

ABSTRACT

The Zombie Argument and Its Implications for Naturalism

Autumn S. Averitt

Director: Todd Buras, Ph.D.

In the discipline of philosophy today, the nature of consciousness is an extremely popular topic. Additionally, there is a general consensus on the truth of naturalism. This thesis investigates whether a naturalist conception of consciousness is tenable. The zombie argument against the naturalist program is very striking, as it clearly delineates the apparent differences between the phenomenal and psychological concepts of the mind. This thesis explores the naturalist program and the nature of consciousness. It then outlines the zombie argument and evaluates naturalist responses to the argument. It concludes that the strongest naturalist response still falls short of defeating the zombie argument. Naturalism therefore fails to adequately account for consciousness.

APPROVED BY DIRECTOR OF HONORS THESIS:

Dr. Todd Buras, Department of Philosophy

APPROVED BY THE HONORS PROGRAM:

Dr. Andrew Wisely, Director

DATE: _____

THE ZOMBIE ARGUMENT AND ITS IMPLICATIONS FOR NATURALISM

A Thesis Submitted to the Faculty of

Baylor University

In Partial Fulfillment of the Requirements for the

Honors Program

By

Autumn S. Averitt

Waco, TX

May 2014

TABLE OF CONTENTS

Chapter One: Naturalism and Its Constraints	1
Chapter Two: The Phenomenon of Consciousness	19
Chapter Three: The Zombie Argument	28
Chapter Four: Naturalist Responses to the Zombie Argument	45
Bibliography	60

CHAPTER ONE

Naturalism and its Constraints

Introduction

Within philosophy, discussion of the mind is dynamic and growing. Neuroscience is advancing rapidly, and discovering many new neural pathways, neurotransmitters, etc. These physical components of the brain seem readily investigable. Yet even as we use our minds constantly, they remain stubbornly mysterious to us. Philosophers therefore ponder questions about the inner life of the mind. These questions are more abstract and less easy to observe. A question that is currently in hot debate is that of consciousness. What is it, and how can it be explained? The reason this topic is so popular is that many philosophers today consider themselves naturalists. Considering oneself a naturalist entails certain things about the kinds of explanations to which one can appeal. One entailment is that reality need be explained in purely physical terms. This thesis investigates whether naturalists have an adequate answer to the question, “Is consciousness physical?” This thesis begins investigation with a parsing out of the relevant terms “naturalism” and “consciousness.” The thesis then takes a single argument against the physical nature of consciousness, the zombie argument, and examines it. The final chapter of the thesis studies whether any naturalist response to the argument is satisfactory. It concludes that all naturalist positions fail to adequately explain the phenomena of consciousness. Therefore it is likely that consciousness is non-physical, and naturalists must come up with new theories or give up their physical concepts of consciousness.

Definitions of Naturalism

The Vagueness of Naturalism

Naturalism is very intricate. Typical formulations regard naturalism as entailing a particular view encompassing the entire universe. However, this wide scope is not a necessary consequent of naturalism. It is possible for someone to refer properly to himself as a naturalist even if he restricts his naturalism to certain features of reality. A problem arises as various philosophers each understand naturalism to be differing sets of assertions about what applies to the basic make-up of reality. As these philosophers with differing positions refer to themselves as “naturalists,” the confusion multiplies and a proper definition of naturalism becomes very difficult. The first goal of this chapter is to clarify the vague topic of naturalism by discussing two relevant types of naturalism.¹ The two types are methodological naturalism and metaphysical naturalism. The types overlap to some extent, and so it is necessary to begin with a rough definition of each to avoid confusion. *Methodological naturalism* is the assertion that science is conducted properly only when it appeals to mechanistic explanations and avoids agential explanations. *Metaphysical naturalism* is the conjunction of the claims that (1) science, as it currently stands, has revealed a nearly complete picture of the universe, and (2) the complete picture of the universe will not differ much from the current picture in general character. What follows is an examination and refinement of these rough definitions. These refined definitions will then be instrumental in the later discussion of consciousness.

¹ These two types of naturalism are the types relevant to the discussion of consciousness, but Michael C. Rea suggests there may be more types in use, such as epistemological naturalism or linguistic naturalism (Rea 2002).

Methodological Naturalism

It is easiest to begin with an examination of methodological naturalism. This form of naturalism has a smaller scope and is therefore more easily understood. Furthermore, an understanding of metaphysical naturalism depends in part on an understanding of methodological naturalism. Methodological naturalism is an assertion about the norms of the procedures of science. Methodological naturalism holds that science is properly conducted when it refuses to appeal to purposes, intentions, agents, and similar concepts in its explanations. To fully explore the implications of eliminating such phenomena, this chapter will examine the position first by understanding the difference between mechanistic and purposive explanations, then understanding methodological naturalism as a methodology, understanding its roots in science, understanding the motivation for eliminating agential concepts, and finally conclude with reasons to distinguish between methodological naturalism and naturalistic thought with a broader scope.

It is important to note the difference between mechanistic and purposive explanations in order to understand methodological naturalism. C. Stephen Layman, referencing Richard Swinburne, makes this distinction clear. He states that methodological naturalists "think of science as an attempt to explain things in terms of natural as opposed to supernatural factors."² Natural factors consist of purely physical entities and the relations between them. Layman takes physical entities to be "composed entirely of entities whose only built-in tendencies are mechanistic."³ Mechanistic

² C. Stephen Layman, *Letters to Doubting Thomas: A Case for the Existence of God*, (New York: Oxford University Press, 2007), 19.

³ *Ibid.*, 17.

tendencies are most clearly understood as those processes that are reactions without choice or purpose to external stimuli. Mechanistic tendencies are either deterministic or probabilistic.⁴ Deterministic tendencies occur when all the physical facts of a previous time entail the physical facts of the time in question. Probabilistic tendencies occur when the physical facts entail that there are multiple ways for events to proceed, and that the way things proceed is determined solely by chance. Both mechanistic and deterministic tendencies lack purpose and intention; instead there is only the causal story.⁵

Michael C. Rea develops a clear picture of a methodology in his book *World Without Design*. He writes, "methodological assumptions generally are taken to be either background presuppositions about what the world is like that guide and constrain the process of inquiry, views about how inquiry should be conducted, or views about what sorts of inquiry are likely to be fruitful."⁶ If a methodology takes certain assumptions to influence inquiry in certain areas, it seems that methodological naturalism would take the assumptions of naturalism to influence certain kinds of inquiry. This, however, leads to the question of what the naturalistic assumptions would be. These constraints involve the empirical nature of science. Rea explains empiricism as being "standardly defined as the

⁴ Ibid.

⁵ I thank Dr. Alexander Pruss for reminding me that there are some naturalists who explain even purpose and intention in terms of natural processes such as evolution. For example, Fred Dretske addresses intentionality in his "A Recipe for Thought," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002). According to these philosophers, intention and purpose are composed of these mechanistic tendencies. Note, however, that it must be the case that the tendencies in and of themselves still lack intention and purpose.

⁶ Michael C. Rea, *World Without Design: The Ontological Consequences of Naturalism*, (Oxford: Clarendon Press, 2002), 64.

view that, at most, only analytic truths can be justified a priori."⁷ In other words, empiricism is the thesis that non-analytic truths must be justified through observation.⁸ Methodological naturalism therefore holds that the procedures of science are properly conducted when inquiry is dependent on evidence gained through observation rather than a priori reasoning.

Methodological naturalism rules out purposive scientific explanations in favor of a united form of explanation: one that only appeals to third-person observation. D.M. Armstrong addresses why this is appealing in his essay "The Nature of Mind." In this essay he explains the motivation behind methodological naturalism by citing the "authority of science,"⁹ or science as the source of reliable knowledge about contingent matters. For Armstrong, science is related only to the mechanistic physical, electrical, and chemical properties, and this includes accounts of humankind.¹⁰ He believes that the strength of science is demonstrated in humankind's ability to come to widespread intellectual consensus within science (particularly the hard sciences). The ability to "decid[e] disputed questions"¹¹ is provided by the *method* of science. He admits humanity's fallibility in interpretation within the discipline,¹² but still holds the sciences to be the greatest hope for discovering truth. The powerful evidence in favor of science

⁷ Ibid., 38.

⁸ Ibid., 39.

⁹ D.M. Armstrong, "The Nature of Mind" in *The Nature of Mind and Other Essays*, (New York: Cornell University Press, 1981), 2-4.

¹⁰ Ibid., 1.

¹¹ Ibid., 3.

¹² Ibid.

(the ground that we have covered and the consensus at which we have already arrived) leads Armstrong to ask "must we not therefore appeal to these relative certainties for guidance when we come to consider uncertainties elsewhere?"¹³ The success of science sheds some light on the way the scope of methodological naturalism is often (but not always) expanded from an account of the proper way to conduct science to an account of a proper account of everything.

Given the extensive evidence in favor of using methodological naturalism as the sole procedure in advancing knowledge, it seems difficult to imagine why anyone would restrict its domain to science alone. Alvin Plantinga, in his article "Methodological Naturalism?" argues that while methodological naturalism attempts to rule out bias in investigation, it ultimately fails. He argues that it does not have a global scope, or in other words, that it cannot be used accurately with regard to all parts of reality. He agrees with methodological naturalists that science is at its basis empirical, and states that "science is in some special way related to the deliverances of experience, in particular the deliverances of sense."¹⁴ However, he asserts that methodological naturalism excludes many other kinds of knowledge from areas in which these varieties are much needed. He argues that these kinds of knowledge cannot be excluded from explanations of certain phenomena, such as the knowledge of biblical truths in explaining altruistic Christians' behavior. Simply because an explanation is not empirically verifiable, Plantinga argues, does not mean that it cannot be used adequately for the phenomena at hand. *Any* agential

¹³ Ibid., 4.

¹⁴ Alvin Plantinga, "Methodological Naturalism?" in vol 1. of *Facets of Faith and Science*, ed. Jitse M. van der Meer (New York: University Press of America, Inc., 1996), 208.

explanations involving motivation or purpose at the fundamental level would be in conflict with methodological naturalism if methodological naturalism's scope is widened to a global domain. Plantinga thus concedes to what he calls a "Duhemian science"¹⁵ that is neutral on topics not obviously within the domain of empiricism. This might be regarded as the *bare* methodological naturalism, one that does not expand its scope beyond what is necessary. Thus methodological naturalism makes no assertions as to the adequacy of agential explanation in other areas, but avoids such explanations only within the domain of science. The reasons for this are practical: the greatest number of scientists working in conjunction will be likely to produce truth more quickly than only a few scientists will. The easiest method to obtaining unity among scientists is to eliminate conflict, and avoiding assertions about agential explanations in dubious areas is conducive to such a process. The ideals of Plantinga's "Duhemian science" are a clear guide to methodological naturalism, and will help in reflecting on the distinctions between methodological naturalism and metaphysical naturalism with its wider scope.

