Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Olivia Odoffin April 17, 2013

Toward an Interdisciplinary Theory of Consciousness: Methodology, Reductionism, and Mechanistic Analysis

by

Olivia Odoffin

Dr. Robert McCauley
Adviser

Neuroscience and Behavioral Biology Department

Dr. Robert McCauley

Adviser

Dr. Richard Patterson

Committee Member

Dr. Hillary Rodman

Committee Member

2013

Toward an Interdisciplinary Theory of Consciousness: Methodology, Reductionism, and Mechanistic Analysis

by

Olivia Odoffin

Dr. Robert McCauley

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience and Behavioral Biology Department

2013

Abstract

Toward an Interdisciplinary Theory of Consciousness: Methodology, Reductionism, and Mechanistic Analysis

by Olivia Odoffin

Many approaches can be taken with regard to creating a theory of consciousness. A theory of consciousness, in my view, should take into account findings from scientific studies of conscious experience by cognitive science. A theory of consciousness, then, will be a scientific theory. Problems in the philosophy of science regarding the construction of scientific theories are especially relevant to illuminating the way in which a scientific theory of consciousness will be formed. I argue that a theory of consciousness will be constructed with the foundation of mechanistic analysis rather than traditional views of strict reductionism. Since mechanistic analysis emphasizes the importance of multiple levels of analysis, a theory of consciousness will be interdisciplinary. Traditional views of theory construction in science fail to address the criteria needed for interdisciplinary theories as opposed to theories involving one discipline. An interdisciplinary of consciousness will need to incorporate bridge sciences to connect experimental data from the various disciplines involved in consciousness studies.

Toward an Interdisciplinary Science of Consciousness: Methodology, Reductionism, and Mechanistic Analysis

by

Olivia Odoffin

Dr. Robert McCauley

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience and Behavioral Biology Department

2013

Acknowledgements

This thesis is the product of many years of intellectual development and transformation. The Mellon Mays Undergraduate Fellowship (MMUF) provided me with the resources to cultivate my burgeoning passion for the philosophical debates about conscious experience. The first stages of this thesis began nearly two years ago as a prospectus for the MMUF Summer Institute. My interest in consciousness eventually transformed into the present topic by way of the numerous conference papers written for MMUF. MMUF provided me with invaluable opportunities to consistently fine-tune my intellectual interests. I am grateful for the consistent support and encouragement of the MMUF family here at Emory.

I would also like to thank my thesis adviser and Mellon mentor, Dr. Robert McCauley, for his patience, understanding, and persistence for the past two years. His scrupulousness with regard to my writing has provided me with a glimpse of the rigorousness of philosophical scholarship. As a result, I have become a better scholar than I was at the beginning of this two-year-long journey, and I look forward to continuing to grow and develop philosophically.

Finally, I want to thank my friends Camille Maddox, Michael Harris, Aerhealle Hampton, and Tiana Boykin for their emotional support through this final semester of my undergraduate career.

Table of Contents

Introduction

Descartes, widely considered to be the father of modern philosophy, substantially influenced modern metaphysics (Secada, 2000). Particularly, his infamous phrase "*Cogito ergo sum*[1]" continues to impact contemporary conceptions of the mind/body problem, which lies at the intersection of metaphysics and philosophy of mind. Descartes' expression is representative of Cartesian dualism, the belief that the mind is fundamentally distinct from the body. Scholars have largely abandoned Cartesian dualism in contemporary philosophy, but a sustained interest in the centrality of consciousness to cognition remains[2]. Contemporary formulations of the mind/body problem are concerned with the relationship between consciousness and the brain.

Consciousness studies is becoming an increasingly popular area of research. An offshoot of cognitive science, consciousness studies involves the efforts of philosophy, neuroscience, and psychology in an investigation of conscious experience. Albeit from a different perspective than traditional philosophy, cognitive scientists also seek to determine how neural activity gives rise to conscious experience. As I will show, the interdisciplinary nature of consciousness studies is potentially problematic for the formation of a sufficiently comprehensive theory of consciousness[3].

Due to the innumerable approaches for analyzing conscious experience, it is necessary to limit the scope of the present discussion. For this reason, many assumptions about the nature of

---

[1] English translation: "I think, therefore, I am."

[2] By this statement, I do not intend to assert that interest in consciousness is a result of Descartes. At the very least, people have been interested in consciousness since antiquity. However, Descartes' formulation of the mind/body problem has had the most direct impact on contemporary views of consciousness.

[3] The phrase "sufficiently comprehensive theory of consciousness" appears throughout this thesis. By "sufficiently comprehensive," I simply mean a theory of consciousness that incorporates data from all of the disciplines included in consciousness studies.

conscious experience, the relationship between the mind and the body, the ability to create a scientific study of consciousness need to be addressed.

An underlying, and controversial, assumption of this thesis is that a theory of consciousness will be a *scientific* theory. Although the word "consciousness" does not often appear in scientific literature, phenomena that involve conscious experience—such as awareness, explicit memory, and decision making—are commonly studied. Thus, despite philosophical objections, we are already on our way to an understanding of conscious experience through the use of scientific methods. For the purposes of this thesis, the point is not to assert that there are metaphysical barriers to understanding consciousness, especially regarding its phenomenology. It may be the case that there is some essential aspect of conscious experience that cannot be described by physical theories (Jackson, 1986). It may be the case that we are fundamentally unable to understand all of the features of our consciousness (McGinn, 1989). It may be the case that science is trying, and currently failing, to provide a sufficiently descriptive account of conscious experience (Chalmers, 1996).

Even if some, or all, of these predictions are eventually shown to be accurate, I believe it would be a mistake to halt the current progress researchers are making in developing an account of consciousness. Consciousness studies, as it currently operates, is not committed to any assumptions about the intractability of consciousness. Certainly, developing a theory of consciousness is not an easy task. Even considering all of the progress made in consciousness studies, much more needs to be done before a comprehensive theory of consciousness can be formulated. Despite the difficult task ahead, the core assumption of the mind/brain sciences is that mental activity is a result of brain activity. Keeping this in mind, it is important for scholars to continue to search for an explanation of how neural activity can give rise to conscious experience.

In short, I do not intend to settle metaphysical and epistemological debates about the nature of consciousness. The purpose of this project is to develop criteria for a theory of consciousness. Thus, I approached this thesis with an interest in the methodological barriers to a science—and eventually, theory— of consciousness. While numerous metaphysical and epistemological problems and assumptions underlie this discussion, the main concern is methodological. I argue that mechanistic analysis, rather than strict reductionism, is better suited for the biological and psychological sciences, and thus, for a theory of consciousness.

This thesis is grounded in a discussion of central issues in the philosophy of science that will be useful to unravel the problems associated with theory formation. Chapter 1 serves to foreground the relevant problems in the philosophy of science. Chapter 1 begins with a description of consciousness studies[4]. Then, to introduce the problem of demarcation, I discuss the difference between colloquial uses of "theory" and the uses of theories in science. The criteria for scientific theories are intimately tied to differentiating scientific statements from non scientific statements. The relevant criteria for scientific theories are also relevant for a theory of consciousness; these criteria will ensure that a theory of consciousness will be scientific.

The interdisciplinary nature of consciousness studies raises additional concerns about the relationship between theories in different scientific disciplines. Chapter 1 goes on to discuss the deductive-nomological model of explanation and traditional reductionism. I raise the theoretical

---

[4] The interdisciplinary nature of consciousness studies discussed in the beginning of Chapter 1 serves as a recurring theme throughout this discussion. The emphasis on interdisciplinarity in the study of consciousness will, as I argue, constrain what theoretical foundations are appropriate for a theory of consciousness (i.e. reductionism versus mechanical analysis) and add criteria to a theory of consciousness.

problems with the expectation of intertheoretic reduction in science before asserting that a new model of mechanistic analysis better represents methodologies in the biological and psychological sciences.

In Chapter 2, I detail my argument for mechanistic analysis over strict reductionism by appealing to arguments raised by Paul and Patricia Churchland. Since the Churchlands are receptive to strict reductionism—and in some cases, eliminativism—they provide a foundation for illustrating that mechanistic analysis better models current research strategies in the mind/brain sciences than strict reductionism. Chapter 2 shows that many of the Churchland's assertions about consciousness are compatible with the type of mechanistic analysis advocated by William Bechtel.

Under the tools of mechanistic analysis, Chapter 3 explains that the interdisciplinary nature of the science of consciousness will lead to an interdisciplinary theory. I appeal to Owen Flanagan's description of the natural method to show that a theory of consciousness will incorporate data from phenomenological, psychological, and neuroscientific data[5]. Valerie Hardcastle's criteria for interdisciplinary theories in cognitive science are applicable to a theory of consciousness. Appealing to Flanagan and Hardcastle, I explain that a theory of consciousness will be interdisciplinary. Overall, I hope to show how philosophical questions in conjunction with scientific methodologies can illuminate our understanding of conscious experience.

[5] This discussion emphasizes the contribution of data from psychology and neuroscience in a theory of consciousness. Consciousness studies, however, more disciplines are involved in consciousness studies than psychology and neuroscience. I limit this discussion to neuroscience and psychology purely for practical purposes. A theory of consciousness should include other domains such as anthropology and computer science. In the third chapter/ the epilogue, I argue that a theory of consciousness—under the criteria I present—will be extendable to the other disciplines involved in consciousness studies.

Chapter 1: Scientific Theories and Reductionism

"Consciousness studies" is an umbrella term used to describe the interdisciplinary effort to provide an account of consciousness. Consciousness studies is a subset of cognitive science. Cognitive science is an interdisciplinary examination of multiple aspects of cognition. Although consciousness has not always been categorized as a viable area for scientific investigation, consciousness is now being investigated by an ever increasing number of cognitive scientists. Several disciplines in cognitive science are investigating consciousness—including neuroscience, psychology, and philosophy.

As a result of the multiple disciplines involved in consciousness studies, consciousness is typically not investigated as a unitary phenomenon. The disciplines involved in consciousness studies attempt to investigate all of the features of consciousness. There are several ways of construing conscious experience. David Rosenthal has conceptualized consciousness into creature consciousness and state consciousness (2002). Creature consciousness is attributable to the entire organism, while state consciousness corresponds to mental states that are said to be "conscious." Examples of creature consciousness would be different levels of arousal, such as unconsciousness versus wakefulness, and self-consciousness. Many features of consciousness fall under Rosenthal's category of state consciousness. State consciousness includes qualitative states, or qualia, and access consciousness.

