Introduction

Felipe De Brigard and Walter Sinnott-Armstrong

Philosophers and neuroscientists try to answer many of the same questions: How can consciousness emerge from brain activity? How are perception, memory, and imagination related to what we know? How do brain development and activity influence moral choice, behavior, and character? Are specific decisions based on general intentions? Is brain activity determined or stochastic? Do we control our choices and actions? And so on.

Each field can learn a lot from the other about these and other topics of mutual interest. Unfortunately, researchers in these disciplines rarely work together or even understand each other. Fortunately, recent developments signal the coming end of this misfortune. New methods as well as theories in both neuroscience and philosophy increase openness on both sides and pave the way to mutually beneficial collaborations. Nonetheless, significant impediments remain. We need to understand both the promise and the obstacles in order to make progress.

The Growth of Neuroscience

The past few decades have witnessed an exponential increase in neuroscience research. Whether this growth is due to its brilliant investigators from various disciplines, to its multiple synergistic research methods, or to its numerous promises of profound results, neuroscience is now the flagship among biological and medical sciences.

The White House confirmed this prominent position in April 2013 when it unveiled the Brain Research through Advancing Innovative Neurotechnologies—or BRAIN—Initiative (https://obamawhitehouse.archives.gov/BRAIN). This Initiative hopes to enable us not only to understand better the relationship between the mind and the brain but also to "foster science and innovation" and "to inspire the next generation of scientists"—two of the so-called Grand Challenges of the twenty-first century.

Encouraging as this plan might sound, the truth is that we are far from possessing a detailed picture of how the brain works and even farther from understanding how brain functions relate to operations of the human mind. This gap is no secret, as neuroscientists will be the first to admit. Part of the problem lies in the sheer difficulty of developing noninvasive technologies that are powerful and sensitive enough to allow us to measure brain activity in vivo and at different levels of complexity, from molecules to large-scale brain areas. Similarly, neuroscientists face the daunting task of systematizing, analyzing, and modeling enormous amounts of data. Above and beyond these fundamental yet tractable practical difficulties rests a less discussed but equally pivotal challenge: that of understanding how answers about the brain can resolve questions about the mind and vice versa. No matter how much neural data we amass and no matter how sophisticated or computationally powerful our statistical models become, the goal of endeavors such as the BRAIN Initiative will be out of reach unless we have a clear theoretical framework that reveals how the brain relates to the mind. This relation is the bailiwick of philosophers.

A Role for Philosophy

Since Plato's Phaedo, philosophers have repeatedly asked how our minds are related to our bodies. This perennial problem has been reshaped in the context of twenty-first-century neuroscience into the question of how our brains can cause or realize our minds. Many profound and original answers have been proposed by philosophers in recent years, including not only subtle variations on reductionism but also forceful arguments against reductionism, as well as several sophisticated alternatives to reductionism that take neuroscience into account. These philosophical theories could potentially provide the kind of framework that the BRAIN Initiative needs. Thus, neuroscience can gain many benefits from working with philosophers.

Philosophers also stand to gain insights from neuroscience. Any approach to the philosophy of mind needs to be consistent with what we know about how the brain works, but neuroscience also affects many other central issues in philosophy. Epistemologists try to develop theories of knowledge that need to take into account how information is processed in human brains, including the neural systems that enable perception, memory, and reasoning. Philosophers worry about free will and responsibility, and their views can be tested against discoveries in the neuroscience of decision, action, self-control, and mental illness. Again, moral philosophy—including practical ethics, normative theory, and meta-ethics—often hinges on empirical assumptions that can be tested by neuroscience together with psychology. Are moral beliefs based on emotion or reasoning? Do people have moral virtues and vices that remain stable across situations? Are moral judgments unified by a shared substrate in the brain? These questions (and many more) are asked by philosophers, and their answers depend on empirical claims that can and should be tested scientifically. In these ways, philosophers can benefit tremendously by working with neuroscientists.

Barriers to Collaboration

At this juncture, philosophy and neuroscience are poised to work together, learn together, and advance together. However, just as philosophers have come to appreciate the results coming from neuroscience, and just as neuroscientists have come to appreciate the importance of asking foundational questions to grasp the significance of their results better, both fields are keenly aware of the substantial gaps between their vocabularies, research methods, and cultures. These gaps stand in the way of progress.

One barrier is language. Each discipline creates and sometimes insists on its own proprietary vocabulary. Yet, interdisciplinary collaboration requires mutual understanding of each other's technical terms. Disciplines also develop distinctive methodologies, standards, and practices, and they cannot work together without becoming familiar with these central features of their collaborators' fields. One aim of each chapter in this book is to help readers from both fields become familiar and comfortable with each other's vocabularies, methods, standards, and practices.