Thus methodological naturalism has two central tenets: (1) science is reliable for understanding contingent matters and (2) mechanistic explanation is ideal within science. These views may give way to views about a larger domain than just the sciences, but it is not necessary that they do so. When philosophers do enlarge the scope of methodological naturalism, it is due to their commitment to use a single methodology for learning about reality. As Armstrong pointed out, the evidence leads many to take science as the best candidate for such a methodology. When the domain of

¹⁵ Ibid., 211.

methodological naturalism is expanded to have a global scope, or to encompass all of reality, then the position leads to metaphysical naturalism.

Metaphysical Naturalism

Just as an understanding of methodology helps with an understanding of methodological naturalism, so an understanding of metaphysics helps with an understanding of metaphysical naturalism. Metaphysics itself is very hard to define. For the purposes of this thesis, it is sufficient to consider metaphysics as the study of the ultimate truths of reality.¹⁶ Roughly, theories resulting from the study of metaphysics are ontologies, accounts of the nature of being, which are comprehensive and aspire to include all of reality.¹⁷ This definition is not completely satisfactory because there are philosophers whose metaphysical theory is to deny that anything exists. Nevertheless, the idea of general ultimate truths will get the point across well enough. Similarly metaphysical naturalism is a comprehensive claim about what is true of reality: roughly that methodologically naturalistic science is the only tool to discover significant truths about any aspect of reality.

Many philosophers consider themselves metaphysical naturalists, but what they consider metaphysical naturalism to be often varies. The landscape of metaphysical naturalism becomes even more varied as metaphysical non-naturalists offer their own definitions. Some of the confusion is a result of the connotations of “naturalism,” but other sources of confusion are the varying commitments to methodological naturalism

¹⁶ See qt. Peter Van Inwagen in “Metaphysics,” in *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/metaphysics/#NatMet>

¹⁷ Ibid.

and what these commitments imply about metaphysics. It is very appealing to be able to streamline the investigation of reality into a single method. This method would then be the default plan of attack for understanding anything in the universe. Science is within humankind's capabilities and has shown much progress, and therefore is one of the most appealing options for streamlined investigation. Michael C. Rea thus dubs naturalism a "research program."¹⁸ Someone who views such a program as the only method to gaining truth would be a "full-blooded" metaphysical naturalist.¹⁹ Full-blooded metaphysical naturalists view methodological naturalism as having a global domain. Other philosophers are not convinced methodological naturalism has such a wide domain, and hold that there are other methods of investigation necessary to explaining parts of reality. Therefore these naturalists are qualified metaphysical naturalists or even metaphysical non-naturalists. Such philosophers might be similar to Alvin Plantinga or Pierre Duhem, accepting methodological naturalism within a restricted domain and expecting agential explanations to hold in other domains of reality. Clearly there is a wide variety of types of metaphysical naturalism, with some philosophers applying it to wider domains while others restrict its application. In order to understand the relevant distinctions along the metaphysical naturalism spectrum, it is necessary to have a full understanding of the definition of metaphysical naturalism.

Metaphysical naturalism is a claim about what makes up reality. John Searle calls this the "basic facts" of reality. What he has in mind are those facts that relate the "mindless, meaningless, unfree, nonrational, brute physical particles." These entities and

¹⁸ Michael C. Rea, *World Without Design: The Ontological Consequences of Naturalism*, (Oxford: Clarendon Press, 2002), 66.

¹⁹ I thank Dr. Todd Buras for the phrasing.

the relations between them are the “basic structure of the universe,” and there is nothing else. Anything beyond or incompatible with this basic structure is ruled out. Peter Unger also gives a clear understanding of the basic facts in his article “Free Will and Scientiphicalism.” He states, “distributed differently in space at different times, there is physical stuff or matter in the world; as it can exist whilst never experienced by any sentient being, this matter is mind-independent.”²⁰ In both definitions, mindedness is something *excluded* from naturalist thought about the basic constituents. Unger makes a more positive claim in a version of naturalism’s implications for humans: “wholly composed of such mindless physical parts as electrons, you are a being whose powers are all physical powers, physically deriving from the powers of your parts and their physical arrangements.”²¹ What “physical stuff” is composed of, however, is still best described by what it is not. This different physical stuff interacts as determined by physical laws and gives rise to humans. Thus, humans are nothing more than the base physical matter interacting in mechanistic ways. The metaphysical naturalist would believe this not only about humans, but also about everything else in reality.

Unger and Searle make the foundational claim of metaphysical naturalism very clear, but another philosopher, Alex Rosenberg, emphasizes an important point. Rosenberg considers himself a “scientist” because his assertion is that “science’s description of the world is correct in its fundamentals; and that when ‘complete,’ what

²⁰ Peter Unger, “Free Will and Scientiphicalism” *Philosophy and Phenomenological Research* 65 (2002): 1, accessed September 21, 2013, <http://philosophy.fas.nyu.edu/docs/IO/1179/freewill.pdf>.

²¹ Ibid.

science tells us will not be surprisingly different from what it tells us today.”²² The first half of Rosenberg’s claim relates that the “basic facts” are the essential components of reality. Science has painted a picture of a massive organization of mindless particles related only by mechanistic tendencies, and Rosenberg here claims that this description is the fundamental picture of the world. The second half of Rosenberg’s claim can be formulated more vividly as: “nothing at the unsettled frontiers of physics challenges the parts we’re going to make use of.”²³ The assertion that science’s general character is fixed is a crucial aspect of metaphysical naturalism. Some philosophers are open to the possibility of science including mental or non-material entities as basic posits of science. However, the full-blooded metaphysical naturalist instead views science as correct thus far and as in a position to make very good estimates about what completed science will look like.

These two assertions (science is the tool to understand reality and everything we discover will fit with its general character) are the central tenets of metaphysical naturalism. These assertions have implications for how the metaphysical naturalist should account for parts of reality that appear non-mechanistic. John Searle says it becomes necessary for the naturalists to “square” non-mechanistic aspects of humans with the basic non-purposive structure.²⁴ To accomplish the task, metaphysical naturalists must appeal back to methodological naturalism. They must investigate reality

²² Alex Rosenberg, *The Atheist’s Guide to Reality: Enjoying Life without Illusions*, (New York: W. W. Norton & Company), 6-7.

²³ Alex Rosenberg, *The Atheist’s Guide to Reality: Enjoying Life without Illusions*, (New York: W. W. Norton & Company), 21.

²⁴ John Searle, “Philosophy and the Basic Facts” in *Freedom and Neurobiology*, (New York: Columbia University Press, 2007), 4-6.

through science to discover the basic mechanistic, physical entities at play in humans. Simon Blackburn realizes the constraints of a naturalistic program and therefore describes metaphysical naturalism in a way that focuses on these problematic areas. He asserts naturalism to be, “see[ing] human beings as frail complexes of perishable tissue, and so part of the natural order... it is above all to refuse any appeal to a supernatural order.”²⁵ Borrowing terms from Wilfrid Sellars, Blackburn calls the common conception of humans a part of the “manifest image” and the naturalist understanding of reality the “scientific image.” The manifest image describes the common agent’s perception of the world as meaningful and purposive, whereas the scientific image describes the world in terms of mechanistic tendencies. These images need to be reconciled by closing the gap between them. Many naturalists opt to close the gap through pure reduction, or a redefinition of all aspects of the manifest image in terms of the scientific image. Blackburn offers another important option as he discusses ethics, an aspect of the manifest image. He argues that reduction to the physical is not necessary, but *some* kind of reconciliation is necessary. The manner in which the reconciliation occurs determines whether the theory is naturalistic. For Blackburn, so long as the scientific image is the base line, then the theory is metaphysically naturalistic. This understanding of metaphysical naturalism explains how there can be various levels of qualified metaphysical naturalism.

Therefore metaphysical naturalism has two views as well: (1) the “basic facts” of reality are mechanistic, non-purposive entities as revealed by completed science, and (2) there is nothing that does not depend upon these entities. Clearly this constrains the

²⁵ Simon Blackburn, *Ruling Passions*, (New York: Oxford University Press, 1998), 48.

naturalist's understanding of reality. These restrictions are important to understand before delving into naturalist explanations of consciousness. The next section relates the types of responses metaphysical naturalists may have to any given phenomenon, as dictated by the constraints of their central two views.

The Constraints of Naturalism

It should now be simple to move from the definitions of methodological and metaphysical naturalism to the constraints placed by them on metaphysical theories. Recall that a comprehensive methodological naturalism often leads to metaphysical naturalism because of its assumptions about the natural world. Since metaphysical naturalism is the stance with the more fundamental implications than methodological naturalism, it is the stance to be discussed for the remainder of this thesis. Thus the term “naturalism” will henceforth be shorthand for metaphysical naturalism. The question is perhaps best phrased by Searle as: if everything really is brute, physical meaningless entities governed by regular laws that serve no purpose, how do we fit in?²⁶ The common conception of humans involves agential terms that at first glance do not seem to fit the scientific image. If humans reduce to nothing more than an agglomeration of physical particles interacting in regular ways, then it is difficult to see how agential concepts could apply. It is also difficult to perceive how naturalists could explain how this illusion of agential concepts could arise from the reality they claim underlies it. But they must reconcile both of these gaps. In this section the (metaphysical) naturalists' options will be explored. Essentially, a naturalist has three options for accounting for

²⁶ John Searle, “Philosophy and the Basic Facts” in *Freedom and Neurobiology*, (New York: Columbia University Press, 2007), 4-5.

anything from the manifest image. These options are: (1) reduce the phenomena to the physical, (2) eliminate the phenomena, or (3) posit a non-mechanistic primitive. These three positions cover the naturalists' main range of possible responses, although the third option covers ground that may betray naturalism. Each of these options will now be examined in turn.

Option 1 for the naturalist is to reduce the phenomena to the physical, or mechanistic, workings of the universe. In his article on reduction in the *Routledge Encyclopedia of Philosophy*, Jaegwon Kim defines reduction as a "procedure whereby a given domain of items... is shown to be either absorbable into, or dispensable in favor of, another domain."²⁷ There are multiple ways one domain can be reduced into another. One way is through derivation, such as the reduction of the laws of planetary motion from principles of Newtonian mechanics and force laws. In this way, the laws can simply be an extension of a preexisting theoretical framework.²⁸ Another method of reduction is through definition. If concepts in one domain can be defined using only concepts of another domain, then those concepts are reduced into that domain. Slightly different from either of these methods of reduction is Ernest Nagel's model of reduction. While derivation and definition both play a role in his thinking, neither is required here. Instead, Nagel proposes reduction through bridge laws, or connecting laws that link the vocabulary of the reduced domain to the domain to which it is to be reduced (the reduction base). The stringency of the connection depends on the nature of the domains

²⁷ Jaegwon Kim, "Reduction, problems of," in *Routledge Encyclopedia of Philosophy*, ed. E. Craig, (London: Routledge, 1998), accessed November 13, 2013, <http://www.rep.routledge.com/article/Q089>.