Section I: Distinguishing Scientific Theories from Non-Scientific Theories

Consciousness studies is a fairly recent development; prior to the 1980s, scientists considered consciousness too subjective for scientific investigation. The scientific method is traditionally thought to be "objective." This presumed objectivity of science corresponded to the idea that subjects for scientific inquiry were supposed to be mind-independent. Consciousness is a perspective that is at least intimately tied to humans, if not other non-human animals.

A fundamental assumption of this thesis is that consciousness is able to be investigated scientifically. Since I aim to explore the requirements for a theory of consciousness, it is important to note that a theory of consciousness, in this view, will necessarily bear some relation to the general criteria for scientific theories. Philosophers of science are concerned with determining what separates scientific theories from non-scientific theories. Statements that are amenable to scientific inquiry have certain qualities non-scientific statements do not seem to have. A theory of consciousness will have to conform to the expectations of scientific theories in general, so it is important to note what criteria demarcate a theory as "scientific."

In contrast to the negative assumptions about "theory" in everyday language, theories in science are the most reliable form of scientific knowledge. Claiming an idea to be a "theory" in colloquial language disparages that idea as speculative or ill-formed. In this informal context, a theory is an idea without adequate evidence to support it. A "theory," in this instance, is an idea that has little explanatory value due to a lack of evidence. In contrast, scientific theories are the converse of this ordinary conception of "theory." "Theory" is used in science to describe an idea that is supported by evidence, and thus, has explanatory value. Scientific theories provide frameworks for explaining a phenomenon or set of phenomena.

Scientific theories organize scientific knowledge, and for this reason, theory formation is a common goal of scientific enterprises. Researchers investigate phenomena through narrow examination. The natural sciences in particular must worry about keeping control over the parameters of an experiment. Additionally, technological constraints and a commitment to precision often render it unfeasible to investigate more than one or a few specific parts of a phenomenon. Due to the narrow purview of most experiments, the resulting data are often disjointed and is difficult to consolidate into a larger picture. Scientific theories are perhaps the best way to unify disparate data in order to build a more complete explanation of some phenomenon. Theory formation is a difficult and lengthy process, but it is a common enterprise because it is a way of unifying knowledge. Chemists develop theories about chemical systems, economists develop theories about decision making, and cognitive scientists develop theories about cognitive faculties. A primary goal of consciousness studies, much like other forms of scientific investigation, is to develop a *theory* about conscious experience

Scientific theories are in contrast with the judgment of "theory" in everyday language because researchers systematically test scientific theories using experimental procedures. These procedures include mathematics, interviews, observation, and experimentation. The method used to test a hypothesis or theory depends upon the phenomenon to be tested. Regardless of the method of inquiry, one of the primary goals of scientific investigation is to provide an explanation of phenomena. As previously stated, scientific theories contribute to this primary goal of science. Because the sciences purport to explain regularities in the world and scientific theories are the most organized formulation of scientific knowledge, theories in science must systematically explain events in the world.

The connotation of "theory" used in science and "theory" used in everyday language differs due to the assumed commonalities among scientific theories. A central area of concern in the philosophy of science is to figure out what differentiates scientific statements from non-scientific statements. Philosophers of science examine scientific theories in order to determine the common features that are found in all scientific theories. Ideally, once scholars illuminate the reason scientific theories are "scientific," they will be better able to formulate new hypotheses and theories for investigation while also excluding non-scientific statements from scientific inquiry. A few of the most well known and debated criteria for determining a theory to be “scientific” include the support of observational data, the principle of falsifiability, and independent testability.

Observational support. Systematic observation is integral to the traditional formulation of the scientific method, and scientific theories, in turn, need to be supported by observational data. Scientific theories need to be able to explain why the world appears as we perceive it. They also need to be able to explain phenomena in the world that aren’t apparent from pure sensory observation. For example, atomic theory posits that matter is composed of atoms (although we now know they are not indivisible). Atoms are not viewable by the naked eye, but atomic theory is able to explain why matter is composed in a specific way. Data itself does not need to be observable by the senses. Mathematical formulas, for instance, are a form of evidence that is not directly observable by the senses. Observational data takes many forms beyond perception, and it includes evidence generated from scientific investigations that are not readily perceivable without the use of sophisticated instrumentation. Non-observable data, however, are not by itself sufficient to support a scientific theory. Non-observable data, such as mathematical formulas, also need to conform to the observable phenomena that it tries to support.

Scientific theories unify data, and thus, are formed after a systematic combination of observation and experimentation. Experimentation begins and ends with observation; experiments are conducted based on observational data, and the results of experiments are observed and recorded. Scientific theories attempt to synthesize observational data measured both before and after experimentation. Observational and experimental data that initially appear to be disparate can be explained by a sufficiently comprehensive theory. The best scientific theories are able to explain any discrepancies between observational data before experimentation and data from experimental results. For example, heliocentric theories are able to explain the observational assumptions of geocentric theories, such as why it appears that the earth is stationary while the heavenly bodies are in constant motion[6].

The problem of demarcation and the principle of falsifiability. Inductive reasoning is commonly thought to distinguish scientific enterprises from non-scientific enterprises. Inductive reasoning, in contrast to deductive reasoning, derives general statements from specific instances. Scientific reasoning appears to operate in this manner; theories seem to develop only after observing specific instances. Even though induction seems to be commonplace in science, philosophers have noted what is known as “the problem of induction.” The 18th century philosopher David Hume provided a clear formulation of the problems with inductive reasoning. According to Hume, no logical justification can be provided for asserting a causal relationship for some particular event (2001). That is, it is not justifiable to predict the effect of some future event because we cannot be completely certain of the regularity of nature. We believe induction to be a viable form of reasoning only because we have observed some habitual relationship between two events.

Karl Popper maintained that the central problem in the philosophy of science was to distinguish between scientific statements and non-scientific statements[7]. Karl Popper describes the difficulty of distinguishing a scientific statement from non-scientific statements—including

[6] The discussion of the problem of demarcation that follows is based upon Popperian arguments. Popper, however, would not agree with the primacy of observation in science; for Popper, science begins with problems rather than observation. Popper disagrees that observation is prior to theories. According to Popper, our perception of the external world is necessarily theory laden. Because of this, observation does not differentiate scientific theories from other theories (Popper, 1963).

[7] The logical empiricists addressed this issue prior to Popper, but for the purposes of this thesis, I am using Popper’s formulation of the problem of demarcation.

religious, pseudo-scientific, and metaphysical statements—as the "problem of demarcation" (1963). He states, "in order [for a statement] to be ranked as scientific, [it] must be capable of conflicting with possible, or conceivable, observations" (Popper, 1963, p.51).

Popper's solution to the problem of demarcation was strongly influenced by Hume's assertion that inductive reasoning was unjustifiable. Inductive reasoning was thought to be central to scientific reasoning, but since Hume claimed that inductive reasoning was faulty, Popper sought an alternative solution to the problem of demarcation. Observation is central to scientific methods, but Popper maintained that scientific questions began by thinking of problems instead of seeking an explanation for observable phenomena (1963). Without accepting induction or observation as central to science, Popper had to seek an additional characteristic of science to solve the problem of demarcation.

Popper asserts that the problem of demarcation is solved by the principle of falsifiability. He famously argued that all scientific theories are "falsifiable." The principle of falsifiability claims that scientific theories need to be articulated such that they can be shown false (Popper, 1963). According to Popper, the testability of a scientific theory relies on its ability to be falsified[8]. If a theory is formulated in such a way that observational data can neither confirm nor deny its claims, the theory is not falsifiable, and thus, cannot be tested by scientific means. Scientific data cannot conclusively show that a theory is correct. Evidence garnered from experimental studies, at most, provides support for scientific theories or hypotheses. However, the principle of falsifiability claims that scientific theories can be shown to be false with a sufficient amount of contrasting evidence. Scientific theories do not need to be actually shown to be false in order to be considered falsifiable; at most, the principle of falsifiability states that

[8] For criticisms of Popper's criterion of falsifiability, see: Kuhn, 1970 and Lakatos, 1970.

theories need to be able to be shown to be false *in principle*. It only needs to be possible that new observational data can come along and show a theory to be false. If a theory or hypothesis is not falsifiable, then it cannot be tested by experimentation.

Independent testability. Theories may be applied outside of the domains for which they were originally intended. This possibility for theories to explain phenomena outside of their targeted domain is known as independent testability. According to Philip Kitcher, a theory has independent testability because auxiliary hypotheses are capable of being added to the theory (2002). These auxiliary hypotheses allow the theory to be updated to include new phenomena that were not included in the original formulation of the theory. Both scientific hypotheses and theories are falsifiable, but theories are broader in scope than scientific hypotheses. Hypotheses predict the results of a single experiment, study, or investigation. Hypotheses are not extendable to evidence outside of the study for which they are formulated. Scientific theories, on the other hand, are extendable to domains that they were not originally meant to encompass. The formation of scientific theories often involves the results of multiple hypotheses from several experimental studies. The applicability of a theory may be tested after it has already formed by testing it with a new hypothesis. If a theory's explanatory scope is able to account for the data from studying new hypotheses, then the theory can be extendable to a different domain.

The number of domains a theory can be extended to is a sign of the theory's strength. The ideal scientific theory is elegant, simple, and able to explain a wide number of phenomena. Old scientific theories are often disregarded if a new theory is able to explain the same events of the old theory in addition to other phenomena that the old theory did not explain. For most purposes, Newton's laws are sufficient to explain many regularities of motion. Newton's theory was hailed by many to be the ultimate physical theory. Once physics advanced, researchers discovered anomalous phenomena that Newton's theory did not account for. Researchers attempted to account for the deficiencies in Newton's theory by adding additional axioms to explain irregular patterns of motion. Eventually, Einstein's theory of relativity encompassed the same domains as Newton's theory of motion, but it also explained the irregularities that were not explained by Newton's theory. Thus, Einstein's theory of relativity has a greater explanatory scope than Newton's theory of motion. Although auxiliary hypotheses were added to Newton's theory, a new, more general theory developed with the ability to describe more domains in physics.

A theory of conscious experience would follow these general criteria for scientific theories. A theory of consciousness would need to be supported by observational data, be falsifiable, and be able to account for new evidence. A theory of consciousness will likely incorporate data across domains; many disciplines are involved in consciousness studies, and a theory of consciousness would incorporate the data found in these disciplines. Data found in neuroscience and psychology would need to be unified by a theory of consciousness. Essentially, an overarching theory of consciousness would serve as a framework for explaining all observable dimensions of consciousness, including both state and creature consciousness. Additionally, an all encompassing theory of conscious experience would need to be formulated in such a way that it can, in principle, be shown to be false by scientific data. The principle of falsifiability separates a non-scientific theory from a scientific theory. In order for an account of consciousness to be based upon scientific findings, it needs to be formulated in a way that is empirically testable. Lastly, a theory of consciousness would be differentiated from a mere hypothesis about consciousness if it is independently testable—i.e., is capable of being extended into a new domain.