Successful collaboration also requires mutual respect, which comes from appreciating the value of the other field. Real appreciation, however, cannot come from a single lecture or conference. It takes time. That is why each of the chapters in this book was written by a team that includes at least one philosopher and at least one neuroscientist. These teams had worked together before on research projects and then spent many months writing a chapter together. The authors of each chapter also met for an intense week in which they helped each other to improve their chapters and make them more intelligible to readers who were trained in different disciplines or worked on different topics. This lengthy and intense process of collaboration helped them appreciate each other other's disciplines and led them to respect each other. This mutual appreciation is clearly evident in their chapters. They provide tangible proof that interdisciplinary collaboration between neuroscientists and philosophers is feasible and fruitful (as well as fun!).

The teams of authors of these chapters were drawn from the fellows in Summer Seminars in Neuroscience and Philosophy (SSNAP), which we codirected in 2016–2018 with generous support from the John Templeton Foundation (for which we are very grateful). Fellows in these seminars studied each other's fields intensely for two weeks, and then most of them collaborated in research projects over the next year or two. Competition for slots in these seminars was fierce, and these authors represent some of the best fellows in that elite group. They are all future leaders in neuroscience and philosophy who will shape interactions between these disciplines for years to come. If you want to know how this interdisciplinary field will develop, they will show and tell you. That is another goal of the chapters in this book.

In addition, this book displays the wide variety of topics that fall under the general heading of neuroscience and philosophy. Although each chapter combines both disciplines, they use them in different ways to make progress on different issues.

Part I: Social Neuroscience and Philosophy

The first group of chapters addresses basic issues about our social and moral lives: how we decide to act and ought to act toward each other, how we understand each other's mental states and selves, and how we deal with pressing social problems regarding crime and mental or brain health.

Moral Judgment

Since ancient Greece, philosophers have debated about which acts are morally right or wrong. Moral controversies have also polarized whole societies, not only recently but throughout history and around the world. Neuroscience by itself cannot solve these problems, but it might be able to help us understand what makes people hold strong but contrary views about many important moral issues. Some skeptics have suggested that knowledge of the neural processes that produce moral disagreements should make us doubt that any moral belief is true or known to be true.

In their chapter, **May, Workman, Haas, and Han** survey what neuroscience has learned so far about these issues. They divide neuroscientific studies of morality into three periods: the first period from 1990 to 2000 emphasized gut feelings, the second period from 2000 to 2010 introduced dual process models using brain imaging and manipulations, and the third period from 2010 to 2020 added computational models and theories of how adults and children learn their moral views. In the end, they argue that neuroscience does not undermine all moral judgment but instead can help us to determine when moral judgment is trustworthy or not.

Empathy

Moral judgments often concern harm to other people. As a result, many philosophers see empathy as central to moral judgment. However, morality is also often said to be impartial, and several critics have argued that empathy can interfere with impartial moral judgments when we empathize only with individuals or groups whom we favor. These debates have been clouded by imprecision and ambiguity, since empathy can have many different meanings.

In their chapter, **Spaulding, Svetlova, and Read** try to untie these Gordian knots by distinguishing automatic versus conscious empathic reactions, cognitive versus affective and motivational aspects of empathy, and self-oriented versus other-oriented concerns. They trace the development through childhood of self-awareness and other-awareness as well as concern for others, sympathy, and mentalizing (theory of mind). Contrary to some recent claims, they argue that these abilities cannot be understood solely in terms of mirror neuron activations, for various kinds of empathy also involve the right temporal parietal junction, precuneus, and superior temporal sulcus, among other brain regions. In the end, they bring this neuroscientific and philosophical theorizing to bear on the issues of how (and whether) empathy is related to morality.

Agency

In addition to substantive moral judgments about actions, we also judge some people to be morally responsible for some actions but not for others.

These judgments depend on various aspects of our actions, such as who does the act or causes its effects, whether the agent knows or could know what was being done, which factors affect what the agent does, and whether the agent controls the act or its consequences. These issues have been investigated in depth by decision neuroscientists, and their findings have far-reaching implications for philosophy of action as well as philosophical theories of free will and moral responsibility.

In their chapter, **Waller and Brager** summarize recent findings about two main issues: the sense of agency (the feeling that I did an action) and the initiation of physical movements (which bears on control of whether and when an agent acts). Then, they go into more detail about the previously unexplored issue of how sleep deprivation affects the brain processes that underlie the sense of agency and the initiation of action. This case study raises questions about whether we should excuse mistakes by agents who are required to make important decisions without adequate sleep—including military personnel, shift workers, college students, and parents of young children. These questions cannot be answered without empirical information from neuroscientists together with normative reflection from philosophers.