²⁸ Ibid.

being discussed, but Kim articulates that “each predicate of the theory to be reduced must be connected, via a *biconditional law*, to a nomologically coextensive predicate of the reducer.”²⁹ Thus the concepts in each domain must be interchangeable with every instance of the other domain’s concepts. The difference between reduction by bridge laws and reduction by definition, however, is that the translation is based on the meaning equivalences of the bridge laws, rather than by empirical laws.

Reduction is the ideal option for naturalism. When fully accounting for the world, the more that is incorporated into an explanation the better. This lends the thesis explanatory power, or higher probability in being correct due to its incorporating more evidence. Option 1 therefore enables the naturalist to have a better explanation because they incorporate more difficult phenomena into their naturalistic explanation. With option 1, the naturalist can satisfy our intuitions that the world appears a certain way.

Naturalists, by reducing phenomena from the manifest image, do not deny the world’s appearance but instead incorporate it into their explanation. Strong reductive accounts also give an explanation for how and why things appear as they do. These kinds of explanations account best for agential perspectives, and therefore are considered more plausible. Occasionally these explanations are less explicit in their account of agential perspectives. An example would be Colin McGinn, with his mysterian understanding of consciousness. In his account, he argues that option 1 is possible for consciousness but humans will never know it because of our cognitive closure, or our restriction to our own perspective.³⁰ While he has confidence in the truth of metaphysical naturalism, he has no

²⁹ Ibid.

³⁰ Colin McGinn, *The Mysterious Flame: Conscious Minds in a Material World*, (New York: Basic Books), 1999, 45-46.

explanation of how reduction of consciousness could be true. Such a stance obviously has less explanatory power than a stance that explains exactly how consciousness fits into a metaphysically naturalistic world. It is crucial for explanatory power that the phenomena under consideration are made clear to us as humans. Claiming that humans are incapable of such understanding with regard to phenomena is something that should only be conceded when all other alternatives have been explored, so as to prevent us from giving up on finding an explanation of a phenomenon too early.³¹

Option 2 (eliminating the phenomena) is another method naturalists utilize for phenomena outside their mechanistic metaphysics. Georges Rey writes in his article on eliminativism for the *Routledge Encyclopedia of Philosophy* that “eliminativism about anything, x , is the view that x 's do not exist.”³² This is not the same as reduction, because reduction actually grounds or supports belief in one domain by identifying with the (confidently held) reduction base. Eliminativism argues instead that the concepts of a certain domain do not apply to anything in one's ontologies, and so do not apply to reality. Option 2, while often counterintuitive, is just as naturalistic as option 1. By eliminating everything but mechanistic explanations, the naturalist still maintains a coherent explanation of reality that fits with the central tenets of naturalism. However,

³¹ It is important to note that even though McGinn claims no knowledge towards reducing consciousness, he gives good reason (namely the structure of the nature of mind and the structure of our theorizing) to think that such knowledge is impossible. He also asserts that we may proceed with neurophysiology and psychology, and that we may come to know the form of the solution to conscious reduction. It is only the specifics that escape our grasp. Such positions are crucial for McGinn to remain naturalistic rather than skeptic, and lend him more explanatory power than other mysterian positions.

³² Georges Rey, “Eliminativism” in *Routledge Encyclopedia of Philosophy*, ed. E. Craig, (London: Routledge, 1998), accessed November 14, 2013, <http://www.rep.routledge.com/article/W012>.

this option is less convincing to naturalists' opponents, as it lacks a certain explanatory power. Often even naturalists view eliminativism as inadequate. The failure lies in the fact that eliminativism fails to do what option 1 can: account for the appearance of things. If so much is an illusion, the theory that explains *why* the illusion takes place has much more explanatory power than the theory that offers no consolation.

Option 3 cannot be taken lightly, as it effectively qualifies or limits one's naturalism. Such a concession can vary in degree. Many naturalists qualify their naturalism to a lesser degree by incorporating into their worldview things such as numbers, universals, and sets. Such concessions venture less from pure naturalism than others do since these kinds of things are not causally active. They play no real role in the causal order of the world, but rather are structural or abstract features of the world. All non-naturalistic posits concede something not already explained by science. However, these posits vary in character. These varying "grades" of non-naturalism include supervenience explanations, epiphenomenalism, and non-reductive physicalism. If a naturalist concedes a non-mechanistic property's causal efficacy their metaphysics becomes full-blown non-naturalism. Additionally, if brute, the posit does not have explanatory power. Brute posits that do not fit the general character of the naturalist theory are not an option for the naturalist.

Such are the three options that naturalists have to account for phenomena. It is important to keep these three options in mind for the upcoming discussion of consciousness. Ultimately, the best option is the one with the most explanatory power for the phenomena in question. However, before the naturalist positions on consciousness

are explained, it is necessary to understand the phenomena of consciousness. This is the subject of the next chapter.

CHAPTER TWO

The Phenomenon of Consciousness

The Two Problems

Introduction

In the last chapter we learned that the metaphysical naturalist has one goal for his account of the universe, namely, to eliminate the purposive in favor of the mechanistic picture. Since humans function in daily life under the purposive picture, it behooves the naturalist to explain any apparent discord between these two explanations. Thus the naturalist avoids using lay vocabulary of purposive or agential concepts and instead relies on mechanistic vocabulary. Many phenomena are difficult to fit into this framework, but the biggest challenge for naturalists is the phenomenon of consciousness. This thesis will explore why the phenomenon is so challenging and if there are any possible solutions available to the naturalist; but first, this chapter will explain what the phenomenon is. A substantive definition of consciousness is the goal for the first section of this chapter, after which we will consider its implications. Since the concept of consciousness is so difficult, however, our definition must arise from distinctions between two commonly conflated phenomena. For this we will appeal to both experiential and technical distinctions.

The Experience of Consciousness

The phenomenon of consciousness is difficult to differentiate. Many philosophers and scientists in the past have mistaken it for various kinds of higher-order thinking.

However, as we will soon show, consciousness is best identified by the subjective character of experience. This qualitative feel is the distinct difference between seeing green and white, tasting a lemon and a cherry, or hearing the note F# and the note Bb. Higher-order thinking includes functions such as self-consciousness, awareness, introspection, knowledge, control, etc.¹ At their basis these are causal or functional concepts, which D.M. Armstrong defines as, “the concept of a mental state [that] essentially involves, and is exhausted by, the concept of a state that is *apt to be the cause of certain effects or apt to be the effect of certain causes.*”² Higher-order thinking’s functional analyzability makes reduction of it to mechanistic terms simpler than reduction of subjective experiences. At the same time, these functions are intricately related to our lives and experience. It is natural to consider the functions a central part of our experience, and so conflation of higher-order thinking with consciousness is hard to avoid. To avoid conflation, it is necessary to have a basic working knowledge of each concept. We will follow David Chalmers’s work in *The Conscious Mind* in order to get the clearest picture possible of the distinctions. In so doing, we will adopt his terms for each concept: psychological consciousness for higher-order thinking, and phenomenal consciousness for the subjective character of experience.

One reason for the conflation of psychological and phenomenal consciousness is due to the inaccuracy of language. Chalmers refers to this as “the double life of mental

¹ David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, (New York: Oxford University Press, 1996), 26-27.

² D.M. Armstrong, “The Causal Theory of the Mind,” in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 82.

terms.”³ In everyday life people use mental concepts that consist both of a psychological and a phenomenal part. An example he gives is that of pain. Obviously when we say that we are in pain, we are mostly referring to the phenomenal quality that “hurts.” But when we talk of pain in animals, we may only mean that the animal has a natural aversion to those kinds of stimuli which are counterproductive to its health and flourishing.⁴ We could also talk of this psychological aspect of pain in humans, say, when someone is “jumping up and down in pain.” It is clear that the statement about pain is more a description of a causal-functional role than one about how the person feels. However, it seems that there are often mixed references in our speech as well. We can use the word “pain” to refer both to the psychological and phenomenal occurrences of the same exact event. Such an example might be that “he was in so much pain from his backache he couldn’t focus on his work.” We are referring both to how the man is feeling and how the pain manifested in terms of functional causes and effects. Presumably underneath the surface of the statement lies the description that the man is aching and uncomfortable, as well as the one that a problem with his back is inducing a mental state that ensures his attention is on how to fix the problem. Chalmers goes on to discuss how some mental concepts have an obvious “double life,” whereas others lean more towards one side or the other. Nevertheless, our language does not make sharp distinctions between the two types of consciousness, and so our concepts are limited in a similar manner.⁵

³ Ibid., 16.

⁴ Ibid., 17.

⁵ Ibid., 18.

Another reason psychological and phenomenal consciousness become conflated is due to their instantiation at the same time. While it may not be logically necessary that the two types of consciousness occur together, it is how they occur in every instance we observe. As Chalmers puts it, “conscious experience does not occur in a vacuum.”⁶ Our experience is related to our functional states, and likely arises from them in one way or another. Since phenomenal consciousness seems dependent on the psychological, any distinction between the two may be more conceptual than actual. However, since we are dealing with conceptions of consciousness, these distinctions are still crucial.

Phenomenal consciousness also seems to have another disadvantage in language. Because experience is so immediate, it is somewhat “ineffable.”⁷ The intrinsic aspects are inexplicable, whereas the extrinsic aspects (what the experience is *of*, or *in reaction to*) are more apparent and communicable. This makes it even more difficult to distinguish the differences of phenomenal consciousness from psychological consciousness, since we often lack the vocabulary to isolate the phenomena about which we wish to talk.

Given all of these difficulties, it may seem that these two concepts of consciousness are an arbitrary distinction and not real at all. In order to emphasize the contrary, this chapter will now move into more substantive descriptions of each concept. Along the way, the chapter will also discuss how naturalistic reduction is able to handle each phenomenon. The chapter will conclude by explaining why phenomenal consciousness presents a challenge to the naturalist.

⁶ Ibid., 22.

⁷ Ibid., 22.