A theory of consciousness is the Holy Grail of consciousness studies, and so far, the theories of consciousness that exist emphasize one area of inquiry over another. A few examples include Bernard Baars' "Global Workspace Theory" (1997), the various representational theories, and even quantum theories of consciousness, which emphasize cognitive psychology, philosophy, and physics respectively. A theory of consciousness that attempts to account for data found across the many disciplines involved in consciousness studies is a daunting task. Scientific theories are primarily about the unification of empirical findings. Considering that current scientific knowledge about consciousness only amounts to snippets of data, it should be no surprise that there has not yet been a comprehensive theory of consciousness.

Section II: Raising Problems for Reductionism in the Philosophy of Science

Many levels of explanation are present both between disciplines and within disciplines. As a result, a fundamental issue in the philosophy of science must be raised: how can evidence and theories from different disciplines be integrated to explain a particular phenomenon, such as consciousness? In consciousness studies, neuroscience investigates consciousness from the level of computational neuroscience, cellular neuroscience, systems neuroscience, and behavioral neuroscience. Psychology investigates consciousness from the level of cognitive psychology, social psychology, and developmental psychology. In addition to these levels of inquiry formed from investigating consciousness' relation to the brain, consciousness has an additional level of explanation: subjective experience. Subjective experience seems to add an additional problem to the formation of a theory of consciousness outside of issues relating to disciplines themselves. Before the problem of subjective experience is raised, the nature of theoretic reduction between and within sciences needs to be discussed.

As traditionally conceived, the ideal of reduction rests on the assumption that scientific disciplines are related to each other. Scientific disciplines are often conceived as having a hierarchical structure. In this conception, physics is considered the most fundamental science. The principles in physics are conceived of as being universal, and often, eternal. The rest of the scientific disciplines have often been conceived as special applications of physical principles. In this manner, the physical laws are thought to be the foundation of all other levels of scientific inquiry. Thus, physics is at the bottom of the disciplinary hierarchy. From the foundation of physics, the rest of the hierarchy can be constructed—chemistry is at the level above physics, then biology, then psychology, then sociology, and so on. The theories in each of these disciplines occupy a "level" of explanation relative to the levels of the other sciences. A theory in physics is at a "lower-level" than a theory in chemistry, which is at a "higher-level" than a theory in physics. Scientific theories across all disciplines can be constructed in this manner. For example, psychology is at a higher-level than biology, but it is at a lower level than sociology.

The doctrine of "reductionism" rests on the idea that theories at the more fundamental level can explain theories at a higher level. In its most extreme form, reductionism will lead to all theories being reduced to physical laws. Ernest Nagel was a famous proponent of scientific reduction. Using examples from the history of science, he proposed that the unification of scientific disciplines rests on the ability for a higher-level theory to be reduced to a lower-level theory (Nagel, 1979). A paradigmatic example that is used to support reductionism is the reduction of Kepler's laws of motion to Newton's laws of motion. Within astronomy, Newtonian laws of motion are broader and more encompassing than Keplerian laws of motion. Newton's laws of motion were able to explain Kepler's laws of motion. Because of this, Kepler's laws of motion were reduced to Newton's laws of motion. This historical fact provides a good example of theoretic reduction within a discipline.

Reductionism has another implication that is an offshoot of the unification of scientific theories. As envisioned by Nagel (1979), theoretic reduction also signifies scientific progress. It was considered a sign of scientific progress when Newton's laws of motion superseded Kepler's laws of motion. It was a further sign of scientific progress when Einstein's theory of relativity explained Newton's laws of motion and more. Thus, as understood by Nagel, scientific progress occurs through the reduction of scientific theories.

The "deductive-nomological model" (D-N model) is the formal terminology for Carl Hempel and Paul Oppenheim's (1958) proposal for scientific explanation. In the D-N model, scientific explanations are both deductive and nomological—"deductive" meaning that they should follow the format of a deductive argument in formal logic, and "nomological" meaning that the premises of the argument should be formulated in terms of universal laws. In short, according to the D-N model, scientific explanations should follow the form of a deductive argument, and the premises should be in the form of laws. The *explanans* and the *explanandum* are two parts of the formulation of scientific explanation. The explanandum is the phenomenon to be explained, and the explanans is what provides the explanation.

The D-N model of explanation fits the reductionist ideal. Theoretic reduction can occur through the DN-model. The higher-level theory would be the explanandum, while the lower level theory would be the explanans. Appealing again to the Newton and Kepler example, Newton's laws of motion are the explanans and Kepler's laws of motion are the explanandum. When the theory in the explanandum can be explained by laws in the explanans, the theory in the explanandum often loses its explanatory value. When this is the case, the theory in the explanandum is reduced to the theory in the explanans. The theory used in the explanandum is eliminated.

The appeal to laws in the D-N model of explanation may work for certain disciplines (i.e., physics) but it may not work for other disciplines (e.g., biology). As the disciplines move higher up the hierarchy, generalizations become more prominent than laws. Generalizations allow for exceptions; generalizations are not formulated to encompass every single instance of a particular phenomenon. Because of this, generalizations do not describe universal principles. Laws of nature do intend to describe every single instantiation, without exception, of a particular phenomenon. Laws seek to reflect universal principles. Laws are more common in physics than biology or psychology, which explain by way of generalization. The D-N model of explanation fails to account for the emphasis on generalizations rather than laws in the biological and psychological sciences. The appeal to laws is one of the downfalls of the D-N model.

In the strictest form of reductionism, theories in all disciplines are reducible to theories in physics. As noted by many scholars, theory reduction often occurs *within* sciences, but it seems much less likely to occur *between* sciences (McCauley, 2007). Jerry Fodor (1974; 1997) noted that it is not the case that higher-level theories seem to neatly map onto lower level theories. Using the terminology of “unwieldy disjunction” Fodor claims it is unlikely that theories in “special sciences,” and therefore the special sciences themselves, are reducible to theories in other sciences. Fodor asserts that the special sciences are not, even in principle, reducible to physics[9].

The main disciplines involved in consciousness studies do not have the discovery of laws as their main research goal. The possibility of the reduction of psychology to neuroscience is

[9] This is a very brief treatment of Fodor’s argument. I discuss this argument in more detail in Chapter 3.

commonly discussed in philosophy of cognitive science. However, if the reduction of psychology to neuroscience rests on the deductive-nomological model of explanation, it seems unlikely that psychology is reducible to neuroscience. Two reasons in particular seem to solidify psychology's independence; first, neuroscience and psychology appear to operate in terms of generalizations and mechanistic explanations rather than laws, and second, in the Fodorian sense, there will likely be "unwieldy disjunctions" between neuroscience and psychology.

The first, and perhaps most condemning reason that psychology seems unlikely to be wholly reducible to neuroscience is that the main research goal for both disciplines is not to discover laws about nature. For consciousness studies in particular, this becomes obvious through the language of "neural correlates of consciousness." When neuroscientists investigate consciousness through experiments on the brain, they are not looking for laws. Christof Koch defines a neural correlate of consciousness as "the minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept" (2004, p. 16). The search for a neural correlate of consciousness is a search for a neural mechanism of consciousness. The use of mechanistic explanations in neuroscience is not specific to consciousness; there are attempts to find the neural mechanisms of attention, face perception, object recognition, empathy, decision making, among a host of other mental activities.

William Bechtel (1994; 2005; 2007) defends the view that the search for mechanisms is different from the search for laws. The search for law-like explanations is a search for universal principles that are able to explain a class of phenomena and events. The search for mechanisms, in contrast, is not an attempt to discover universal principles. According to Bechtel, mechanistic explanations are confined to a specific set of parts that produces a given phenomenon. In relation to consciousness studies, in both neuroscience and psychology, discovering mechanisms is the driving research strategy. Bechtel's conception of mechanisms is reductive, but it allows for the independence of multiple levels of inquiry. Mechanistic reduction, however, differs from classical reductionism. There are multiple levels within a mechanism with relations between each level. Reductionism in mechanistic explanations happens within the mechanism; because the relations are causal, there need to be appeals to the lower-level of the mechanism. The key feature of Bechtel's conception of mechanisms is that levels of inquiry carry a certain amount of independence. The system needs to be studied both as a whole and at independent levels within the mechanism[10].

[10] This is a very brief treatment of the main features of mechanistic analysis. I discuss this argument in much greater detail in Chapter 2.

Chapter 2: Pitting Reductionism Against Mechanistic Analysis

Paul and Patricia Churchland are known for their revisionist views about everyday conscious experience. They both advocate the scientific study of conscious experience, and they both expect that neuroscientific and psychological experiments will deepen our understanding of consciousness. Philosophers' contentiousness about the Churchlands' positions begins with the Churchlands' position on the reduction—and in some cases, outright elimination—of some aspects of our current conception of conscious experience (Campbell, 1986; Searle, 1992). I argue that with the background of mechanistic analysis rather than intertheoretic reduction, phenomenal experience is the overall activity of a mechanism, and the individual levels in the explanation of the mechanism cannot be reduced to the lowest level of its individual parts.

In Section I, I describe the direct and indirect approaches to a scientific study of conscious experience. In Section II, I introduce the importance of mechanisms in the biological and behavioral sciences and recast the distinction between the direct and indirect approaches in mechanistic terms. In Section III, I outline the implications of mechanistic explanation for the reduction of consciousness to neurobiology compared to traditional formulations of reduction that use laws in explanation. Section IV applies the current discussion of mechanistic reduction to phenomenal experience more explicitly. Finally, Section V addresses accusations regarding the elimination of consciousness.

Section I: Toward a Science of Consciousness

Patricia Churchland outlines two methodological approaches to the scientific study of consciousness—the direct and indirect approach (Churchland, 2002). According to Churchland, the direct and indirect approaches describe two different research methods utilized in consciousness studies. The direct approach and the indirect approach are not mutually exclusive; they are concurrent research strategies in consciousness studies. In fact, both the direct approach and indirect approach are commonly used in both neurobiological and psychological research on consciousness.