Self

We make moral judgments not only about what we and others do in the present, but also about what we and others did in the past—sometimes the distant past. Therefore, we need to ask whether the agent is the same person today as they were when they did the act in the past. If not, the person at present does not seem responsible for what the other person did in the past. Philosophers have addressed this issue of personal identity over time with elaborate thought experiments, but they have rarely used empirical methods—until recently.

In their chapter, **Everett, Skorburg, and Livingston** review psychological studies of how common folk identify people over time when those people have undergone major life changes, such as those that result from severe brain damage. They focus on the perhaps surprising finding that changes in moral characteristics, especially in a negative direction, affect folk judgments about personal identity more than other changes, including memory loss and disruptions in non-moral personality traits. This "moral self" effect has been replicated in many groups, including real families of people with brain damage, and its neural basis is starting to be studied. It raises profound challenges to traditional philosophical theories of moral responsibility as well as personal identity, and it suggests that our notion of personal identity over time is not a descriptive matter of metaphysics alone.

Mental Illness

Mental illnesses provide more striking examples of when people might say that an agent was not herself when she did something wrong. Her illness did it, not her—but when is she "not herself"? That question cannot be answered properly without a deeper understanding of mental illness. Mental illnesses are, of course, related to brain dysfunction, and neuroscientists have made significant progress in understanding and treating several mental illnesses. Neuroscience alone cannot determine when a brain abnormality is a mental illness, disease, disorder, or dysfunction, but it can discover neural changes that lead to the feelings, thoughts, and behaviors that usually constitute mental illnesses.

In their chapter, **Washington, Leone, and Niemi** address these and other issues with respect to obsessive-compulsive disorder (OCD), schizophrenia, and addiction, focusing on deep brain stimulation (DBS) as a treatment for OCD and heterogeneity within schizophrenia and addiction as illuminating for what (if anything) unifies a mental illness and for how brain states are related to mental states. The authors discuss what neuroscience has found out about these three examples, but then they go on to show how neurobiological models of these and other mental illnesses have affected philosophical theories, scientific methods, medical treatments, government policies, court procedures, and public opinions in ways that have been controversial and problematic. These applications illustrate some ways in which neuroscience can be misused and can raise important moral issues.

Crime

People with certain mental disorders are sometimes confined against their will because they are dangerous to themselves and others. Such "involuntary commitment" is based on a prediction of future behavior. Our legal system also incarcerates criminals for longer or shorter periods based on predictions of future crimes when defendants are granted or denied bail, when those found guilty are sentenced to punishments, and when inmates do or do not receive parole. These predictions are often criticized as inaccurate, especially when they are based on subjective assessments by judges or forensic clinicians. Actuarial statistics can be more objective and accurate, but these run into new criticisms because they do not treat people as individuals and often use classifications that are indirect proxies for race and socioeconomic disadvantage. Some neuroscientists have recently attempted to use brain scans to improve the accuracy and fairness of these predictions, but neuroprediction remains controversial.

In their chapter, **Aharoni, Abdulla, Allen, and Nadelhoffer** explain the methods that statisticians and neuroscientists use to predict crime, critically assess some limited successes using these tools, and suggest directions for improving both legal procedures and statistical and neuroscientific methods. They respond to criticisms by neuroscientists as well as ethicists and lawyers, but they conclude optimistically that neuroprediction has the potential to become more accurate and at least as fair as the methods of prediction that courts use today.

Optogenetics

In addition to prediction in law, neuroscience can also be used for treatment in medicine. Pharmacological interventions on brain function have been around for a while and have helped to ease some mental illnesses, but new techniques are quickly coming online. Among the most striking are electrical DBS, transcranial direct current stimulation, gene therapy, and now optogenetic DBS. Each new treatment raises important new ethical issues.

In their chapter, **Felsen and Blumenthal-Barby** explain the basic neurobiology behind optogenetics—how light can affect brain function—along with the potential advantages of this technique over other methods of treatments for various conditions, including retinitis pigmentosa, Parkinson's disease, depression, and OCD. Next, they address ethical concerns involving how to select subjects fairly, how to weigh risks against benefits, how to ensure scientific validity, and how to ensure that users and stakeholders know enough about the treatments to give valid consent. As with several other chapters in this volume, the authors aim not to settle these issues but only to begin an informed discussion.