Psychological Consciousness

As previously mentioned, psychological consciousness is higher-order thinking. There are several kinds of higher-order thinking that we often call consciousness.⁸ All of these processes are *functional*, meaning that they are defined by the role they play rather than their intrinsic nature. The most basic psychological consciousness is merely being awake as opposed to asleep. Being awake is often confused for consciousness; however, we have conscious experience even while sleeping, which is why dreams are so surreal. Introspection, reportability and self-consciousness's functions are each about the intake of data from ourselves. Introspection allows us to be aware of what mental state we are in and use such information, whereas reportability is the ability to report the contents of our mental states. Introspection and reportability are both psychological consciousness because they are functional attributes. Some philosophers attempt to explain phenomenal consciousness in terms of introspection or reportability. Nevertheless, introspection and reportability themselves are not phenomenal. Self-consciousness is what distinguishes us from most animals, enabling us to think about ourselves as distinct from the world around us. Many confuse this function to be phenomenal, but it is not. Within the sphere of deliberation or intentionality, we have attention and voluntary control. Attention has to do with the directedness of our thought, or how it is always *about* something. Voluntary control is how our acts are caused in the right way by previous thought. Knowledge involves our belief or assent to the truth, and is usually thought to be needed for consciousness of that truth. These concepts are also all functional, which distinguishes them as part of what Chalmers calls the "easy problem" for naturalists.

⁸ Ibid., 26-27.

The “easy problem” is to explain these functional aspects of consciousness in terms of the naturalist framework.⁹ As noted in chapter one, any attempt to reduce a phenomena to fit the naturalist picture must be explicable in purely mechanistic terms. For the concept in question to reduce into the naturalist picture, it must be replaceable by the mechanistic explanation in every instance that it occurs. We should also be able to replace the mechanistic explanation with the concept in question wherever it occurs. Such a relation is an identity relation, and this is what is required for the naturalist picture to succeed. Obviously this is no simple task when it comes to the mind. There is much unexplored in psychology. We have much to learn about the interactions of chemicals and electrical impulses in our brains. There are undoubtedly thousands of relations between neurons that we have never encountered. Nevertheless, this is still considered the “easy problem” of consciousness. The name is due to the possibility of achieving this understanding one day. Science seems to have all the right tools for the job. Chalmers refers to these aspects as “puzzles rather than mysteries.”¹⁰ As discussed, functional concepts are *already* defined only in terms of their causal role or relation to the external world. Science’s definitions are always in these terms, and so it does seem evident that given enough time and research, science will make headway in defining higher-order thinking. We can expect science to reveal laws reflecting the identity relationship of these functional aspects of the brain with certain neurons behaving in certain ways because we expect *something* to fill the role. Indeed, we can have much confidence in this type of identification because it has already occurred in some instances: for example,

⁹ Ibid., 24.

¹⁰ Ibid.

the identification of dopamine as one of the neurotransmitters involved in addiction. Therefore there is hope, so to speak, when it comes to understanding the mental states of psychological consciousness.

Phenomenal Consciousness

Phenomenal consciousness has a completely different character. As we mentioned before, it consists of the qualitative feel of subjective experience. This experience is the most immediate thing in our lives, and it is also what makes our experience uniquely ours. It is impossible to share phenomenal consciousness with anyone else, as nobody else can access it immediately the way you do. Chalmers runs through a laundry list of the most obvious of these experiences.¹¹ On the list are included all sensory experiences: visual, auditory, tactile, olfactory, taste, temperature, pains and other sensations. Each sense has a unique character and a unique external stimulus inducing the reaction in us. Examples are the taste of a granny smith apple as opposed to a red delicious apple, or the feeling of a soft cotton blanket as opposed to the scratchiness of wool or slippery smoothness of silk. Other phenomenal experiences include the mental images we can pull up at will. If I say “picture a tree” and you do so, then undoubtedly you have some representation that is phenomenally unique. It is uniquely *your* representation of a large branched plant, and it is uniquely characterized by qualitative traits of an oak, aspen, pine, etc. Emotions are also phenomenal. There is something distinctive about anger as opposed to frustration, though they may produce similar behavior. Disappointment is very different from hope, which is very different from happiness. Moving into less obvious areas, certain thoughts can be phenomenal.

¹¹ Ibid., 7-10.

This is most noticeable if the thought is actively thought to oneself, because there “is *something* it is like to be having such thoughts.”¹² The final sense Chalmers discusses is the sense of self. This is not the same as self-consciousness, but rather is the “background hum”¹³ of experience, the “what it is like”¹⁴ to be you.

These experiences are not so obviously mechanistic in nature. They are not functional, in that they are not defined in terms of their causal role or relation to the external world. Instead, these qualitative experiences, or qualia, are known by their intrinsic character. In the words of Thomas Nagel, these experiences are subjective and therefore require a point of view.¹⁵ It makes little sense to talk of my experiences as separate from me. It also makes little sense to consider my experiences as immediately accessible to you the way they are to me. This inaccessibility obviously greatly inhibits naturalistic reduction, which requires third-person, objective explanations and descriptions. Hence Chalmers refers to this as the “hard problem.”¹⁶ Reductions of phenomenal experience seem to be beyond our ken, and possibly necessarily so. It is not hard to conceive the possibility that these phenomena may never be reduced in the way psychological consciousness will. We have no idea how physical matter could give rise to such experiences, whereas we have an idea with psychological consciousness. Science no longer seems to have the right tools. This difficulty is highlighted specifically by

¹² Ibid., 10.

¹³ Ibid.

¹⁴ Thomas Nagel, “What Is It Like to Be a Bat?” in *The Philosophical Review*, 83: 4, (1974), 436, accessed December 5, 2013: <http://www.jstor.org/stable/2183914>.

¹⁵ Ibid., 438.

¹⁶ David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, (New York: Oxford University Press, 1996), 25.

several thought experiments. Now that we understand the phenomena at hand, we will be able to understand the thought experiment and understand its implications. We will do this in the next chapter.

CHAPTER THREE

The Zombie Argument

The Zombie Argument as a Thought Experiment

Introduction

The previous chapters dealt with the crucial definitions to the investigation of naturalistic thought of consciousness. With a working account of both naturalism and consciousness at hand, it is now possible to delve into the specifics of what problems there are for a naturalistic definition of consciousness. This chapter relates one argument against naturalism of the mind: the zombie argument. This argument uses a thought experiment to claim that it is possible that phenomenal consciousness is not identical to psychological consciousness. It then uses that possibility to show that they are not identical. In order to proceed, then, it will be necessary to investigate what thought experiments are, how conceivability relates to possibility, and how some kinds of identity require necessary identity.

Thought Experiments

Before proceeding to the zombie argument, it is important to explain the method of investigation used for its construction. Much of philosophy uses thought experiments, or experiments that are performed by thinking about how concepts apply in a given set of circumstances. Through these experiments it is possible to gain new understanding

without new empirical data.¹ This new understanding can be achieved because of our storehouse of instinctive knowledge or intuitions that are gained through experience.² As we posit a definition of a concept, we have the ability to test it against our intuitions about reality. If the definition tells us that the concept applies in *all* the same situations that we intuitively know the concept to apply, then it meets a standard called the sufficiency condition. If the definition also tells us that the concept applies *only* in the situations in which we intuit that the concept applies, then the definition also meets the necessity condition. One purpose of this chapter is to discuss these two conditions. Obviously thought experiments' success depends upon our intuitions. Intuitive knowledge will vary greatly from person to person, and therefore thought experiments are often highly contentious. Thought experiments still help to focus discussion and attention on the most difficult cases, however. Some cases do not exist in reality, but instead are only possibilities. These cases are the most contentious, and are usually the cases where various definitions of a concept conflict the most. Sometimes definitions line up perfectly in reality and pick out the same phenomena for a concept. These same theories, when applied to possible worlds, then differ greatly. This tension is where the true philosophical discussion begins. Later in this chapter an example of a contentious possible world will help to clarify the implications of naturalistic thought for the philosophy of mind. However, a few remarks on necessary and sufficient conditions (the conditions which make for necessity and sufficiency relations) are required first.

¹ James Robert Brown and Yiftach Fehige, "Thought Experiments," in *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition), ed. by Edward N. Zalta, <http://plato.stanford.edu/archives/fall2011/entries/thought-experiment/>.

² Ibid.

Although thought experiments may seem vague, they help to distinguish rather precisely the necessary and sufficient conditions of when concepts apply. A necessary condition for a concept to apply might be thought of as an essential building block of a concept. In order for the concept to apply correctly, the phenomena in question must have this certain feature or aspect. An example might be that a necessary condition for life is cell growth. A certain thing cannot be said to have life unless it has cell growth. This might be translated into a standard “if-then” statement: “If a thing is alive, then it has cell growth.” This restricts the possible applications of the concept of “life” to a narrower range of things: namely, those with “cell growth.” A sufficient condition, on the other hand, is an aspect of the phenomena that enables an automatically correct application of a concept. An example is that anything that “breathes” has “life.” The conditional statement might then look like this: “If something is breathing, then it has life.” Although not all things that have life breathe, all things that breathe have life. Sufficiency conditions provide features which always enable the correct application of the concept. A concept might have both a necessary and sufficient condition, and indeed this is true of “cell growth” and “life.” Something cannot have “life” without “cell growth,” and cannot have “cell growth” without thereby having “life.” “Cell growth” might then be taken to be part of the fundamental definition of “life.”

Thought experiments help to distinguish necessary and sufficient conditions because they help clarify what features *enable* the correct application of a concept and what features *are required for* a concept’s correct application. We compare which way a case is sorted by a necessary or sufficient condition with the way that our intuitions sort the same case. As we repeatedly appeal to a wide base of various cases, or situations

involving the concept in question, we better understand the concept and when, exactly, it applies. These cases may be either actual or possible. In actual cases, we draw on experience. In looking at plants, they clearly have “cell growth,” and so we apply the concept of “life” to them. This supports the necessary condition of “cell growth” for “life.” We can also support the sufficient condition of “breathes” for “life” by observing humans, birds, dogs, fish, and other creatures that have “life” and “breathe.” If any creature was found to be both breathing and dead, then it would contradict the sufficient condition.

In thought experiments, we also appeal to possible cases. As mentioned previously, these possible cases broaden our understanding of the concept. Possible cases allow us to determine the differences between definitions that do not vary in their application to actual cases. These are cases which may not happen in the world or universe that we live in, but might have happened, had the universe been different. For possibility, we often use the guide of conceivability. Conceivability does not necessarily entail possibility, but it at least lends support to a case’s possibility. Possible cases are those that can be described with no internal contradictions. These cases are important because they often exemplify situations where the necessary and sufficient relations come apart. We might imagine some substance, for example, that contains no cells but exhibits some other sort of growth. Or we might consider aliens with the same description. Either way, we would have to decide what our definitions of our concepts decide (neither are alive) and what our intuitions would decide (the substance may or may not be alive, but the aliens seem to be). Thus this possible case supports both conditions discussed

earlier, as it is an instance of both a possible living thing with cell growth and a possible breathing thing that is alive.

