acting impulsively and then gropes a neighbor. He is discovered to have a brain tumor, and when it is removed, he once again becomes considerate and law-abiding. Eight months later, he is arrested for groping a sales assistant. Evidence that Carl's brain tumor has now returned would seem, prima facie, relevant to his responsibility for the assault.

We think the pertinent question is not *whether* brain science can inform responsibility assessments, but in which sorts of cases, and to what extent. If facts about the brain bear on facts about the mind, then neuropsychological findings must in at least some cases inform assessment of defendants' mental states, so long as we can link facts about brains to facts about the capacities necessary for responsible agency. All we need, it seems, is to determine which mental capacities are necessary to responsible agency, and which facts about brains are relevant to those capacities.

This line of thought is disquieting to many prominent philosophers interested in agency and responsibility.[6] Stephen Morse, for example, worries about the spread of "brain overclaim syndrome" (Morse 2006a) into the legal realm. Morse argues that some commentators allow their enthusiasm for the fast-moving world of neuroscience, with its seductive fMRI images and buzzy breakthroughs, to lead them astray. They apply neuropsychological findings incautiously and overestimate their true relevance to questions of criminal responsibility. Though he is not hostile to neuroscience

in general, Morse is decidedly skeptical about the evidential significance of neuroscientific results in criminal trials. In his survey of *Roper v. Simmons*, a case in which the U.S. Supreme Court eventually ruled that the death penalty for crimes committed while the defendant was under the age of 18 was unconstitutional, Morse discusses the neuroscientific findings adduced in amicus briefs to the Court:

> Assuming the validity of the neuroscientific evidence, what does it add? The rigorous behavioral studies already confirm the behavioral differences [between adolescents and adults].... At most, the neuroscientific evidence provides a partial causal explanation of why the observed behavioral differences exist and thus some further evidence of the validity of the behavioral differences. It is only of limited and indirect relevance to responsibility assessment, which is based on behavioral criteria. (Morse 2006a, 408–9)

Morse's worries about neuroscience's incursion into the criminal law mostly concern the prospect that overzealous legal scholars will wrongly inflate the significance of neuroscientific findings, or that neuroscience—which is after all a rather young scientific field—may get some things wrong altogether. That is, Morse worries that some are claiming the relevance of brain science to legal responsibility without building an appropriate bridge between the two types of facts. At times, he even seems to worry that such a bridge cannot be built because "the way the brain enables the mind continues to be a mystery" (Morse 2013b, 512).[7] We would caution against thinking of brain and mind as two separate but causally related entities, though, and suggest that a good portion of the mystery derives from this thinking. But luckily, we will argue, we need not enter into such metaphysical disputes in order to develop an effective and sound approach to responsibility grounded in science. Although Morse may be right to worry about certain aspects of the relationship between brain and mind—the mystery of how brain matter gives rise to conscious phenomenal experiences, for example—progress in neuroscience and cognitive science continues to inform our understanding of how the brain enables our cognitive capacities and functions. How a physical system can fulfill certain functions is not a great source of mystery, and we will argue (primarily in chapter 5) that our remarkable abilities to be responsible derive from these cognitive functions, rather than our equally remarkable ability to be conscious. Given this, if we are clear on the capacities that matter to responsible agency, we should be increasingly able to use neuroscience to understand

how brain function supports or undermines these capacities as science progresses.