The direct approach is to "identify the [neural] substrate as a *correlate* of phenomenological awareness, then eventually get to a reductive explanation of conscious states in neurobiological terms" (Churchland, 2002, p. 134). The direct approach is "direct" because it associates some specific process(es) with some specific psychological event that involves phenomenal experience. The correlate of the phenomenal experience will be the neural activity that happens in accordance with the psychological experience. The physical correlate does not have to be confined to a particular neural system. For example, a particular psychological experience could correspond to the firing pattern of a set of neurons distributed throughout multiple brain areas (Churchland, 2002). Similarly, the correlate does not have to be found only at the level of individual neurons. As Churchland describes it (2002), the correlate to a conscious experience could be found at any level of investigation, whether the molecular, cellular, pathway, or systems level.

The direct approach follows experimental methods in the biological and psychological sciences. In these two areas, research on a particular phenomenon is often conducted through this piecemeal approach. Experiments are designed to investigate an extremely narrow aspect of the particular phenomenon at hand. After close investigation, researchers begin to notice common—or sometimes, disparate—themes in experimental findings. The neural mechanism of the phenomenon is discovered by examining the neural activation patterns that occur at the same time as a corresponding psychological event. After multiple experimental studies activation patterns are consistently associated with a particular psychological event, and a neural mechanism for that event is found. From this, investigators are able to create broader explanations of the particular mechanism's function or functions. Since this general method is common to neuroscience and psychology, it is no surprise that investigations in consciousness studies often follow what Churchland outlines as the direct approach.

The search for the neural correlate of consciousness (NCC) follows the direct approach to a scientific investigation of conscious experience. A NCC associates a psychological state with a neural state. A neural correlate, or mechanism, is considered an NCC if it is both necessary and sufficient for producing the associated conscious psychological event. According to Churchland, "what we seek is the *identification* of some perceptual class of neural activity with perceptual awareness," (2002, p. 154). Francis Crick and Christof Koch (1990), who found that the oscillation pattern of neurons during a conscious experience may be such a correlate of visual conscious experience, exemplify use of the direct approach. Crick and Koch focused particularly on visual awareness and approached the study of consciousness in a piecemeal fashion. This is central to the direct approach; rather than creating a theory about all conscious experience, the direct approach tries to find a mechanism for one instance of consciousness, in this case visual awareness. Following Crick and Koch's study, other searches for a NCC have continued the pattern of isolating mechanisms for different conscious experiences.

The discovery of commonalities between different types of conscious experiences aligns with the indirect approach to the study of consciousness. As described by Churchland (2002), the indirect approach is another method that can be used to investigate consciousness. For the indirect approach, researchers devise theories about conscious experience once they have already developed theories about general functions of the brain. A holistic theory of consciousness will develop after more information has been discovered about mental activities that appear to involve consciousness. This approach is indirect because consciousness itself is not the main phenomenon of investigation. Unlike the direct approach, the goal of the indirect approach is not to find a neural correlate of consciousness, but to create a comprehensive theory of conscious experience.

The indirect approach in consciousness studies expects a theory of consciousness to be formed after the development of theories about other functions that involve conscious awareness. For example, attention, working memory, and decision making are cognitive acts that often involve consciousness. Investigating these processes is not equivalent to investigating the direct correlates of conscious awareness. Because these processes involve consciousness, an investigation of conscious activity that utilizes the indirect approach may try to connect the functions of these processes. For example, Global Workspace Theory proposed by Bernard Baars (1997) uses the indirect approach to theorize about the function of consciousness. Baars' model describes consciousness as the global access to cognitive functions distributed throughout the brain. His theory explains the function of consciousness in terms of cognitive acts that involve conscious awareness, such as working memory and attention. Although theories regarding attention and working memory are not yet complete, Baars' formulated his theory of consciousness in relation to other theories about the general functions of attention and working memory.

Churchland (2002) asserts that the direct approach and indirect approach are not mutually exclusive. In fact, both approaches are useful for a theory of consciousness. According to Churchland, the indirect approach will likely take more time than the direct approach. Research in consciousness studies will primarily occur by means of the direct approach since it takes less time than the indirect approach. A comprehensive theory of consciousness will develop after the discovery of mechanisms of specific types of conscious experiences, such as awareness. Thus, it is possible that progress in consciousness studies will result from progress in both the direct approach and the indirect approach to consciousness studies.

Section II: Mechanisms, the Direct Approach, and the Indirect Approach

The direct approach explicitly involves discussion of mechanisms, but the indirect approach can be recast in the terminology of mechanistic analysis. Recent formulations of mechanistic explanation in the philosophy of science will illuminate these research strategies that Patricia Churchland has outlined. First, the components of the mechanistic framework in neuroscience and psychology need to be discussed.

The popularity of mechanisms has changed over several centuries. In the Renaissance, Descartes popularized a mechanistic view of the world. His physical explanations of the cosmos invoked mechanistic analysis. More importantly, Descartes used mechanisms to explain biological processes. For example, Descartes was famous for his mechanical view of animal behavior. Although he attributed sensations and emotions to animals, he did not consider them to have higher cognition or consciousness (Radner & Radner, 1989). Descartes took the view of animals as mere automata, preferring to describe animal behavior in terms of mechanistic concepts such as reflexes. For humans, Descartes' dualism partially stemmed from his inability to conceive of mechanisms as able to give rise to mental experience. For Descartes, mechanisms were purely physical, and mental abilities seemed purely nonphysical.

The current conception of mechanisms has radically changed from the heyday of mechanistic philosophy during Descartes' era. Now, mechanisms are readily considered to underlie many neural processes—including cognitive and conscious phenomena. This difference is partially due to a change in the conception of the abilities of mechanisms. Descartes' mechanisms were inflexible and, as he imagined them, could not account for higher cognitive abilities. Current talk of mechanisms, as evidenced by the neural mechanisms of consciousness, are said to underlie more complex cognitive processes. Although there are differences among the ways that current scholars describe mechanisms, mechanistic formulations in the philosophy of science generally share a common theme. As described by Wright and Bechtel (2006, p. 45), the resurgence of mechanistic explanation indicates that "many target phenomena and their associated regularities are the functioning of *composite hierarchical systems*" [emphasis in the original]. The description of mechanisms as "composite hierarchical systems" emphasizes the parts within the mechanism, the organization of those parts, and the function of the mechanism as a whole. Bechtel describes a mechanism as follows:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena (Bechtel & Abrahamsen, 2005, p. 423).

Organization of the parts within a mechanism. Crucially, a mechanism cannot be understood only by examining its parts; the organization of the parts within the mechanism must also be understood. The mechanism for a particular phenomenon is designated as such because of its specific structure; the interactions between the different parts of a mechanism are unique to that mechanism. Changing the organization of a mechanism's parts alters the relationship between the parts, ultimately affecting the mechanism's activity.

The mechanism as a whole. The significance of the word "composite" means that, in addition to the function of its individual parts, the function of a mechanism has to be considered holistically. In some mereological relations, the whole is not reducible to its individual parts; the activity of the whole is greater than the activity of its unarranged parts. This is the case for mechanisms. The parts of a mechanism interact with its other parts such that the entire mechanism gains additional functions not attributable to any individual part. The mechanism's overall activity is the result of the complex interactions between its parts.

Levels within a mechanism. The terminology of "levels" is pervasive in the philosophy of science. In the traditional conception of reduction, levels of scientific analysis usually attempt to capture ontological facts about the organization of reality (Nagel, 1979). In one conception of levels, the academic disciplines correspond to structural levels found in the universe. Physics, for example, would be at the lowest-level of explanation because it describes the most general and fundamental facts about the physical world. Levels in all of the other academic disciplines are simply special instances of more foundational physical laws. In the model of traditional reduction, the levels are ontological. Levels can also map onto physical size, closely—but not completely—corresponding to the sub-disciplines within a particular academic discipline. Within biology, the lowest level of explanation would be molecular biology since it investigates the most fundamental parts of the cell. The other sub-disciplines of biology would include and build upon this molecular level.

Understanding levels based upon ontology and size quickly runs into problems. For example, there are varying levels of analysis within individual disciplines. Theories in the biological sciences span both small and large scales of analysis. Watson and Crick's DNA model is at a lower level than Darwin's theory of evolution (McCauley, 2007). Psychological theories are considered to be higher than biological theories. Darwin's theory of evolution is an extremely large scale theory. Evolutionary theory in the biological sciences cannot be clearly organized with respect to a higher level psychological theory. It is not clear that the hierarchical structures of the academic disciplines' ontological levels are neatly structured, and the complexity of the layers within each sub-discipline furthers complicates matters.

Fortunately, the view of levels in mechanistic analysis does not make sweeping claims about the fundamental nature of the physical world. As Bechtel describes, "it is the set of working parts that are organized and whose operations are coordinated to realize the phenomenon of interest that constitute a level" (2007, p. 146). When discussing mechanisms, "levels" refer to local, mereological relationships of the mechanism's parts. In other words, the parts of a mechanism are localized to the particular mechanism they comprise, and they constitute the mechanism in a part-whole relationship. Levels in mechanistic analysis are *compositional.* Unlike ontological levels, levels within a mechanism do not encompass phenomena located outside said mechanism—hence, they are localized. The individual parts of a mechanism, and especially biological mechanisms, do not need to be unique to the particular phenomenon that the mechanism gives rise to. For example, a neuronal pathway may be a part of multiple mechanisms if it is activated during different psychological events. Unlike levels that correspond to physical size, the levels within a mechanism do not need to be of similar sizes. Compositional levels merely correspond to the individual parts that make up the mechanism's hierarchical organization.

The theoretical framework of mechanistic explanation easily conforms to Churchland's description of both the direct approach and the indirect approach.

The direct approach. Discovering the mechanism for a phenomenon mirrors the direct approach because correlating the phenomenon to its mechanism occurs prior to an explanation of the phenomenon. Researchers search for the underlying mechanism of a particular phenomenon in small steps. When investigating a new phenomenon, neither the mechanism's parts nor the organization of those parts are known. Scientists often restrict the scope of an unknown problem so that the problem becomes more manageable. When looking for a mechanism, only fragments of the mechanism are discovered at one time. As a result, explanations of each level of a mechanism are constructed in a piecemeal fashion (Wright & Bechtel, 2006). The overall picture of the mechanism begins to appear only after experimentation. The explanation or the theoretical framework for a phenomenon is developed after its mechanism is discovered.