Part II: Cognitive Neuroscience and Philosophy

The second group of chapters addresses basic issues about our mental lives: how we classify and recall what we experience, how we see and feel objects in the world, how we ponder plans and alternatives, and how our brains make us conscious and create specific mental states.

Touch

Our senses enable us not only to gather information about the world, but also to navigate surroundings successfully. Given the pivotal role our senses play in our mental life, it is unsurprising that philosophers' interest in understanding them is as old as philosophy itself. Indeed, philosophers' views on the reliability of our senses marks a foundational distinction in the history of philosophy. Empiricist philosophers, on the one hand, hold that our sensory interactions with the external world are our main—and sometimes only—source of knowledge, while rationalists, on the other hand, highlight the unreliability of our senses as a reason to distrust them as dependable sources of knowledge.

Historically, both the philosophy and the sciences of the mind have been much more interested in vision than in the other senses. However, in the past few years, there has been growing interest in some of the more neglected senses, such as somatosensation. In their chapter, Cheng and Cataldo survey the history of the psychology and neuroscience of somatosensation, and hone in on a particular kind of somatosensation: touch. Next, they suggest ways in which scientific research on touch can have important implications for traditional and contemporary questions in the philosophy of perception.

Sight

As mentioned before, vision is the most extensively studied of all of our senses, and researchers have approached the study of vision from a wide variety of perspectives. Much of that research has been aided by the development of precise mathematical and computational models that help to account for a number of behavioral findings, as well as to predict performance in a variety of tasks. These models have been used to capture, for instance, reaction times under different conditions, perceptual sensitivity, discrimination among different stimuli, and precision as a function of different visual features.

It is less clear, however, how these different mathematical and computational models relate to a critical feature of our visual experience: its phenomenological dimension. In their chapter, **Denison, Block, and Samaha** survey four commonly used quantitative models of visual processing: signal detection theory, drift diffusion, probabilistic population codes, and sampling. Then, based upon research in the philosophy and the science of visual experience, they discuss possible ways in which these models can map onto different features of the contents of our visual experiences. While they find that no model seems to fit all relevant features of our visual contents equally well, they offer several avenues for continuing to explore the relationship between quantitative models of visual processing and the phenomenology of visual perception.

Consciousness

Research on the neural substrates of our conscious experience has surged in the last few decades. Nevertheless, the question as to how our brains give rise to our conscious, subjective experience remains unanswered. A common approach to investigate the neural substrates of conscious experience is to analyze differences in brain activity associated with stimuli of which participants are consciously aware relative to those of which participants are not consciously aware. The thought here is that such differential activity would reflect the neural mechanisms responsible for the conscious awareness of the stimuli.

Although widely used, this strategy is almost always confounded with perceptual task performance (i.e., the measured degree to which a participant succeeds at achieving the experimental goal), which has been shown to dissociate from conscious awareness. In their chapter, **Morales, Odegaard, and Maniscalco** explore the issue of task performance as a confound in research on conscious perception. Specifically, they critically evaluate the strategy of performance matching, which has been thought to provide a solution to the problem of performance confounds, and suggest a signal detection theoretical model that can help to overcome the challenges associated with this experimental strategy in particular and with experimental design in conscious perception in general.

Memory

We are conscious not only of what occurs in our surroundings but also of what happened in our past. Memory, after all, allows us to bring back to mind previous experiences. It is therefore tempting to think of memory as a cognitive faculty dealing primarily with time. Although how exactly the brain manages to support our recollections is still a matter of debate, most researchers agree that the hippocampus and adjacent medial temporal lobe regions are necessary for encoding and, likely, retrieving memories too. Surprisingly, however, for the first two thirds of the twentieth century, work on the neurophysiology of the so-called hippocampal complex showed it to be strongly associated not so much with time but with space. A little more than forty years ago, the theory of the hippocampus as a cognitive map brought together time and space in the hippocampus in a profoundly influential way.

In their paper, **Robins, Aronowitz, and Stolk** critically revisit the notion of a cognitive map and explore its uses in cognitive neuroscience above and beyond merely being a representation of space. As they show, many researchers use the term "cognitive map" to cover representations of nonspatial information as well. How do cognitive maps scale up from representing space to capturing nonspatial information? As their chapter shows, the answer to this question is complex and has important consequences for understanding not only memory but also several other cognitive processes.