Other scholars have staked out positions at the other end of the continuum from Morse and worry that neuroscience will get it right—all too right, one might say, since they see the relevance of neuroscience to questions of responsibility and the criminal law as not only direct but radically disruptive. Scholars harboring these concerns tend to see our everyday folk-psychological responsibility assessments as deeply committed to a libertarian conception of free will, which is threatened by the findings emerging from brain science.[8] If their view is accurate, then a complete neuroscientific understanding of human agency might well supplant our cherished commonsense notions of free will, agency, and responsibility, which ground our legal practices of blame and punishment. This may then lead to verdicts about responsibility that differ strongly from our commonsense notions of who should be held responsible. A prominent line of conjecture holds that neuroscientific data may require us to shift away from retributive (blame-based) punishment and toward a rationale for punishment that is exclusively concerned with deterrence and/or rehabilitation.[9]

Defenders of this sort of view ignore the proliferation of work articulating compatibilist theories of responsibility, however. A large and diverse group of scholars see the folk intuitions about agency and responsibility as compatible with physical determinism and believe that what libertarian commitments exist can be revised without much disruption to the structure of responsibility assessments.[10] That is, the folk-psychological concepts that underpin assessments of responsibility—where mental capacities and states are attributed to an agent to determine the level of praise or blame that constitutes an appropriate response to a particular action—do not seem to be undermined by the brute fact that these capacities and states are realized in the brain (which is a physical object subject to certain deterministic rules or laws).

For example, the most popular philosophical account of the capacities necessary for responsibility, the "reasons-responsiveness" account made famous by Fischer and Ravizza (1998), understands such capacities as compatible with determinism. On one version of a reasons account, a person is moderately responsive to a reason for action if he or she would have recognized that particular reason as relevant to his or her action in a "close possible world," which means a world we can imagine that is very like our

own except for particular details (this is called a counterfactual reasons account) (Vargas 2013). That is, under other similar circumstances the person would have recognized the relevance of the reason to his or her actions, possibly causing the person to inhibit the action altogether. For example, in the actual world, Mary doesn't remember to pick her child up from soccer because she is trying to finish work due by the end of the day. But if, in another possible world, her deadline were to be extended, she would have remembered. This shows she is appropriately sensitive to reasons to pick up her child, even if she forgets under certain conditions. On this sort of view, even claims that a person "ought to have been more careful" or acted differently would seem compatible with the truth of determinism. We will discuss philosophical theories of responsibility and free agency in more detail in chapter 3, but it will suffice to say here that the compatibilist theories advocated by many philosophers leave wide open the possibility that responsibility assessments may be informed by brain science without undermining attributions of praise and blame, or application of punishment, and in so doing depriving us of justified retributive punishment.

In this book we argue for what we see as a moderate position: Neuroscience is both relevant to responsibility and consistent with our ordinary "folk" conceptions of it. Evidence from cognitive science and neuroscience can illuminate and inform the nature of responsibility and agency in specific, testable ways. We are not alone in this view. Neuroscientist and philosopher Adina Roskies, for instance, has conjectured that "neuroscience might enable us to develop a more sophisticated view of responsibility that takes into account both the cognitive demands and the control demands made by intuitive and legal notions of responsibility, and reconciles them with a scientifically informed view of the brain as a physical system that governs our actions" (Roskies 2006, 423).

The majority of our argument in this book constitutes our effort to take up this challenge. Of course, there are existing accounts, especially from philosophers of law, examining the psychological capacities that ground criminal law verdicts (Duff 2004, Moore 1997, Morse 2006b). Our theory's novelty lies in its effort to inform what exactly these capacities are and how they can make us responsible with findings from neuroscience, many of which are recent. The criminal law, as a high-stakes repository of folk-psychological judgments about responsibility, carries an implicit commitment to a view of human agency as basically free and reason-based.

The image of human agency given by contemporary brain science seems irreconcilably different: that of beings whose actions are governed by the mechanistic churn of an immensely complex physical system. Our goal is to show that these two conceptions of human agency are not in fact incompatible. We will offer, in Roskies's terms, a sophisticated view of responsibility that can both serve our folk and legal purposes and be reconciled with and reinforced by a mature scientific understanding of the brain. Hence, our title: *Responsible Brains*.