The indirect approach. The indirect approach is used to find the organization of a mechanism's parts before the development of a theory to explain the mechanism's function. Mechanistic explanation involves the analysis of the connection between a lower level and higher level of organization (Wright & Bechtel, 2006). In the indirect approach, explanations are formulated prior to a theory relating them together. Simply figuring out the parts comprising a mechanism may not be enough to know how the levels connect to each other. *Post hoc* theorizing is often necessary in order to figure out how activity at a mechanism's lower-level can give rise to activity at a mechanism's higher-level, or even to the mechanism's overall activity. The unification of different levels within a mechanism resembles the description of the unification between phenomena in different theories (Darden and Maull, 1977). Similar to the indirect approach in consciousness studies, different aspects of a particular phenomenon are discovered before those aspects are unified.

Section III: Implications for Reduction

Traditional notions of reductionism and mechanistic reduction differ regarding the independence of higher levels of explanation. Because of the emphasis on the organization of the parts within the mechanism, higher levels of analysis have more of a role in explanation than in traditional models of explanation. In traditional models of reductive explanation where scientific laws are emphasized, there is an expectation that the higher-level is simply a *special case* of the laws at the lower-level. In this case, the lower-level law would encompass the higher-level law. However, in the case of mechanistic analysis, the lower-level of the mechanism does not encompass the higher-level of the mechanism. For mechanisms, all levels of organization are important for the mechanism's function. Each of the mechanism's levels has an indispensable role in explaining how the mechanism produces a certain phenomenon.

For both the direct approach and the indirect approach, Churchland expects that discovering the neural mechanisms of conscious experience will allow us to "eventually get to a reductive explanation of conscious states in neurobiological terms," (2002, p.134). Churchland's expectation of explaining conscious phenomena in neurobiological terms is certainly reasonable, but it is unclear what is to be expected of a reductive explanation in neurobiological terms. Is the neurobiological explanation going to occur from only one level of explanation? Will it involve multiple levels of analysis? Or, will conscious phenomena only be explained in terms of the lowest neurobiological level?

Wright and Bechtel explain that mechanistic explanations can be considered both reductionistic and non-reductionistic. In particular, they state:

> Accordingly, in one sense, a mechanistic explanation is through-and-through reductionistic: it appeals to increasingly finer-grain component operations and parts in explaining the activity of a mechanism. But in another sense, a mechanistic explanation is non-reductionistic: explanations at a lower level do not replace, sequester, or exclusively preside over the refinement of higher-level explanations, because mechanisms are hierarchical, multi-level structures that involve real and different functions being

> performed by the whole composite system and by its component parts. Rather than serving to reduce one level to another, mechanisms *bridge* levels (Wright and Bechtel, 2006, p. 55).

Mechanistic explanations do appeal to a mechanism's constituent parts at lower levels, but these parts at lower levels do not take precedence over the mechanism's parts at higher levels. Mechanisms must be understood as a whole. Accordingly, the mechanism needs to be understood at multiple levels and in its specific context (Machamer, Darden, and Craver, 2000). Mechanistic analyses are appealing precisely because they emphasize the unification of different levels of organization and analyses (Glennan, 2002). Reductionism unifies different levels of analyses, but only at the expense of higher levels.

In mechanisms, higher levels play a crucial role in the organization of the mechanism, and are therefore, indispensable to the mechanism itself and to explanations about the mechanism's function. Activity at a higher level of the mechanism builds upon, and is irreducible to, activity at a lower level of the mechanism. Bechtel uses an example of the cellular mechanism that synthesizes ATP to illustrate this point (2008). ATP is a coenzyme that is essential for providing energy to cells, and needs to be synthesized from the energy present in food. ATP can only be synthesized by the interaction of the enzymes that constitute its mechanism; an individual enzyme is not sufficient to produce ATP. Hence, the overall function of ATP's mechanism cannot be explained solely in terms of the individual functions of the mechanism's parts. Because of the irreducibility of the levels within a mechanism, each level of the mechanism adds some additional feature to the mechanism's function. An explanation of the mechanism's overall activity needs to take into account the contributions of each level within the mechanism.

The possibility of applying reductionism to a mechanistic framework depends upon the relationship between a mechanism's compositional levels and the levels of analysis in scientific disciplines. In an ideal situation, if a mechanism's levels neatly correspond to levels of analysis (which would then map on to the ontological levels), then the structure of the mechanism's higher-levels should be explainable in terms of the highest relevant level of scientific analysis. Correspondingly, the structure of the lowest organizational level in the mechanism should be explainable in terms of the lowest relevant level of scientific analysis. For example, imagine that the lowest-level of a mechanism of consciousness is explainable in terms of molecular neuroscience, and the next highest level is explainable in terms of cellular neuroscience, and the higher levels of the mechanism are explainable in terms of the higher levels of neuroscience or psychology. In a strictly reductionistic framework, all of the levels of organization within the mechanism would be explainable in terms of the lowest, molecular neuroscience level. In a mechanistic framework, all of the levels of organization have a level of independence such that none of the higher-levels would be reducible to the lowest, molecular neuroscience level.

In a realistic situation, there is no neat correspondence of levels of organization within the mechanism and levels of analysis in scientific disciplines. For starters, there has been no uncontroversial organization of the hierarchy of the levels of analysis in the scientific disciplines. Additionally, it is possible that levels in a mechanism of consciousness do not correspond to the levels of the scientific disciplines. This would be the case if the oscillation pattern of neural firing across brain structures is a mechanism of consciousness. In the realistic situation, reduction is unlikely because levels of analysis and levels of the neural mechanism may not neatly correspond to each other. In the ideal situation, reduction is unlikely because the levels of a mechanism need to be understood both independently and together.

Section IV: Mechanisms and Phenomenal Experience

Considering the importance of a mechanism's organization, at the very least, a theory of conscious phenomena based upon a mechanistic framework will not be able to appeal only to the lowest level of explanation. If conscious activity is the result of one or more mechanisms, the *whole* of the mechanism will need to be considered. All of the levels of the mechanism will feature in a scientific explanation of the mechanism that generates the phenomenon. For phenomenal experience, this would mean that a theory of consciousness would be an explanation of the mechanism(s) that generate the experience; phenomenal experience would be the explanandum of a theory of consciousness.

If phenomenal experience is the overall activity of some neural mechanism(s), then it can be considered an emergent property of the mechanism(s) from which it results. Emergence, as used here, only refers to a behavior or phenomenon that arises as a result of the activity of a complex system (Bechtel, 2008). In this sense, an emergent property is not a supernatural or fundamental metaphysical property. Similarly to the common usage of "emergence" in philosophical discourse, this description of phenomenal consciousness as an emergent property of a complex system does contrast with strict reductionism. As previously stated, the function of an entire mechanism is an aggregation of the functions of its individual parts. Moreover, the behavior of the entire mechanism is more than simply an aggregation of its individual parts. With increasing levels of a mechanism is increasing complexity of the behavior exhibited at each hierarchical level.

Describing phenomenal consciousness as an emergent property of a biological mechanism does not entail that phenomenal consciousness is a special phenomenon that cannot be explained by physical descriptions. On the contrary, the emergence of phenomenal consciousness, on this account, only requires that phenomenology should be examined at an independent level in addition to other levels at which it might be considered (Bechtel, 2008). Because phenomenal consciousness is the overall behavior of some mechanism or mechanisms, it occupies a certain amount of explanatory independence from the mechanism's behavior at other levels.

Thus, contrary to Churchland's (2002) assumption that "the direct and indirect approaches predict that reductive explanations will proceed stepwise from highest to lowest," explanations regarding consciousness, phenomenology in particular, will probably appeal to all levels within the mechanism. Because phenomenal consciousness is likely a behavior that emerges from an aggregate of other behaviors in the mechanism, phenomenal consciousness can be considered a part of the mechanism as a whole. For mechanistic explanation, phenomenal consciousness is a whole that is unlikely to be completely reduced to its lowest level of explanation.

Section V: Eliminating Consciousness?

The Churchlands expect that findings in consciousness studies will revise many of the common sense concepts that we currently posses about our experience. They often draw parallels between the misleading perceptions that we have about external events in the world with the misleading intuitions we have about various mental states. In particular, they assert that "*our common-sense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity*," [original emphasis] (Churchland, 1988, p. 43). Because of this, experimental findings in neuroscience and psychology will, at the very least, illuminate and deepen the intuitive concepts we have about the world. For many theories and conceptions in folk psychology, the Churchlands expect that scientific findings will overtake and completely eliminate some of the concepts we associate with our cognitive processes.

The Churchlands have often been accused of advocating the elimination of consciousness. It is likely that this accusation is at least partially due to their use of the phrase “eliminative materialism.” They consider the phrase “revisionary materialism” to be a more accurate description of their stance (Churchland & Churchland, 1996, p. 298). Although they expect many of our concepts about conscious experience to be changed by advances in consciousness studies, they have no strong ideological stake in the amount of revision, stating that it is “an empirical question how much revision a theory and its concepts will undergo...” (Churchland & Churchland, 1996, p. 298). Despite the use of the term “eliminative,” the complete elimination of all concepts related to consciousness seems unlikely. They expect that the concepts we bring to our experience of the world will likely be revised, to some extent, by scientific discoveries. If this means the complete elimination of our concepts regarding consciousness, then this is just a side effect of scientific advancement.

The Churchlands do expect our current concepts about conscious experience to change in some way, but this fact should not be confused with expecting the elimination of consciousness itself. They do not deny that consciousness exists. Consciousness, as we currently understand it, will not cease to be a phenomenon just because our theories about it change. Our understanding about consciousness may change, but this does not necessitate that consciousness itself will change.

Mechanistic analysis is consistent with the possibility that our understanding of phenomenal consciousness will change based upon our understanding of the mechanisms that generate the conscious experience. It is also consistent with the Churchlands' assertion that phenomenal consciousness itself will not cease to exist if our conceptions about phenomenal consciousness change. Mechanistic analysis will create a richer understanding of conscious experience than a strictly reductionistic viewpoint. Because the levels of organization will be kept intact in an explanation of the mechanisms responsible for the generation of the conscious experience, there will be a greater understanding of the relationship between the mechanism's levels and the conscious experience itself.

Chapter 3: Multidisciplinary Theories for Multidisciplinary Sciences

I have shown that mechanistic analysis is a foundational tool that will inform how consciousness is investigated scientifically. Moreover, the tools used in mechanistic analysis will frame how we think about consciousness. The formulation of scientific theories is heavily influenced by the conceptual frameworks that scientists bring to investigations. The development of evolutionary theory is an obvious example of one of the most far-reaching paradigm shifts that changed many areas of research. Scientific ideas before and after the advent of evolutionary theory possessed such radically different frameworks that, according to some, the theories were incommensurable (or, incomparable) (Kuhn, 1970).