Just as in science experiments, even one legitimate instance of the condition failing to correctly sort a thought experiment can result in one of two things: a weakening of the condition or a deeming of the experiment to be an anomaly. This leads to the difficult question of what to do in such a situation. Perhaps the simplest solution is to “throw out” such a thought experiment, arguing that it fails to represent a genuine possibility. However, this often requires some sort of explanation for why it fails to be possible, or why it is not a good example of the conditions. Another solution is to change the conditions slightly so as to sort the case correctly. Such an amendment might be very small. Sometimes, however, the change is considerable, and thus the condition may need to be completely revised. The method of determining which solution is best in a given situation has no established rules. Fortunately there is a principle to help guide philosophers through this gray area: the principle of reflective equilibrium. Essentially this principle advises moderation when dealing with adjustments to theories. One should be flexible in changing one’s theories to suit the phenomena, but also, as the theories become stronger, our understanding of the phenomena may be impacted by the theories. Because this is such a difficult balance, different philosophers approach the same problems differently. These differences are typically due to the philosophers’ commitments in other areas, the strength of their intuitions, or the strength of their confidence in the conditions proposed. These background commitments are the main reason why there is so much contention surrounding unclear thought experiments. The

zombie thought experiment is highly debated, and this is why clarity on necessary and sufficient conditions is paramount to sorting through the arguments.

The Possibility of Zombies

The zombie argument is one of a family of arguments. These arguments are known as the conceivability arguments. Conceivability arguments involve the conceivability of a being that is physically identical but not phenomenally identical to a normal human being.³ The being's phenomenal experience may differ by being completely absent, being inverted, or being partially absent.⁴ These scenarios all attempt to prove the same thing, i.e. that the phenomenal is independent from the physical world. They also use the same means to prove it, namely, they imagine a changed phenomenal experience yet identical physical interactions. This shows the independence of the two kinds, and therefore the non-physicality of the phenomenal. This chapter will use the zombie argument as a representative for the family, but the other arguments would proceed similarly.

The zombie argument represents a logical extreme. The argument involves conceiving of a being with normal, functioning physical responses to its environment while remaining phenomenally "empty." Some might question whether this begs the question against the causal efficacy of phenomenal experience. While the zombie possibility is certainly easier to imagine if phenomenology is not causally active, it is also possible to imagine a zombie without this assumption. To do so, imagine a zombie

³ David J. Chalmers, "Consciousness and its Place in Nature," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (Oxford University Press: New York, 2002), 249.

⁴ Ibid.

nearly physically identical, but also having an added physical something that plays the role qualia normally play in the causal chain. Either way, this being has thoughts, feelings, and emotions insofar as it has the same neurological processes working in their brain. The difference lies in the *experience* of these thoughts, feelings, and emotions. Serotonin would be produced in the zombie as well as it is produced in the human, but no “warm fuzzies” would be experienced. A perfectly physiologically identical zombie, however, would still be able to report that she is happy, that she has reasons for being happy, and perhaps even describe the feeling, i.e. describe that she has “warm fuzzies.”⁵ The zombie would thus be functionally identical to a human being, having all of the proper physical responses to her environment. Not only would the zombie be responsive to the environment, but also she would be in perfect relation with the physical world, able to interact with her surroundings in the same way a human can.

An actual case which draws near the zombie case is known as the long-distance truck driver.⁶ Say you get into your car and you begin to drive home from work. You have made this trip a thousand times, and you begin thinking of things other than the road ahead and the cars around you. You function just as well as ever, maintaining control of the vehicle and changing lanes when necessary, using turn signals, etc. You are the

⁵ This last feature of the zombie is unclear, because it leads to issues with whether the zombie is making a false statement. (See Robert Kirk, "Zombies," in *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), edited by Edward N. Zalta, <http://plato.stanford.edu/archives/sum2012/entries/zombies/>) For the sake of the argument we will assume these statements to be true. Based on the relevant naturalists' positions such language just is about the functional processes zombies would have anyway.

⁶ What follows is a modification of Armstrong's example in D. M. Armstrong, "The Nature of Mind" in *The Nature of Mind and Other Essays*, (Humanities Press: New York, 1968).

model citizen of someone who drives his or her car. And yet, you have had an exhausting day and so you begin to think less and less about anything in particular. Soon you are just driving. You are not consciously thinking about anything. In fact, you are not even paying attention to any inputs from your environment at all.

If this thought experiment is to be brought to the level of a full-blown zombie, it would need to be the case that you do not *experience* anything at all. Your phenomenal world is empty. Your neurons continue firing and you continue driving perfectly. Perhaps you even see a friend on the way home, honk, wave, and smile. But you do not experience any of it; it is only your physical system interacting with the world via cause and effect.

To return to the real-life situation, you pull into your garage and as you turn off the car you suddenly “wake up.” You are aware, in the phenomenal sense, again. You can feel the steering wheel and the key in your hand, and the leather of the seat beneath you. You smell the pine of the air freshener, and you open your eyes wide in the darkness of the garage as your garage door closes. The true zombie would never “wake up,” but would live her whole life phenomenally empty. The striking feature of the thought experiment draws out the difference between phenomenal consciousness and psychological consciousness previously discussed. Phenomenal consciousness is the characteristic feel of interacting with the world around you, or the “what it is like” to be you.⁷ Psychological consciousness is the functional role your brain plays as you go about life, composed of the physical reactions (such as neurons firing) to your environment that then produce your bodily reactions.

⁷ Thomas Nagel, “What Is It Like to Be a Bat?” in *The Philosophical Review*, 83: 4, (1974), 436, accessed February 7, 2014: <http://www.jstor.org/stable/2183914>.

Another phenomenon similar to zombies is the case of blindsight. In blindsight, the patients' visual cortexes are damaged, resulting in the patients' report that they are unable to see anything in certain areas of their visual field.⁸ These gaps are called "scotoma," and seem to be holes in conscious perception. Nonetheless, when presented with certain challenges or problems to solve that involve the information accessible only within the scotoma, the patients perform well above mere chance. They are able to grab items, identify items, and make choices with information retrieved from such areas. This implies that they have psychological awareness of the stimuli but not phenomenal consciousness of them. A zombie would have a scotoma with a range that takes up the full range of their visual field. In other words, they would lack all conscious vision. Furthermore, a zombie would possess similar phenomena with regard to every other sensation of which they should be conscious, resulting in a completely phenomenally "empty" twin.

Chalmers succinctly describes zombies this way, "these systems will look identical to a normal conscious being from the third-person perspective: in particular, their brain processes will be molecule-for-molecule identical with the original, and their behavior will be indistinguishable. But things will be different from the first-person point of view."⁹ These thought experiments enable us to test the conceivability of zombies, and it seems that they are indeed conceivable. Furthermore, the fact that these *actual*

⁸ Colin Allen and Michael Trestman, "Animal Consciousness," in *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), ed. by Edward N. Zalta, forthcoming URL = <<http://plato.stanford.edu/archives/sum2014/entries/consciousness-animal/>>.

⁹ David J. Chalmers, "Consciousness and its Place in Nature," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (Oxford University Press: New York, 2002), 249.

cases are very similar to what a zombie would be like makes the possibility of zombies more likely. It is difficult to deny that an actual case can be taken to an extreme or used to extrapolate further phenomena. Partial zombies' actuality makes full zombies' possibility more likely. If we have already seen partial zombies, denying full zombies requires having a good reason for their logical impossibility.

The Zombie Argument

In order to understand the implications of the possibility of zombies, it is not enough to conceive one, or think one up. The conceivability of zombies is just one step in a multi-faceted argument. The argument may be formulated as follows:

1. Zombies are conceivable.
2. If zombies are conceivable, then they are metaphysically possible.
3. If zombies are metaphysically possible, then phenomenal consciousness is nonphysical.
4. If phenomenal consciousness is nonphysical, metaphysical naturalism is false.
5. So, metaphysical naturalism is false.¹⁰

This argument is a valid argument, which means that it is impossible for the conclusion to be false while the premises are true. Naturalists hold the conclusion to be false, and so they must counter the argument by rejecting the truth of one of the premises.

The preceding section addressed the relatively uncontroversial 1st premise. However, there are philosophers who think we cannot truly conceive a zombie. These philosophers posit unknown or unknowable naturalistic truths that would produce a contradiction if we knew them. Currently there is not good reason to posit such a truth, as it would beg the question by assuming naturalism. In short, this is not a satisfactory answer. Recall the earlier discussion of conceivability, where the discussion relayed that

¹⁰ Ibid., adapted with the help of Dr. Todd Buras.

if it is possible to describe a zombie without contradicting oneself, then zombies are conceivable. Thinking up a zombie was made possible by extrapolating from the long-distance truck driver case and instances of blindsight. The description given did not seem to be contradictory. This means that if there is a contradiction, it is hidden from our current knowledge. This is why naturalists who attack this premise must posit unknown truths or unknowable ones. Chalmers calls these philosophers “Type-A materialists.”¹¹ Within this category, this paper discusses only the eliminativists. Eliminativists of phenomenal consciousness argue that zombies are not conceivable because our so-called concepts of phenomenology do not refer to anything. Their position will be discussed in greater detail in the next chapter.

Premise 4 follows from the definition of metaphysical naturalism, a concept we explored in chapter one. The premise is construed as a conditional in logic. In order for this premise to be false, the consequent (“metaphysical naturalism is false”) must fail to materially follow from the antecedent (“consciousness is nonphysical”). The only way this is possible is if the antecedent is true and the consequent false. Because of the definition of metaphysical naturalism, metaphysical naturalists believe that only mechanical explanations, fulfilled by physical entities, explain reality. If consciousness is physical, then naturalism is consistent. If consciousness is nonphysical, then naturalism entails a contradiction. The simplest formulation of the contradiction would be that (1) reality is only composed of physical objects and (2) consciousness is nonphysical. This is not possible, and so premise 4 is true. Many philosophers modify this premise in favor of a more defensible naturalist philosophy, namely that the only

¹¹ Ibid., 251.

causally efficacious things in reality are entirely composed of physical objects. This suggests a response by a naturalist. Naturalists need only deny the causal efficacy of consciousness in order to hold both naturalism and nonphysical consciousness. This view is called epiphenomenalism, and will be discussed in the next chapter. Other philosophers may disagree with premise 4 by denying that metaphysical naturalism requires consciousness to be physical. Instead, they claim that conscious experience may be non-physical but interact with the physical in mechanistic ways. To claim this is rather difficult, as the naturalists must posit the nonphysical as something primitive. Both of these positions allow for some nonphysical entity or property, but they differ in how much importance such entities or properties are allowed to have. In addition to the eliminativists, we will explore these positions in the next chapter.