The Road Ahead

In the chapters to come, we will argue that folk conceptions of responsibility, which underpin and are reflected in the structure of criminal offenses and verdicts, implicitly refer to a particular set of cognitive and volitional capacities. These capacities are implemented by brain structures primarily—though not exclusively—belonging to the cognitive control network and are known to the neuropsychological field as *executive functions*.[11] These functions are what allow us to be agents with reasons, plans, and values and to coordinate our behavior accordingly as we move through a complex world. Executive functions—such as attentional control, planning, inhibition, and task switching—are therefore uniquely well suited to ground a reasons-responsiveness account of the capacities necessary for moral responsibility, including both sensitivity to morally or legally relevant reasons and the volitional control to act in accordance with those reasons.

When we analyze paradigmatic cases of legal responsibility, as well as cases "from the margins" of excuse, as Shoemaker (2015) labels them, we find that legal responsibility is contingent upon a person's having the capacity for some baseline level of executive function, either at the time a crime is committed or for some significant period of time before the crime. We claim the law assumes that citizens possess some baseline executive capacity, and that cases of legal responsibility are cases in which an offender either exercised executive functions with regard to the criminal act or could have done so, in the counterfactual sense described above (unless some justification applies). We will argue, further, that cases of legal excuse tend to be cases where an offender had severely compromised executive functions—because, for instance, he or she suffered from schizophrenia or

was too young at the time of the offense to have had a minimally mature set of executive capacities.

Let us map out the road ahead. The early chapters of this book present our executive theory of responsibility, along with its philosophical and scientific foundations. Chapter 2 will give a detailed review of the current neuroscientific research on executive functions—what they are, how they are realized in the brain, and how they are measured and assessed. We will discuss the relationship between the various executive functions, their organization, the extent to which they are separable, and their relevance to reasons-responsiveness. Chapter 3 will motivate our position and place it within the broader philosophical landscape of theories about free agency and moral responsibility. Chapter 4 describes how our theory meshes with legal theory and the criminal law itself.

In chapters 5 through 7, we attempt to further support our theory by arguing its merits compared to those of a prominent competitor: the intriguing consciousness thesis advocated by Neil Levy (2014), according to which consciousness is crucial for responsible agency. We have chosen to discuss Levy's theory as much for its strengths as for what we see as its flaws. His view is nuanced, closely argued, and informed by attention to findings in the cognitive sciences. And indeed, we think Levy is barking up the right tree, because consciousness and executive function are strongly related phenomena. But they are different phenomena. We will argue that it is actually executive functions, which are anatomically and functionally separate from the process of consciousness itself, that do the important work of enabling reasons-responsiveness and, therefore, responsibility.

Having articulated the core of our theory and argued for its merits in comparison to a prominent rival view, we turn in the later chapters to extending and applying that theory. Chapter 6 continues what we see as a fruitful comparison between our theory and Levy's while fleshing out our claim that our theory is folk-consistent, by explicating how it would assign specific truth values to everyday folk claims about responsibility. In chapter 7, we describe two cases in detail—one man who neglected to watch his children over the weekend, and another with a rare sleep disorder who attacked his in-laws in the middle of the night—explaining how our theory's treatment of them is preferable to Levy's. Chapter 8 takes up the puzzle of juvenile responsibility—how, on an executive account of responsibility, the capacities for responsible agency take root and develop in the

maturing brain. The gradual maturation of executive functions calls for us to think of juvenile responsibility in scalar terms, with consequent implications for reforming the machinery of juvenile justice. Chapter 9 explores the ways that mental disorders or diseases may undermine responsible agency, applying the executive theory of responsibility to the doctrine of legal insanity and the question of whether psychopaths are culpable for the harms they cause. Chapter 10 discusses the implications of our view for thinking about criminal punishment. If we take seriously the notion that executive function is the key to responsibility, we may need to revise our sense of why punishment is justified in the first place, and how it should be applied to best serve the proper functions of criminal justice. Chapter 11 will briefly recapitulate our main claims, revisit the cases described in the first part of this chapter, and outline some future avenues of inquiry and practical application suggested by our theory.