The main takeaway point is that thinking about a theory of consciousness with the background of mechanistic analysis instead of traditional reductionism will affect the theory's construction. Individual levels of investigation will play a more prominent role in theories using mechanistic analysis compared to theories using reductionism. I argue that since the scientific disciplines contributing to a theory of consciousness—psychology and neuroscience—use mechanisms instead of laws, a theory of consciousness will be multidisciplinary.

In Section I, I describe the natural method as the appropriate model for a science of consciousness. Section II addresses the potential concern that a multidisciplinary theory of consciousness is problematic because there are tenuous connections between theoretical frameworks in different scientific disciplines. I further elaborate the specific requirements for a multidisciplinary theory of consciousness in Sections III and IV.

Section I: The Natural Method

The relationship between psychology and neuroscience has its own set of problems, but the methodology for studying consciousness is also potentially problematic. Thus far, I have been writing under the assumption that consciousness does not, in fact, present a problem for scientific inquiry. In the previous chapter, the direct and the indirect methods for studying consciousness as outlined by Churchland (2002) are examples of ways to study consciousness scientifically. Of course, there are very few uncontested claims, and the claim that consciousness is not problematic for science is certainly not one of them. On the contrary, only recently has consciousness been deemed a subject matter for scientific inquiry. Prior to technological developments in cognitive science in the 1980s, consciousness was regarded as too subjective to be studied using scientific methods. Thus, prior to this change, consciousness had been a topic only for philosophical treatment.

The phenomenal feature of consciousness, considered by many to be the definitive marker of consciousness, was the biggest obstacle to science. The subjectivity of consciousness was thought to be in tension with the objectivity of scientific methodology. Eventually, cognitivism became dominant and scientists realized that the methods developed in psychology could be used to illuminate our understanding of consciousness. When technologies in neuroscience began to develop, researchers realized that the examination of the brain could illuminate findings in experimental psychology.[11] Both disciplines could, for instance, work together to find an explanation for people that show deficits in conscious experience. Lesion studies, for example, help to correlate deficits in the verbally reported experience of experimental participants with deficits in neural function. With the technological advances of the

[11] This discussion of the beginning of consciousness' inclusion in scientific methods is by no means comprehensive. The overly simplistic view of the development of consciousness studies primarily serves to highlight a change in attitudes towards the compatibility of consciousness and science.

biological and psychological sciences and changes in the characterization of consciousness as incompatible with science, consciousness began to be studied by scientific methods.

At this time, many philosophers expressed metaphysical, epistemological, and methodological reservations regarding the inclusion of phenomenal experience in science. Science has many metaphysical assumptions that were considered problematic for the examination of conscious experience. A common, though not universally held, view of science presumes that everything studied using scientific methods is physical. For practical purposes, researchers must assume that everything in the domain of scientific inquiry is physical. This presumption, when extended to all objects in the universe, is known as physicalism. Even though consciousness is a byproduct of biological processes, consciousness as such seems decidedly distinct from physical stuff.

Thomas Nagel was at the forefront of this debate. He asserted that a purely physicalist approach to consciousness could not accommodate phenomenology (1974). Nagel argues that humans cannot take the perspective of another animal, such as a bat, because our physical constitution differs in significant ways. These biological differences affect consciousness in such a way that it is difficult to imagine the way the world seems to another; we cannot imagine *what it is like to be* another species (Nagel, 1974). This *what it is like* aspect of consciousness is phenomenal experience. Nagel critiqued the physicalist approach to consciousness because it did not take into account phenomenal experience.

When this critique is extended, the physicalist assumptions of science, and science itself, seem to be incompatible with consciousness. A common argument in philosophical literature is that the best physical theories of consciousness will still leave out essential features of phenomenal consciousness (Jackson, 1986). According to this argument, even the most comprehensive physical description of phenomenal experience will not be able to completely describe the subjective features of phenomenal experience. Since phenomenal experience is often considered to be the defining feature of consciousness, the best physical theory of consciousness will always be inadequate.

After Nagel's critique of materialist theories of consciousness, a rash of philosophical scholarship emerged. Some philosophers conceded Nagel's claim entirely and began to look for alternatives to physicalism. Other philosophers continued to believe that the interaction between consciousness and scientific methods was promising, and embraced the union of the two (Dennett, 1992). Alternative positions emerged as well. Colin McGinn, in particular, began a new wave of philosophical positions on consciousness. McGinn did not deny that consciousness was compatible with a physicalist metaphysics. He did, however, assert that it appeared unlikely that scientific methods would lead to a richer understanding of consciousness (McGinn, 1991). McGinn claimed that we are "cognitively closed" to a complete understanding of conscious experience. In other words, according to McGinn, our biological constitution prevents us from fully understanding our mental faculties.

This abridged version of a few reactions to the consciousness-science interface highlights the significance of what Owen Flanagan terms the "natural method." In response to scholars', such as McGinn's, reluctance to embrace the potential for phenomenal consciousness to be examined using scientific methods, Owen Flanagan articulated "the natural method" (1992). Flanagan carved out a middle ground between the extreme eliminativist and extreme dualist positions regarding conscious experience. Flanagan highlights, without overstating, the importance of phenomenology, stating that, "phenomenology alone never reveals anything at all about the mental events and processes involved in conscious mental life…but it is incredible to think that we could do without phenomenology altogether," (1992, p. 12). The natural method is Flanagan's attempt to include descriptions of phenomenal consciousness along with studies in neuroscience and psychology.

Since Flanagan developed the natural method in response to arguments about phenomenal consciousness, it has contentious premises. Three assumptions underlie Flanagan's description of the natural method: that mental activity is a direct result of neural activity, that consciousness is not an *a priori* problem for physicalism, and that the scientific analysis of consciousness does not rest upon a requirement for a unified theory of consciousness.

The first assumption is that there are no metaphysical barriers between the mind and the brain. A science of the mind, consciousness in particular, assumes a direct relationship between brain activities and mental activities. As previously stated, consciousness studies presumes a physicalist conception of the universe. Physicalism does not have the associated problems of substance dualism, as espoused by Descartes, which states that both mental substance and physical substance exist in the universe. The fundamental problem with Descartes' formulation of substance dualism is that it is unclear how the mind and brain interact with each other (Bechtel, 1988). Because only one substance exists, physicalism does not have this problem of interaction between the mind and the brain. Changes in brain states should correspond to changes in mental states, and changes in mental states should correspond to changes in brain states. The purpose of the scientific study of the mind is to determine precisely in what manner brain functions affect mental functions.

The second assumption of the natural method, and perhaps the most important, is that phenomenal consciousness is not an *a priori* problem for physicalism. Phenomenology is simply one aspect of consciousness. Since consciousness is a product of the brain, neural functions need to be understood in order for consciousness to be understood. Phenomenology may be subjective, but this is not a reason to completely renounce a science of consciousness. The natural method approaches phenomenal experience as a natural, not intractable, fact of the universe. Phenomenology needs to be studied alongside the mind/brain sciences so that a complete picture of consciousness can emerge. According to Flanagan, "the phenomenal features are only part of the story," and the other part of the story will be completed through the scientific investigation of both the mind and the brain (1995, p. 1114).

The third assumption of the natural method is that we do not need to expect a unified theory of consciousness before we begin to investigate conscious experience. This expectation by Flanagan corresponds to the methods currently in practice in consciousness studies, and it is compatible with what Churchland describes as the direct and indirect methods. Flanagan expects that, "…a theory of mind, and the role conscious mentation plays in it, will need to be built domain-by-domain with no *a priori* expectation that there will be a unified account of the causal role or evolutionary history of different domains and competences…" (1995, p. 1104). Neither the direct nor the indirect method requires a theory about consciousness to be developed before research into consciousness occurs.

Mechanistic analysis, rather than strict reductionism, is best to use with the natural method. The parts of the mechanism will not be settled before the mechanism has been examined scientifically. A theory of consciousness will describe the functions of both the individual parts of the mechanism as well as the function of the mechanism as a whole. Additionally, mechanisms are compatible with the presumption of a materialist metaphysics. In the previous chapter, I suggested how consciousness can emerge from physical, neural mechanisms. The emergence of consciousness need not demand an additional metaphysical substance. With mechanistic analysis, consciousness and physicalism are compatible. Overall, we can expect a theory of consciousness to be formed using the natural method, and with a background of mechanistic analysis.

The natural method suggests the necessity of an interdisciplinary theory of consciousness. A virtue of the natural method is that it incorporates data from a range of areas. Phenomenology will contribute to a theory of consciousness because it is the starting point for any scientific inquiry about consciousness. A theory of consciousness will need to explain the features of phenomenal experience, and data from the sciences should correspond to (or be used to explain any irregularities in) phenomenal experience. Phenomenal experience needs to be examined in order to correlate the functions of the individual parts of the mechanism with the function of the mechanism as a whole; the interactions of the mechanism's individual parts contribute to the phenomenal experience resulting from the mechanism's overall function. In the field of psychology and in accordance with the natural method, psychological events are correlated with systems level neural activity. In this manner, cognitive psychology is integral to mechanistic analysis; psychological studies initially help mechanistic analysis by determining a mechanism's function. Systems level neuroscience can help determine the mechanism's location. Then, other levels of analysis in neuroscience, such as cellular and molecular neuroscience, help to determine the mechanism's parts. In this manner, both psychology and neuroscience are integral to mechanistic analysis. In accordance with both the natural method and mechanistic analysis, an analysis of phenomenal experience is also necessary. A theory of consciousness will explain how the mechanism of a conscious experience gives rise to the conscious experience. Thus, a theory of consciousness will need to incorporate whichever levels of analysis help generate the conscious experience.

Section II: A Problem for Interdisciplinary Theories

The very reason that it seems unlikely for sciences to unify by reduction is a potential barrier to the creation of a unified interdisciplinary theory. The connections between the scientific disciplines, neuroscience and psychology especially, do not perfectly match up to each other. This imperfect correlation of scientific domains is one of the primary reasons that higher level sciences do not seem reducible to lower level sciences[12]. In addition to creating problems for reductionism, the imperfect correlation between sciences is a barrier to the formation of interdisciplinary theories. Independent sciences have different frameworks that are not easily translatable to each other. However, the notion of a bridge science, much like bridge laws, can serve to connect conceptual differences between scientific disciplines.

Reductionism is problematic for the psychology-neuroscience interface precisely because it does not allow for the independence of multiple levels of analysis. The independence of multiple levels of analysis is crucial for the biological and mental sciences because they use mechanistic analysis instead of laws. The specific reason that reductionism fails needs to be outlined in order to understand the similar problems that an interdisciplinary theory must also face. The most relevant reason is that the predicates of the special sciences do not obviously correlate with the predicates of the lower level physical sciences.