Premise 3 states, “If zombies are metaphysically possible, then phenomenal consciousness is nonphysical.” This premise is a natural consequence of what it would take to be a zombie, and what this implies about our conscious experience. A zombie lacks phenomenal consciousness, or mental content, despite being physically identical to a normal human being. Saul Kripke, in his work *Naming and Necessity*, demonstrated that identity statements flanked by rigid designators require a necessary relation.¹² Take the identity “Obama is president.” The name, or indicator, “Obama” picks out the same individual in every possible world and is therefore a rigid designator. The indicator “president” does not pick out the same individual in every possible world, and therefore it is not a rigid designator. We therefore do not expect this relation to be necessary.

¹² Saul A. Kripke, excerpt from *Naming and Necessity*, in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 329.

Because there is at least one possible world where Obama is not president, we can conclude that the Obama is not necessarily identical with the president. We therefore do not expect the two to have all and only the same properties. For instance, the president has the right to veto, whereas Obama on his own does not. In contrast, the terms involved in the identity statements of phenomenal consciousness are rigid designators, picking out the same phenomena in the relevant possible worlds. "Pain," for instance, is always the thing that feels painful. Additionally, designators of psychological consciousness also pick out the same phenomena in every possible world. To use the dummy case, the designators "C-fibers firing" always picks out the firing of C-fibers. Thus, for the naturalist to identify "pain" as "C-fiber firing" would require the identity to be a necessary relation. The designators would need to refer to all and only the same properties. If they are identical, they are identical in every possible world. The possible world where there is a disparity between phenomenal and physical states shows that the physical-phenomenal identification is not necessary. Zombies provide one example of a world where the disparity exists. If there is no necessary identification between the physical and phenomenal, then this means that phenomenal consciousness is not identical to psychological consciousness. This means that the metaphysical possibility of zombies entails the non-physicality of consciousness. The premise is therefore true.

Finally there is premise 2, "If zombies are conceivable, then they are metaphysically possible." Chalmers refers to the naturalist philosophers who counter this premise as "Type-B materialists." Similar to premise 4, in order for this conditional to be false, the antecedent must be true while the consequent is false. This would mean that zombies are conceivable but nevertheless not metaphysically possible. Type-B

materialists argue that conceivability is not a perfect guide to possibility. They rely on Saul Kripke's work to argue that there are necessary truths that are not knowable *a priori*, or by reason alone.¹³ Kripke's work proves that "water is H₂O" is a necessary identity despite the fact that it is conceivable that water is not H₂O. "Water is H₂O" is an *a posteriori* truth. In order to reason to *a posteriori* truths, we must use information we gain by sense experience. Typically, *a posteriori* truths are not necessary truths, meaning they are not true in every possible world. Instead, the special information we gain about our world only allows us to conclude that these truths are metaphysically possible. Metaphysical possibility indicates possibility for the world that we live in. Metaphysical possibility is too narrow for generalization of the possibility to *all* possible worlds. This might be thought in terms of quantifiers in logic. If we can prove a conclusion about a certain "stand-in" name that is not mentioned earlier in the proof (and therefore has no special information with it), then we can generalize that conclusion to be a universal truth. Similarly, if we can come to know a truth about our world *without* taking into account the special information we have about it, then we may rightly generalize that truth to be necessary.

This train of reasoning makes it seem that no *a posteriori* truths are necessary. Since *a posteriori* truths require the special information that we exclude in universalizing conclusions, they seem to be opposed to universalization. Kripke's work demonstrates otherwise by drawing out the case of H₂O = water. Kripke shows that it is not logically possible for water not to be identical to H₂O, but it is possible to imagine that water is not

¹³ David J. Chalmers, "Consciousness and its Place in Nature," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (Oxford University Press: New York, 2002), 254.

identical to H₂O. If this is the case, then it must be the case that conceivability is not a perfect guide to possibility.¹⁴ Applying Kripke's work to the mind, naturalists say that if qualia are conceivably separate from psychological phenomena, it is only because we lack certain truths that will prove that this separation is not conceivable. They further claim, relying on the "water is H₂O" identity relation, that *a posteriori* truths are sometimes necessary. From this they conclude that there is reason to think that there is a phenomenal-psychological identity and that its necessary identity is an *a posteriori* truth. And if this identity between qualia and psychology is *a posteriori*, then any conceivability cases will not be able to show us whether or not there is a necessary identity. If conceivability cannot help us determine identity, the zombie cases are not useful in the discussion of whether consciousness is physical.

The problem here is that Kripke himself does not agree with the parallel between zombie cases and H₂O cases. Instead, he argues that there is a distinct difference between the water = H₂O case and the phenomenology = psychology case. The difference lies in the indicators on either side of the identity relation. As we discussed during our explication of premise 3, an identity relation flanked by rigid designators must be a necessary one. People considered the identity water = H₂O to be a contingent truth because it needed to be discovered empirically before we could know it.¹⁵ Kripke argues that while water = H₂O identity is necessary and has two rigid designators, our deception

¹⁴ Hilary Putnam, "The Meaning of 'Meaning,'" in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers (New York: Oxford University Press, 2002), 590.

¹⁵ Joseph LaPorte, "Rigid Designators", in *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition), ed. by Edward N. Zalta, URL = <<http://plato.stanford.edu/archives/sum2011/entries/rigid-designators/>>, accessed April 15, 2014.

arises in our knowledge of the term “water.” When we say that water could be XYZ, we are confusing the term “water” as something that only picks out that thing in the world that feels wet, tastes refreshing, flows in streams, etc. These are merely roles played by whatever fills the designator, and therefore it is conceivable that very many things could satisfy the description in other possible worlds. However, after science tells us about everything that water does, it becomes apparent that “water” is shorthand for a molecule with very many complex jobs. When we know the *full* extent of its role (the way it interacts with hydrophilic and hydrophobic substances, its exact specific heat, etc.), it no longer seems possible for anything else to fill its functional roles. “H₂O,” is also a rigid designator, but one that has a rigidity that is more immediately obvious. It is not possible for the designator to indicate anything other than a molecule composed of two hydrogen atoms and one oxygen atom. Thus the relation initially seems contingent, but upon investigation it is clear that it is necessary. Kripke moves on to argue that the terms of phenomenology and psychology are all rigid designators. Recall our discussion of “pain” and “c-fiber firing” from earlier. Since these designators indicate the same thing in all possible worlds, they are rigid. Thus both water = H₂O and phenomenology = psychology, if identical, are necessarily identical.

The problem for the analogy of the identities, then, is that our lack of knowledge made it *seem* that something else could fill water’s role, and that is why the mistake of thinking it a contingent identity was possible. But for something to *seem* painful *entails* its being painful. Therefore nothing but pain could ever seem to be pain. Our lack of knowledge about pain, then, does not inhibit us from identifying it in all possible worlds. Therefore it cannot be possible that we are misconceiving the zombies by substituting

something else for pain, or qualia, or lack of qualia. We have more reason to accept the conceivability of zombies as a guide to their possibility than we have to reject it. Thus the conceiving of zombies genuinely is a possibility, and therefore a counterexample to the necessary relation of phenomenology and psychology. Premise 2 stands in the argument.

We have just completed a review of the argument from zombies against metaphysical naturalism. Each of the premises has been shown to hold weight, but there are several responses we have not yet developed fully. These responses may all be categorized either as attacks on premise 1 or premise 4. There are a myriad of responses to these premises, and so we will survey and evaluate them in the next chapter. There we will select the best of them all. We will then investigate whether it can be developed to cast doubt on or even overturn the zombie argument. If it can, naturalism has a chance. If it cannot, then naturalists must discover a new way to deal with the gap between the phenomenal and the psychological.

CHAPTER FOUR

Naturalist Responses to the Zombie Argument

The last chapter was about the problem that the possibility of zombies raises for naturalists. The last section examined the argument from zombies against metaphysical naturalism. It evaluated each premise and found them to be true, or at least possibly true. The goal of this chapter is to finish that evaluation. Premises 1 and 4 seem to have possible responses by which a naturalist could escape the argument. These responses are eliminativism, emergentism, and non-reductive physicalism. In examining these positions, it is necessary to return to information in the first chapter: namely, naturalists have three options for dealing with phenomena. The options are to eliminate the phenomena, to reduce it to more fundamental phenomena, or to posit the phenomena (or intermediate phenomena) as primitive. This chapter will discuss each of these possible responses in turn and examine how they play into the naturalist options and into the zombie argument. It will conclude that nonreductive physicalism is currently the best naturalist option on the table. Furthermore, the chapter will examine nonreductive physicalist views to discover the best option within that range, arguing it to be Derk Pereboom's recent work in the open possibility of introspective qualitative inaccuracy. Finally, the chapter will discuss Pereboom's work to see whether it actually succeeds in saving naturalism or if it also has flaws that render it susceptible to the Zombie Argument.

Responses to Premise 1

In the zombie argument, the first premise states that zombies are conceivable. This premise means that our working concepts of phenomenal and psychological consciousness must be correct in some basic way. If our concepts are mistaken, then the whole thought experiment fails to illuminate anything interesting about reality. It is useless to talk of the implications of how our concepts apply if these very concepts are mistaken. Naturalists who attack premise 1 attack our concept of phenomenal consciousness. These naturalists argue that our concept of phenomenal consciousness is fundamentally mistaken. This position is called eliminativism.

Eliminativism

According to Paul Churchland, “eliminative materialism is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience.”¹ Churchland dubs this commonsense conception “folk psychology,” and groups it with other folk theories of the past.² According to the eliminativist, one day we will stop talking in terms of folk psychology. Instead, we will discuss psychological consciousness: neurons, neurotransmitters, and the like. We will claim that we are “low on serotonin” rather than “feeling sad” and that will be *all that there is* to our discussion.

¹ Paul Churchland, “Eliminative Materialism and the Propositional Attitudes,” in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 568.

² *Ibid.*, 569.

We will not then discuss our feelings towards lacking serotonin, unless we also use psychological vocabulary. Our current personal descriptions of consciousness will be “left out” as anything meaningful, and we will turn instead to mechanistic descriptions. Churchland describes this process as one where we will need to “reconceiv[e] our internal states and activities”³ because of the misleading nature of our current concepts. In order for the elimination to be complete, future generations should conceivably even lack the concepts of folk psychology. Future generations will learn only the neurological concepts because only such concepts are true, and they also will be capable of eliminating the false theories from their concepts. According to the eliminativist, nothing will be missing in those future generations’ understanding of consciousness.

Eliminativism is not the same as reduction. Reduction puts up with our current understanding of phenomenology and says that it is reducible to the psychological concepts that science posits. Reduction holds that the psychological domain is the ultimate causal force, but that our expressions of phenomenology are not wrong in any real sense. Instead, reduction understands folk psychology to be a sort of shorthand for the complex detail of science. Eliminativism, on the other hand, does hold our folk psychology concepts to be fundamentally mistaken. On behalf of eliminativism, Churchland gives several reasons to motivate this bold claim.