Concepts

In addition to perceptions and memories of particular events, our minds also seem to be furnished with general concepts. Theories of concepts have long traditions in both the philosophy and sciences of the mind. Philosophers have traditionally wondered how abstract concepts fit into a world of particular objects. Cognitive neuroscientists who study concepts often see themselves as investigating the constituents or building blocks of semantic memories. Unfortunately, not only is there lack of clarity as to how precisely to characterize what a concept is, there is also quite a bit of disagreement when it comes to relating the notion of "concept," as employed in cognitive neuroscience, with its uses in psychology and philosophy.

In their chapter, **Leshinkaya and Lambert** explore the nature of this disagreement first by critically examining the way in which concepts are said to be studied in the cognitive neuroscience of long-term semantic memory. Then, they go on to show why this way of understanding "concept" does not dovetail with the way psychologists and philosophers use the term. But the issue is not simply about words. The concern is that if one wants concepts to perform a certain role in our cognitive ontology, then what cognitive neuroscientists investigating concepts are doing now cannot fulfil that role. At the end, the authors offer a suggestion, inspired by recent developments in philosophical discussions of concepts, according to which concepts exhibit "sharedness"—a feature, they suggest, to which cognitive neuroscientists interested in understanding the neural underpinning of concepts should pay attention.

Mind Wandering

As much as we wish to stay focused and undistracted when carrying out a particular task, we often find that our minds wander about, plagued with unrelated thoughts. This phenomenon is frequent, common, and almost always involuntary. For the past two decades, research on mind wandering has blossomed, with several labs now dedicated to studying not only the cognitive characteristics of our minds' tendency to wander away from a task at hand, but also the neural underpinnings of this tendency. However, despite this explosion in scientific interest, the scientific study of mind wandering has been also heavily criticized, in no small part because of its reliance on self-reports.

In their contribution to this volume, **Murray, Irving, and Krasich** walk us through the history and justification of the use of self-reports in the scientific study of mind wandering. In so doing, they build a case as to why self-reports are, in fact, indispensable. Moreover, they critically survey other methods, which allegedly can offer "objective" measures of mind wandering, and they suggest that their validity ultimately depends on self-reports. Finally, they offer both scientific and philosophical arguments as to why self-reports are critical not only for the study of mind wandering but also for our understanding of this phenomenon itself.

Mind and Brain

An overarching, central question in both the philosophy and the sciences of the mind asks what the relationship is between the mind and the brain. Theories about the nature of this relationship abound, but the prospect of them receiving empirical validation has been less obvious. In a sense, cognitive neuroscience was conceived as offering precisely that: the possibility of bridging the psychological and the neural levels, aided in part by the use of novel and revolutionary technologies for exploring the brain in vivo. Many philosophers and neuroscientists think that in order to reach such bridging, we need first to identify our "cognitive ontology"—that is, a taxonomy of the basic psychological kinds that can be correlated with structural and/or functional brain structures. Unfortunately, despite its promise, cognitive neuroscience has yet to deliver a clear consensus on what that cognitive ontology might be.

In their paper, **McCaffrey and Wright** critically examine different attempts to characterize a "correct" cognitive ontology and instead advocate for a kind of pragmatic pluralism according to which different cognitive ontologies would be required for different purposes. The consequences of such pluralism for cognitive neuroscience are indeed profound, but as argued by the authors, there is much promise in relinquishing the idea that there is a single taxonomy of the human mind.

Conclusion

Overall, these chapters display a wide variety of ways in which philosophers and neuroscientists can learn from each other. Of course, this volume does not claim to cover all of the issues that are shared between these disciplines. Nor do any of these chapters claim to be the last word on its topic. Still, we hope that these chapters will inspire philosophers and neuroscientists to work together to the great benefit of both fields.

This is a section of doi:10.7551/mitpress/12611.001.0001

Neuroscience and Philosophy

Edited by: Felipe De Brigard, Walter Sinnott-Armstrong

Citation:

Neuroscience and Philosophy
Edited by: Felipe De Brigard, Walter Sinnott-Armstrong
DOI: 10.7551/mitpress/12611.001.0001
ISBN (electronic): 9780262367332
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif by Westchester Publishing Services. .

Library of Congress Cataloging-in-Publication Data

Names: Brigard, Felipe de, editor. | Sinnott-Armstrong, Walter, 1955– editor.
Title: Neuroscience and philosophy / edited by Felipe De Brigard and
Walter Sinnott-Armstrong.
Description: Cambridge, Massachusetts : The MIT Press, [2022] |
Includes bibliographical references and index.
Identifiers: LCCN 2021000758 | ISBN 9780262045438 (paperback)
Subjects: LCSH: Cognitive neuroscience—Philosophy.
Classification: LCC QP360.5 .N4973 2022 | DDC 612.8/233—dc23
LC record available at https://lccn.loc.gov/2021000758