Jerry Fodor explains the difficulties in identifying bridge laws between scientific disciplines (1974). Reductionism and physicalism are intimately tied together. If everything in

[12] It is important to keep in mind that this imperfect correlation is a reason that some may advocate for the elimination of the higher level. The elimination of the higher level theory may be possible in some rare cases. In the previous chapter, I explained the benefits of mechanistic analysis over reductionism. Since mechanistic analysis emphasizes unifying different levels of analysis and eliminativism seeks to get rid of certain levels of analysis, I suspect that mechanistic analysis is a more reasonable expectation than eliminativism.

the universe is physical, then everything in the universe also follows physical laws. The most general science, physics, studies the fundamental laws that govern everything physical—which, under a physicalist conception of the universe is literally everything. The fundamental premise of reductionism is that everything is reducible to the physical laws predicated in physics. For this reason, it is presumed that all phenomena studied in the scientific disciplines that are not physics are "special cases" of more general physical laws. Following this, scientific disciplines such as chemistry, biology, and psychology, are just special cases of physics—hence Fodor's term "special sciences" (1974).

Fodor asserts that the predicates of a special science must match up with the predicates of a more general science. In order for this to occur, the events in one science must be identical to the events of another science. This identity requirement is the crux of Fodor's conception of a bridge law (1974). Reduction needs identity between the predicates in different sciences in order to ensure that the events refer to the same thing. In order for reductionism to occur, special science predicates need to be identical to (through bridge laws) physical science predicates. Special science predicates, however, are so general that they do not seem identical to predicates in the physical sciences predicates. It is unlikely that multiple instances of predicates in the special sciences will correlate with only one predicate in a physical science. Because there is no one to one identity between special science predicates and physical predicates, special science predicates will not neatly correspond with physical science predicates. Thus, according to Fodor, the special sciences are immune to reductive physicalism.

The connection between scientific disciplines is also problematic for a non-reductive multidisciplinary theory. In accordance with the natural method, an interdisciplinary theory of consciousness will take into account phenomenology, psychology, and neuroscience. Although the natural method appears to be simple, it is unclear how experimental findings in different scientific disciplines fit together. As succinctly stated by Valerie Hardcastle, "…our investigation leaves us with lots of messy details lumped together from different levels of analysis, different theoretical frameworks, and different investigative questions. The connections at times seem tenuous at best" (1996, pp. 141-2). Psychology and neuroscience may be investigating similar phenomena, but they differ in the manner in which they approach investigation. These disciplines have different aims and goals for investigation. Additionally, the scientific disciplines differ with respect to the conceptual tools they use to investigate a particular phenomenon. The differences in terminology between disciplines reflect these conceptual differences between disciplines.

Neuroscience and psychology have different historical backgrounds. Both disciplines investigate similar phenomena, but they have different methods and attempt to ask different questions. Cognitive psychology and cellular neuroscience, for example, both investigate attention, but they use different experimental methods and hypotheses. Cognitive psychology might measure attentional changes at the level of cognitive changes of an experimental participant while cellular neuroscience might try to figure out the attentional changes at the level of individual cells within a single attentional system. In this instance, each study would have different aims, and the results of each may not resemble each other. The data found in the cellular neuroscience study may be so detailed that there is little resemblance to the concepts and terminology present in cognitive psychology. Although the same phenomenon, attention, is being investigated, it seems unclear how the results from one discipline neatly relate to the results from the other discipline.

Section III: “Two-Part” Interdisciplinary Theory

While keeping in mind these concerns, it is much too simplistic to conceive of a definitive separation between sciences such as cognitive neuroscience and cognitive psychology. In fact, there is a considerable amount of overlap in the theoretical assumptions and methods used in psychology and neuroscience. Both disciplines generally have similar operational definitions regarding attention even though the way in which attention is functionally defined might differ between each discipline. Additionally, both disciplines share broad assumptions about the anatomy of the brain. Cognitive neuroscience is a hybrid of psychological and neuroscientific techniques; there are many similarities between it and cognitive psychology, such as the particular cognitive phenomena that are investigated and the use of lesion studies.

In short, the theoretical and technological connections between psychology and neuroscience are a starting point for connecting these disciplines together. The notion of a bridge science, proposed by Hardcastle, is analogous to the idea of bridge laws in traditional proposals of scientific reduction. Bridge sciences serve as a connection between disparate sciences. She states:

> In brief, a bridge science maintains the proper connections by adopting the background assumptions of both psychology and neuroscience (or whatever two sciences), and the important questions from one of these domains, while also maintaining its own contrast class and relevance relations distinct from either of them (Hardcastle, 1996, p. 143).

Hardcastle’s discussion of “contrast classes” and “relevance relations” borrows from philosophical literature on erotetic models of explanation (1996, p.143). Evidenced by the word “erotetic,” this model of explanation claims that explanations answer why-questions (Bromberger, 1966; van Fraassen, 1980). Scientific explanations, in this view, are also answers to why-questions. An explanation for the green color of grass, for example, would answer the question “Why is grass green?” The explanation would go into detail about the features of the chlorophyll present in grass cells, but this answer would be an answer to that simple why-question.

Finding the correct answer to a question requires the narrowing of the number of possible answers. Assume that the possible answers to the question “Why is grass green?” is represented by the set {Q, R, S, T, U, V}, and the answer to the question can only be one member of the set. After testing each explanation within this set, S is deemed the best explanation and the other members {Q, R, T, U, V} are discarded. {Q, R, T, U, V} are then considered contrast classes. When explanations are considered to be answers to why-questions, the contrast class of an explanation is the set of alternative answers for that particular why-question.

According to Hardcastle, psychological theories can only be reduced to neuroscientific theories if both their relevant and contrast class match. The potential set of answers relevant for an explanation is determined by the historical background of the discipline in which the question is raised (Hardcastle, 1996). It follows that the contrast classes of a theory are also determined by the discipline’s historical background. As I have previously stated, Fodor argued that reductionism needs the correspondence between predicates in higher-level theories and predicates in lower-level theories. All of the predicates involved, including the contrast class, need to be present in the theories of both disciplines in order for reductionism to be viable.

Bridge sciences are unique in that they share theoretical assumptions that are present in the two disciplines that it aims to connect. Bridge sciences cannot be reduced to the other two connecting sciences, however; the bridge science is autonomous from its connecting sciences. Bridge sciences are autonomous because they have their own contrast classes and theoretical assumptions that are not found in the two connecting sciences, and thus, cannot be reduced to either of the connecting sciences.

Hardcastle uses event-related potentials (ERP) as an example of a bridge science. ERP measure the electrophysiological response of neural activity related to a particular behavioral, cognitive, or sensory event. ERP meet Hardcastle's expectation for a bridge science; ERP research shares the theoretical assumptions of psychology and neuroscience while retaining its own assumptions. On the one hand, since ERP is used in both neuroscience and psychology, some theoretical assumptions present in neuroscience and psychology are also present in ERP. Hardcastle explains that ERP research adopts many questions in psychology and connects the answer to neural activity. On the other hand, since ERP research measures the electrical properties of neurons, ERP research has theoretical assumptions that are independent of psychology and neuroscience. ERP research is not the only bridge science, however. Hardcastle also explains that computational neuroscience serves as a bridge science because it connects computer science to neuroscience (1996).

The interdisciplinary nature of consciousness studies is preserved through the use of bridge sciences, making bridge sciences compatible with mechanistic analysis. Bridge sciences are irreducible to their connecting sciences due to the theoretical assumptions specific to a particular bridge science. The autonomy of a bridge science is incompatible with reductionism, but it is compatible with mechanistic analysis. A virtue of mechanistic analysis is that the levels within a mechanism retain a degree of independence. The function of the mechanism depends upon the organizational structure of its parts, and specifically, the connection of those parts to each other. It is plausible to expect that some of the levels within a mechanism of consciousness will correspond to a bridge science. For instance, the highest level of a mechanism of consciousness may be at the level of cognitive psychology, ERP research could be an intermediary level, cellular neuroscience could be at an even lower level, and molecular neuroscience could be at an even lower level. Although this proposal is simply hypothetical, it is clear that the notion of a bridge science unifies levels of scientific inquiry, which contributes to the expectation of a unified, yet interdisciplinary, theory of conscious experience.

Section IV: Criteria for an Interdisciplinary Theory of Consciousness

Bridge sciences are important for multidisciplinary theories because they serve as a unifying principle for disparate concepts in different disciplines. The bridge science provides general principles to unify the sciences. In addition to bridge sciences, Hardcastle asserts that an interdisciplinary theory will need to contain both “general principles” and “specific models.” Specifically, she states:

> On the one hand, there will be more or less a laundry list of general principles governing the phenomena across several higher mammals (or whatever). These principles apply to all physical systems for which we have theoretical models of the phenomena…On the other hand, the theory will also include the models of specific animal systems from which we derived the general principles. These models flesh out our general principles—we use them to make our principles concrete and to guide predictions for future experiments (p. 133, Hardcastle, 1996, p. 133).

The general principles allow the interdisciplinary theory to be applicable to more than one organism. A few of the general principles in the study of human consciousness would include: the gross similarity of neural structure, brain organization, brain function, psychological features, and perceptual features across billions of different people. All of the commonalities of conscious experience would be listed among the general principles included in a theory consciousness.

Models for specific systems will be able to account for deficits in consciousness. Blindsight patients, for example, have deficits in phenomenal visual experience due to neurological damage (Weiskrantz, 1990). A theory of consciousness will be able to be tailored to explain the relationship between the damaged area and the associated psychological deficits. In addition, altered states of consciousness (either due to biological or artificial causes, such as drugs) would be included in the models of specific systems. In essence, any atypical feature of consciousness would be incorporated into a theory of consciousness through the addition of specific models.

The theoretical commonalities between psychology and neuroscience that are located in the bridge sciences fulfill the requirement for the general principles needed in an interdisciplinary theory. This generality is important to increase the range of phenomena and domains that can be used in a theory of consciousness. Although the current discussion has primarily focused on a theory of consciousness for humans, general enough principles could allow a theory of consciousness to be applied to non-human animals.

Epilogue: Beyond Philosophy, Neuroscience, and Psychology

Thus far, this entire discussion has been limited to neuroscience and psychology, but cognitive science involves more than just these two disciplines. Linguistics, anthropology, and artificial intelligence all play prominent roles in cognitive science. Evolutionary considerations in anthropology, biology, and psychology will extend the domain of consciousness beyond humans. Artificial intelligence will extend the domain of a theory of consciousness beyond biological organisms.