First, he claims that our folk psychology often fails by being superficial or incomplete. It sometimes cannot explain or predict what is going on inside our minds, but instead these concepts (sleep, learning, intelligence, memory) remain a complete

³ Paul Churchland, *Matter and Consciousness*, (Cambridge, MA: The MIT Press, 2013), accessed online <http://site.ebrary.com.ezproxy.baylor.edu/lib/baylor/docDetail.action?docID=10744804>, 76.

mystery.⁴ Churchland thus describes folk psychology as “at *best* a highly superficial theory, a partial and unpenetrating gloss on a deeper and more complex reality.”⁵ Second, in the past, significant portions of folk psychology have been mistaken. If one looks on primitive cultures, consciousness is attributed to far too many things; this over-attribution results in people believing in gods in the water and air, and attempting to appease them by various sacrifices. Another example of folk psychology gone wrong might be mental illness. It used to be thought that instances of mental illness were instances of demon possession, but now we understand that these people have brain damage or brain malfunctions that produce abnormal behavior. According to eliminativists, these all are complete failures of folk psychology to explain correctly and sufficiently what occurs in the brain. Not only has folk psychology’s track record been very poor, but other folk theories in the past have also been radically misleading. There is no need to discuss these theories here, but at one time our folk theories included ghosts, alchemy, and other clearly false concepts. Based on our track record, then, there is no reason to expect us to have gotten *the most complex* concept correct when we have gotten the less complex concepts wrong.⁶ The overall failure of folk psychology (and folk theories in general) thus far makes it unlikely to achieve success now.

A second reason folk psychology cannot be maintained is that it fails to fit with the emerging scientific account of nature. Churchland asserts that folk psychology’s “intentional categories stand magnificently alone, without visible prospect of reduction to

⁴ Ibid., 78.

⁵ Churchland, “Eliminative Materialism and the Propositional Attitudes,” 571.

⁶ Churchland, *Matter and Consciousness*, 79.

that larger corpus” of the physical sciences.⁷ Furthermore, whereas science has been advancing through new discoveries and revised explanations, folk psychology is stagnant in its explanations. The current folk psychology is not much different from the time of the Greeks, and parts of it have been on the decline for centuries.⁸ We expect seamless integration of truth, and folk psychology does not fit with the natural sciences. Churchland admits that it is not impossible that folk psychology is reducible to psychological concepts, but it is extremely unlikely given the failures and stagnant history of the theory. Generally, people believe the natural sciences to be our greatest advances towards truth, and so it seems justifiable to conclude that folk psychology, if incompatible, must be false.

All of these reasons give us support for eliminativism and for the thesis that folk psychology, and therefore our phenomenal descriptions, are fundamentally mistaken. However, there are reasons to discount eliminativism as well. There are many challenges to eliminativism, but most fall short by begging the question. For brevity’s sake, we will not deal with these objections. Instead, we will look at a single objection which, even when answered, cripples the force of eliminativism.

The strongest objection to eliminativism attacks eliminativism’s assumption that folk psychology is a theory. Churchland, in his article “Eliminative Materialism and the Propositional Attitudes,” discussed folk psychology’s status in depth. He argued that folk psychology is aimed at creating a unifying theory about human behavior which explains and predicts behavior. Churchland then asserts that explanations presuppose laws, and

⁷ Churchland, “Eliminative Materialism and the Propositional Attitudes,” 572.

⁸ Ibid.

laws are the foundations of a theory. Only if folk psychology is a theory can one aptly apply its theories to people and minds other than one's own. The objection is this: it cannot be the case that folk psychology in its entirety is theoretical. Folk psychology gets its basis in immediate, first-hand experience. This observation is not the product of theoretical postulates, but rather the reliable sense experience that all of science is based upon. In order to discount folk psychology completely it must be the case that we cannot hold sense experience to be reliable, and thus cannot trust much else either.

Churchland's response is that we do *not* infer one another's mental states via "inductive analogy from the perilously isolated instance of one's own case." Instead, we use folk psychology as an explanatory hypothesis that even a Martian with a completely different psychology would be justified in applying to our behavior.⁹ However, it does not seem evident that a Martian *would* be justified to use folk psychology. Nor does it seem evident that we use folk psychology as an explanatory hypothesis. It does, however, seem evident that we have experiences and that these experiences cause us to do certain things. Consider the following: when my legs feel tired, I purposefully look for a chair to relieve the burden of weight from them. It is not the case that I notice myself in a chair and wonder why I am no longer standing. I need not *afterwards* posit an explanation like: I must have had tired legs, wanted to get relief from the tiredness, and believed that sitting in a chair would bring relief. Furthermore, when I see someone else sitting in a chair (especially if I observe him collapse into it and release a sigh of relief), I do not have to *posit* that his legs are tired. I am able to intuit or understand that he is tired because he acts just as I do when I am tired.

⁹ Paul Churchland, "Eliminative Materialism and the Propositional Attitudes," 569.

Churchland responds:

As well, introspective judgments about one's own case turn out not to have any special status or integrity anyway. On the present view, an introspective judgment is just an instance of an acquired habit of conceptual response to one's internal states, and the integrity of any particular response is always contingent on the integrity of the acquired conceptual framework (theory) in which the response is framed. Accordingly, one's *introspective* certainty that one's mind is the seat of beliefs and desires may be as badly misplaced as was the classical man's *visual* certainty that the star-flecked sphere of the heavens turns daily.¹⁰

This assertion has the significant consequence that introspection is never trustworthy.

Humans must wait for our conceptual framework to be developed and adapted by science before they may rely on introspection. But if this is the case, then science cannot rely on introspection as it investigates and forms the new framework. This seems to be impossible, as even scientists rely on introspection in everyday operation. Furthermore, it does not seem accurate to say that it is humans' introspective *judgment* that is causing the intuitions of the existence of the mind. Churchland treats introspection just like every other function of the mind, but this is exactly the question at hand. In fact, it seems that introspection is a more immediate form of observation than sensory knowledge.¹¹ If the judgments formed from introspection are what Churchland has a problem with, he is not properly justified in objecting to introspection itself. Perhaps he should reject the judgments formed in folk psychology, but to deny phenomenal experience altogether is a step too far. It appears that this is the step that results in eliminativism's failure, what

¹⁰ Ibid.

¹¹ Eric Schwitzgebel, "Introspection", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), ed. by Edward N. Zalta, <http://plato.stanford.edu/archives/win2012/entries/introspection/>.

Chalmers calls “denying the manifest.”¹² Conscious experience is too immediate to deny its existence completely, without explanation of how one could be so deceived. The naturalists must turn to another option if they are to hold a tenable and convincing position. If naturalists cannot reject premise 1 by eliminativism, then the premise must remain standing. Since premises 2 and 3 must be true by what has been discussed in previous chapters, the only option that remains for naturalists is to attack premise 4.

Responses to Premise 4

Premise 4 of the zombie argument states that if consciousness is nonphysical, then metaphysical naturalism is false. This premise assumes that metaphysical naturalism requires everything to be physical. As we discussed in chapter 1, this is a fundamental part of the definition of metaphysical naturalism, and is therefore not easily countered. There are several theories that consider a partially non-physical theory as naturalism, however. The first is epiphenomenalism, which denies that the mental is causally efficacious. Epiphenomenalists define metaphysical naturalism as requiring only causally efficacious things to be physical. The second theory is to posit a relation that explains the phenomenal without reducing it to the physical. This method, emergentism, consists of the various supervenience theories. The third theory is to posit a non-physical entity as another basic entity of reality. The non-physical entity would thus be another “basic fact” about what exists. This is non-reductive physicalism. Depending on the form of non-reductive physicalism, this view may become too much like dualism to be

¹² David J. Chalmers, “Consciousness and its Place in Nature,” in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 251.

considered consistent with metaphysical naturalism. This section will discuss these options briefly and focus on the best position available to the metaphysical naturalist.

Epiphenomenalism

It is common for naturalists to allow for causally inefficacious non-physical things in their account of reality. As mentioned in the first chapter, these things are typically numbers, sets, or things similar. These abstract features of the world do not have causal powers and so make little difference to the naturalist theory. Obviously this is an appealing way for naturalists to deal with mental phenomena as well.

Frank Jackson is a naturalist who holds that the causal powers of qualia are not apparent. Instead, he argues that the common reasons for discounting epiphenomenalism are all surmountable.¹³ The first reason epiphenomenalism is often dismissed has to do with the obvious nature of the efficacy of qualia. It is just apparent that the “hurtfulness of pain is partly responsible for the subject seeking to avoid pain.”¹⁴ Jackson responds with a Humean objection that anything can seem to cause anything. All that is needed to overturn this intuition is an underlying causal theory that makes both the hurtfulness of pain and the seeking to avoid pain effects of the same hidden cause. A second objection used against epiphenomenalists is natural selection. It seems that if natural selection were to select for creatures with conscious experience, conscious experience must have some evolutionary advantage, and thus, it must be causally efficacious. Jackson here uses the same response. All that is needed for epiphenomenalism is an underlying

¹³ Frank Jackson, “Epiphenomenal Qualia”, in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 276-277.

¹⁴ *Ibid.*, 276.

process that *is* causally advantageous, and that has the side effect of conscious experience. Then conscious experience will have been selected for not because of its usefulness, but because of the usefulness of the process sufficient to bring it about. Finally, Jackson addresses the issue of other minds. If I see someone jumping up and down and screaming “ouch!” then I am immediately capable of coming to the conclusion that he is experiencing pain. Anti-epiphenomenalists argue that the reason we can know their experience is because their behavior is a direct result of it. Epiphenomenalists like Jackson argue that we are actually reasoning from the man’s behavior of pain to his physical cause of pain, and then from that physical cause of pain to the side effect of conscious experience of pain. Thus the reasoning is valid each time, but we wrongly attribute causal powers to experience rather than the physical process causing the experience.

The last two of the three defenses are plausible, but still this theory of consciousness is likely the one most easily dismissed, if only due to intuition. Epiphenomenalism runs into complex problems of self-defeat,¹⁵ but more immediately it is just strongly counterintuitive. Although it is *possible* that there is an underlying physical mechanism doing the work of the causal processes, we do not know what it is. Our experience tells us that we do not hold our hands to stovetops that are on because it is unpleasant, it hurts. To argue that we are deceived in something so immediately obvious requires a good reason. The epiphenomenalists’ only reasons are those that assume

¹⁵ For example, epiphenomenalism seems to entail the inefficacy of reasons to produce belief. But if this is the case, then there is no reason to believe epiphenomenalism. (See William Robinson, "Epiphenomenalism," in *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), ed. by Edward N. Zalta, <http://plato.stanford.edu/archives/sum2012/entries/epiphenomenalism/>.)

naturalism. Additionally, to posit inefficacious events is to posit things extremely unlike anything else in existence. These mental “danglers”¹⁶ would be strange, and would require laws as odd as they are. They would be consistently emerging but would have no effect whatsoever. Instead, consciousness would be like a background noise accompanying the physical workings of the universe, except that the noise would fail to *in any way* affect the physical realm. Such posits are rather unbelievable, and it is for this reason even most naturalists do not view epiphenomenalism as a genuine explanation.

Emergentism

Emergentism’s supervenience theories are similar to reductionist theories in many ways. They argue against premise 4 by saying that consciousness, although nonphysical in the traditional sense, emerges or arises from the physical and so counts as physical in the relevant ways. The mental and physical are said to be non-identical, and yet there is still a dependence relation. This relation can be formalized a bit further as low-level facts determining high-level facts.¹⁷ This dependence without identity relation is called supervenience. An example might be biological facts supervening on facts from physics: anything with the same physical properties thereby is determined to have the same biological properties.¹⁸ Emergentists claim that the mental supervenes on the physical. Philosophers who favor reductionist theories but also are impacted by the force of Kripke’s arguments against phenomenal-psychological identity often end up

¹⁶ Herbert Feigl in J.J.C. Smart, “Sensations and Brain Processes,” in *Philosophy of Mind: Classical and Contemporary Readings*, ed. by David J. Chalmers, (New York: Oxford University Press, 2002), 61.

¹⁷ David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, (New York: Oxford University Press, 1996), 32-33.

¹⁸ *Ibid.*, 33.

supervenience theorists. This is because supervenience theories claim the dependence of phenomenal on the psychological without the identity entailed in reductionism.

This position is unstable, however. Jaegwon Kim argues that it must collapse back into reductionism or become epiphenomenalism.¹⁹ The reason for emergence's instability is the causal nature of the mental. Kim takes it to be true that either the mental states have causal powers or they do not. If they do have causal powers, then, given the causal closure of the physical, there is nothing to distinguish the mental from the physical, and so they are identical. If they are identical, then emergence counts as a reductionist theory. If the mental states do not have causal powers, then the mental may be distinct from the physical, but causally inefficacious. This means that emergence is an epiphenomenalist theory. Either way, emergence is not an explanation that naturalists can maintain while being rationally consistent.

Non-Reductive Physicalism

There is only one position left for the metaphysical naturalist, that of non-reductive physicalism. This move counters premise 4 because it comes up with an example of how a nonphysical mental event could fit into a physicalist, naturalist theory. Non-reductive physicalism's strength lies in the same areas as those of emergence: a physical thing has two sets of properties (mental and physical), where the mental is dependent upon, but not identical to, the physical. Unfortunately for philosophers like Derk Pereboom, non-reductive physicalism is subject to many of the same critiques as

¹⁹ Jaegwon Kim, "The Myth of Nonreductive Materialism," in *Proceedings and Addresses of the American Philosophical Association*, Vol. 63, No. 3 (Nov. 1989), 40-42.

emergence. The position does not seem to be tenable, because it seems to collapse into either reductionism on the one hand or epiphenomenalism on the other.

Derk Pereboom in his article "Robust Nonreductive Physicalism" states that the mental has causal powers, the physical has causal powers, and these two powers are not reducible to one another.²⁰ They are not reducible because the mental and physical have neither type nor token identities. Nevertheless, these causal powers are constitutionally coincident, meaning that they are constituted by "the same stuff" without being identical.²¹ Pereboom argues that being constitutionally coincident makes for no causal competition, because constitutional coincidence is only so different from identity. An example might be a statue that is composed of granite. The statue is one thing and the granite another. If someone took a jackhammer to the granite and ground it into a million pieces, the statue would no longer remain but the granite would. This means that the statue and granite are not identical. Yet, to say that the statue fell over and cracked the floor is no different than saying the granite did so, *because* they are made of the same material stuff. The differences between the constitutional coincidence and identity are therefore not significant enough to expect a difference in causal competition.²²

In the end, this position sounds really wonderful for the naturalist. It seems as if it bridges the gap between the mental and physical without losing the foundation of reality being physical. However, there are problems with the view. The first problem is that the "base" of the physical seems to be an arbitrary distinction. If the physical and mental are

²⁰ Derk Pereboom, "Robust Nonreductive Materialism," in *Journal of Philosophy* 94 (2002), 499-531.

²¹ Ibid., 505.

²² Ibid.

causally efficacious in their respective domains, why assume the physical is causing the mental rather than the mental causing the physical? Perhaps this is a bit of an outlandish claim, but nothing logically rules it out on non-reductive materialism/physicalism.²³ The second problem with the view is that it goes against the causal exclusion principle. If the physical is completely causally efficacious, and it is sufficient for the occurrence of mental events, it seems that either the mental events do not have their own causal efficacy or there is overdetermination occurring. If mental events remain causal danglers, then they are leftovers while different (physical) phenomena do the real work. If overdetermination is occurring, then causal relations as we know them are altered significantly. This alteration is too extreme and effects too much of our concepts of reality to be worth the price. Therefore nonreductive materialism also fails to adequately respond to the zombie argument.

It seems to be that naturalists have no sufficient response to premise 4. This means that all of the premises are true for the valid argument presented in chapter 3. Since it is valid, this means that the conclusion follows and must therefore be true as well. Naturalism about conscious experience is false.

Conclusion

This thesis began with a discussion of what naturalism is, what consciousness is, and what the zombie argument is. This chapter applied the definitions of naturalism, phenomenal consciousness, and psychological consciousness to the zombie argument in order to find whether the naturalists could have any satisfactory answer to the zombie argument. If the zombie argument is sound, then it refutes metaphysical naturalism. This

²³ See Galen Strawson, "Realistic Monism: Why Physicalism entails Panpsychism," in *Journal of Consciousness Studies*, 13, (2006).

chapter explored several responses of a naturalist which each attempt to defeat the zombie argument. All of these responses are too deficient as they currently stand to be maintained consistently. While not every response was discussed in the fullest detail, enough was made clear that the principles applied here could be applied to any other responses currently given. This thesis therefore concludes that naturalism about the phenomenal aspects of mind is false. Naturalists must propose new explanations of the mind or else become non-naturalists.

BIBLIOGRAPHY

- Allen, Colin and Trestman, Michael. "Animal Consciousness", *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), forthcoming <http://plato.stanford.edu/archives/sum2014/entries/consciousness-animal/>.
- Armstrong, D.M. "The Causal Theory of the Mind." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.
- Armstrong, D.M. "The Nature of Mind." In *The Nature of Mind and Other Essays*. New York: Cornell University Press, 1981.
- Blackburn, Simon. *Ruling Passions*. New York: Oxford University Press, 1998.
- Brown, James Robert and Yiftach Fehige. "Thought Experiments." In *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition), edited by Edward N. Zalta, <http://plato.stanford.edu/archives/fall2011/entries/thought-experiment/>.
- Chalmers, David J. "Consciousness and its Place in Nature." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. Oxford University Press: New York, 2002.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.
- Churchland, Paul. "Eliminative Materialism and the Propositional Attitudes." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.
- Churchland, Paul. *Matter and Consciousness*. Cambridge, MA: The MIT Press, 2013, accessed April 2, 2014, <http://site.ebrary.com.ezproxy.baylor.edu/lib/baylor/docDetail.action?docID=10744804>, 76.
- Derk Pereboom, "Robust Nonreductive Materialism," In *Journal of Philosophy* 94 (2002).
- Jackson, Frank. "Epiphenomenal Qualia." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.

- Kim, Jaegwon. "Reduction, problems of." In *Routledge Encyclopedia of Philosophy*, ed. E. Craig, (London: Routledge, 1998), accessed November 13, 2013, <http://www.rep.routledge.com/article/Q089>.
- Kim, Jaegwon. "The Myth of Nonreductive Materialism." In *Proceedings and Addresses of the American Philosophical Association*, Vol. 63, No. 3 (Nov. 1989), 40-42.
- Kirk, Robert. "Zombies." In *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), edited by Edward N. Zalta, <http://plato.stanford.edu/archives/sum2012/entries/zombies/>.
- Kripke, Saul A. Excerpt from *Naming and Necessity*. In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.
- LaPorte, Joseph. "Rigid Designators." In *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition), edited by Edward N. Zalta, accessed April 15, 2014, <http://plato.stanford.edu/archives/sum2011/entries/rigid-designators/>.
- Layman, C. Stephen. *Letters to Doubting Thomas: A Case for the Existence of God*. New York: Oxford University Press, 2007.
- McGinn, Colin. *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books, 1999.
- Nagel, Thomas. "What Is It Like to Be a Bat?" In *The Philosophical Review*, 83: 4, (1974), accessed February 7, 2014: <http://www.jstor.org/stable/2183914>.
- Plantinga, Alvin. "Methodological Naturalism?" In vol 1. of *Facets of Faith and Science*, edited by Jitse M. van der Meer. New York: University Press of America, Inc., 1996.
- Putnam, Hilary. "The Meaning of 'Meaning.'" In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.
- Rea, Michael C. *World Without Design: The Ontological Consequences of Naturalism*. Oxford: Clarendon Press, 2002.
- Rey, Georges. "Eliminativism." In *Routledge Encyclopedia of Philosophy*, edited by E. Craig. London: Routledge, 1998, accessed November 14, 2013, <http://www.rep.routledge.com/article/W012>.

- Robinson, William. "Epiphenomenalism." *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), edited by Edward N. Zalta, <http://plato.stanford.edu/archives/sum2012/entries/epiphenomenalism/>.
- Rosenberg, Alex. *The Atheist's Guide to Reality: Enjoying Life without Illusions*. New York: W. W. Norton & Company, 2011.
- Schwitzgebel, Eric. "Introspection." *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), edited by Edward N. Zalta, <http://plato.stanford.edu/archives/win2012/entries/introspection/>.
- Searle, John. "Philosophy and the Basic Facts." In *Freedom and Neurobiology*. New York: Columbia University Press, 2007.
- Smart, J.J.C. "Sensations and Brain Processes." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. New York: Oxford University Press, 2002.
- Strawson, Galen. "Realistic Monism: Why Physicalism entails Panpsychism." In *Journal of Consciousness Studies*, 13, (2006).
- Unger, Peter. "Free Will and Scientiphicalism." In *Philosophy and Phenomenological Research* 65 (2002): 1, accessed September 21, 2013, <http://philosophy.fas.nyu.edu/docs/IO/1179/freewill.pdf>.
- Van Inwagen, Peter. in "Metaphysics." In *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/metaphysics/#NatMet>.