Hardcastle's expectation for an interdisciplinary theory includes, on the one hand, "general principles governing the phenomena across several higher mammals" and, on the other hand, "models of specific animal systems" (1996, p.133). In this way, an interdisciplinary theory of consciousness with evolutionary considerations can be applied beyond human physiology. Comparative neuroscience and psychology can give considerable insight into the mental life of other organisms, such as a cat or a rat (Churchland, 1986). The background of evolutionary theory supports the claim that humans are probably not the only organisms with conscious experience. Since species gradually evolve by means of evolutionary processes, it is extremely likely that there are evolutionary precursors to human conscious experience in other animals.

Many experimental studies in neuroscience and psychology cannot be performed with human subjects due to ethical considerations. Lesion studies are vital to neuroscience. These types of studies can only be tested on humans when the lesions are already present through accidental causes, but lesions can be induced in other organisms. The functions of different parts of the brain are often discovered after a particular area has been compromised, and lesion studies are a method of determining what general areas are involved in generating a particular function[13]. Since these lesions cannot be induced in humans, they have to be modeled in other organisms such as non-human primates and mice. Experimental results from non-human animals are then extended to humans. Comparative anatomy can be done in this manner because there are similarities between the gross anatomy of humans and other animals. These similarities in neural anatomy often are reflective of physiological similarities between different species.

A theory of consciousness that includes studies in evolutionary anthropology and evolutionary biology will be independently testable with regard to other organisms. Brodmann areas are a cytoarchitectonic map of distinct neural areas in the brains. Maps of Brodmann areas have been developed for a wide variety of species, including human, cat, rat, dog, and numerous non-human primate species (Bechtel & McCauley, 1999). Brodmann found an astonishing number of similar areas across different species. In accordance with evolutionary theory, primate neural anatomy is especially similar to human neural anatomy. Thus, it is to be expected that there will be similar conscious experiences between humans and other primates.

In fact, many studies of conscious experience are already being performed on non-human primates. In this vein, consciousness studies already carries evolutionary assumptions about the development of both neuroanatomy and cognitive functions. To reference just one of many examples, neuropsychologists commonly use monkeys to study the neural correlates of visual awareness (Kanwisher, 2001). Experimental psychologists are able to train monkeys for studies of psychological phenomena such as binocular rivalry. In binocular rivalry tasks, two competing images are presented to the experimental participant, one to each eye. Conscious perception alternates randomly between these two competing images. Monkeys are trained to press two

[13] Lesion studies alone cannot tell us whether an area is either necessary or sufficient for a particular psychological function. Lesion studies can, at most, tell us if a particular location in the brain is necessary for the psychological function.

different buttons to report changes in their conscious experience. Researchers are able to correlate changes in neural activity with the monkeys' self-reported change in their visual experience. With some restrictions, the results of these studies on monkey are able to inform us about the relationship between neural activity and visual conscious experience in humans. It is not the case that humans and other primates, particularly the great apes, share all of the same features of consciousness. Regardless, the evolutionary trajectory of consciousness will be traceable through experiments on other animals[14].

In a similar vein, a theory of consciousness might also have explanatory value beyond the organisms that currently exist. Computer science is involved in cognitive science, and the field of artificial intelligence is especially concerned with modeling cognition in artificial systems. Currently, the cognitive acts that involve consciousness in humans, such as memory, are being modeled in robotic systems. So far, these acts of cognition have been modeled in robots without involving conscious experience. Barring any metaphysical barriers, once a theory of consciousness has been fully fleshed out, it may be possible to model this important aspect of human cognition in artificial systems. Arguably, consciousness is what differentiates biological, and specifically human, intelligence, from artificial intelligence. The current systems that computer scientists have already built are computationally powerful, but so far they have not been able to match all aspects of human cognition. A theory of consciousness will aid our understanding of consciousness so that we may be able to implement conscious activities into non-biological systems. If this were to happen, machine intelligence should vastly improve.

Many of the above claims related to artificial intelligence are wildly speculative, but these hypotheses serve to illustrate the benefits that an interdisciplinary theory of consciousness

[14] Similarly, the developmental trajectory of consciousness in humans, from infancy to adulthood, could be determined through the use of experiments in developmental psychology.

might have on research possibilities in the cognitive sciences. Consciousness is one of the most intriguing yet most problematic features of biological life. With the combined effort of the disciplines at the center of consciousness studies, we are closer than ever—though still very far from—being able to explain consciousness. Our ability to unravel conscious experience will progress for the foreseeable future.

References

Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4* (4), 292-309.

Bechtel, W. (1988). *Philosophy of mind: An overview for cognitive science*. Lawrence Erlbaum.

Bechtel, W. (1994). Levels of description and explanation in cognitive science. *Minds and Machines*, *4*(1), 1-25.

Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Lawrence Erlbaum.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421-441.

Bechtel, W., & McCauley, R. N. (1999). Heuristic identity theory (or back to the future): The mind-body problem against the background of research strategies in cognitive neuroscience. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 67-72). Mahwah, NJ: Erlbaum.

Bromberger, S. (1966). Questions. *The Journal of Philosophy*, *63*(20), 597-606.

Campbell, K. (1986). Can intuitive psychology survive the growth of neuroscience?. *Inquiry*, *29*(1-4), 143-152.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press, USA.

Churchland, P. M. (1986). Some reductive strategies in cognitive neurobiology.*Mind*, *95*(379), 279-309.

Churchland, P. M. (1988). *Matter and Consciousness, Revised Edition: A Contemporary Introduction to the Philosophy of Mind*. MIT press.

Churchland, P. S. (1983). Consciousness: The transmutation of a concept. *Pacific Philosophical Quarterly*, *64* (1983), 80-95.

Churchland, P. S. (1994). Can neurobiology teach us anything about consciousness?. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 67, No. 4, pp. 23-40). American Philosophical Association.

Churchland, P. S. (2002). *Brain-wise: Studies in Neurophilosophy*. MIT press.

Churchland, P. S., & Churchland, P. M. (1996). *The Churchlands and their critics*. R. N. McCauley (Ed.). Cambridge, MA: Blackwell.

Craver, C. F. (2007). *Explaining the brain*. Oxford: Oxford University Press.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences* (Vol. 2, No. 263-275, p. 203).

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of science*, *75*(5), 958-969.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 43*, 44-64.

Dennett, D. C. (1993). *Consciousness explained*. ePenguin.

Flanagan, O. J. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT press.

Flanagan, O. J. (1995). Consciousness and the natural method. *Neuropsychologia*, *33*(9), 1103-1115.

Flanagan, O. J. (1995). Deconstructing dreams: The spandrels of sleep. *The Journal of philosophy*, 5-27.

Fodor, J. A. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese, 28*(2), 97-115.

Fodor, J. A. 1997. “Special Sciences: Still Autonomous After All These Years.” *Philosophical Perspectives*, *11*, 149-163.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*(S3), S342-S353.

Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*(2), 362-381.

Hardcastle, V. G. (1992). Reduction, explanatory extension, and the mind/brain sciences. *Philosophy of Science*, 408-428.

Hardcastle, V. G. (1996). *How to build a theory in cognitive science*. SUNY Press.

Hume, D. (2001). *An enquiry concerning human understanding* (Vol. 3). Oxford University Press, USA.

Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy*, 291-295.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*,*79*(1), 89-113.

Kitcher, P. (1982). *Abusing Science: The Case Against Creationism*. MIT Press: Cambridge, MA.

Koch, Christof (2004) *The quest for consciousness: a neurobiological approach,* Roberts & Company Publishers: Englewood, CO.

Kuhn, T. (1970). *The Structure of Scientific Revolutions.* Chicago and London.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. *Criticism and the Growth of Knowledge*, *4*, 91-195.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science,* 1-25.

McCauley, R. N. (2007). Enriching Philosophical Models of Cross-Scientific Relations: Incorporating Diachronic Theories. *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience and Reduction*. M. Schouten and H. Looren de Jong (eds.). Oxford: Blackwell Publishers, 199-223.

McGinn, C. (1989). Can we solve the mind–body problem?. *Mind*, *98*(391), 349-366.

McGinn, C. (1991). *The problem of consciousness: Essays towards a resolution*. B. Blackwell.

Nagel, E. (1979). *The Structure of Science: Problems in the Logic of Scientific Explanantion.* Hackett Publishing Company Incorporated.

Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, *83*(4), 435-450.

Oppenheim, P. and Putnam, H. (1958). "Unity of Science as a Working Hypothesis," *Minnesota Studies in the Philosophy of Science--Volume II.* H. Feigl, M. Scriven, and G. Maxwell (eds.). Minneapolis: University of Minnesota Press, 405-428.

Pajunen, K. (2008). The nature of organizational mechanisms. *Organization Studies*, *29*(11), 1449-1468.

Popper, K. (2002). *The logic of scientific discovery* (No. 117). Routledge.

Popper, K. (1963). Conjectures and Refutations: The Growth of Scientific Knowledge: Harper Torchbooks: New York, NY.

Radner, D., & Radner, M. (1989). *Animal consciousness*. Buffalo: Prometheus Books.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*(3), 329-359.

Rosenthal, D. M. (2002). Explaining consciousness. *Philosophy of mind: Classical and contemporary readings*, 406-21.

Salmon, W. C. (1984, January). Scientific explanation: Three basic conceptions. In *PSA: Proceedings of the biennial meeting of the philosophy of science association*. Philosophy of Science Association, 293-305.

Schouten, M. K., & De Jong, H. L. (1999). Reduction, elimination, and levels: The case of the LTP-learning link. *Philosophical Psychology*, *12*(3), 237-262.

Searle, J. R. (1992). *The rediscovery of the mind*. MIT press.

Secada, J. (2000). *Cartesian metaphysics: the scholastic origins of modern philosophy*. Cambridge University Press.

Sklar, L. (1967). Types of inter-theoretic reduction. *British Journal for the Philosophy of Science*, 109-124.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press, USA.

Weiskrantz, L. (1990). *Blindsight: A case study and its implications* (Vol. 12). Oxford University Press, USA.

Wright, C., & Bechtel, W. (2006). Mechanisms and psychological explanation. In: *Philosophy of Psychology and Cognitive Science: A Volume of the Handbook of the Philosophy of Science Series*. Eds. Gabbay, D. M., Woods, J., & Thagard, P. Elsevier. pp. 31-79.