Conceptualizing Consciousness: What Must We Explain?

by

Gage Goulet

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Carleton University

Ottawa, Ontario

© 2022

Gage Goulet
ABSTRACT

Explaining consciousness seems like an incredibly daunting task. I suspect the reason for this is because of what our explanatory theories of consciousness are expected to involve and what they are expected to look like. With the hope of deflating what is apparently required from an explanation of consciousness, I show that (i) we can shed peripheral and unproductive explanatory tasks from the enterprise of explaining consciousness, and (ii) demonstrate a means of developing much simpler explanatory theories of the phenomenon. In completing the first, I argue that our explanatory theories need not serve as interpretive tools to understand what experiences are like, and that we need not expand our scientific inquiry in search of “extra ingredients”. As for the second, I motivate a semantic view of theories, wherein our explanations of consciousness may be agnostic about potential answers to other persisting philosophical problems such as the mind-body problem.
I wish to sincerely thank all of those who have contributed to and supported the production of this project. I truly would not have been able to do it without you:

My supervisor, Andrew Brook, for his mentorship, incredible feedback, support, and patience throughout. The skills and insight that you have imparted onto me through our time together on this project are invaluable.

My mother, Angela, my partner, Danielle, and my friends, for being my biggest cheerleaders throughout the process and tolerating my passionate philosophical ramblings.

My colleagues within my cohort and the professors within my department, for inspiring me through insightful discussions and lectures.

Thank you.
# TABLE OF CONTENTS

Title Page: “Conceptualizing Consciousness: What Must We Explain?” ....................i
Abstract................................................................................................................................ii
Acknowledgements...........................................................................................................iii
Table of Contents...........................................................................................................iv

Chapter 1: Introduction: Concepts of Consciousness & The Inflated Approach.......1
  1.1. Conceptualizing Consciousness...........................................................................2
  1.2.0. The Inflated Approach....................................................................................8
  1.2.1. First Facet: We Must Explain What Experiences are Like............................8
  1.2.2. Second Facet: The Extra Ingredient..........................................................10
  1.2.3. Third Facet: Emphasizing Metaphysical Explanations................................14
  1.3. Summary & Plan...............................................................................................15
  1.4. Conclusion.......................................................................................................17

Chapter 2: Describing What It Is Like vs. Explaining Consciousness.........18
  2.1. The Question: What Is It Like?.........................................................................21
  2.2. Description vs. Explanation..............................................................................23
  2.3. The Failure of Nagel’s Question.......................................................................24
  2.4. Further Concerns & Objections......................................................................39
  2.5. Conclusion.......................................................................................................40

Chapter 3: Without The Extra Ingredient; An Indirect Approach...............42
  3.1. The Hard Problem & The Extra Ingredient.....................................................44
  3.2. Anti-Cognitive Approaches.............................................................................47
  3.2.1. The Form of The Arguments.......................................................................47
  3.2.2. A Priori Thought Experiments.....................................................................50
  3.3. Denying Extra Ingredients & Motivating Cognitive Approaches..................61
  3.4. Conclusion.......................................................................................................67

Chapter 4: Metaphysical Theories & Agnostic Explanations of Consciousness..69
  4.1. Metaphysical Theories of Consciousness.......................................................71
  4.1.1. The Problem with Metaphysical Theories of Consciousness.......................74
  4.2. Agnostic Explanations of Consciousness.......................................................77
  4.2.1. Levels of Explanation and the Semantic View of Theories.........................78
  4.2.2. Agnosticism about the Metaphysics of Consciousness.............................84
  4.3. Conclusion.......................................................................................................90

Chapter 5: Conclusion...............................................................................................92

Works Cited.................................................................................................................96
CHAPTER 1
Introduction: Concepts of Consciousness & The Inflated Approach

Consciousness seems to pose a unique problem for researchers. When it comes to studying the mind, our modern empirical methods are incredibly young, and our concepts and theories seem to frame it in a variety of conflicting ways. For the former, it is not yet clear that our empirical methods actually track consciousness, or instead, track some other neuronal or cognitive phenomenon that is not itself consciousness. This leads to a variety of epistemological problems for our research. For the latter, the conceptual work surrounding consciousness has become oversaturated with definitions of the phenomenon, varieties of it, and more; leading to a space where it is common to talk past one another, enter into verbal disputes, conflate definitions, or find oneself in a gridlock or standoff of conceptual analysis. Importantly, these problems seem to bolster our common intuitions regarding consciousness—that it is a phenomenon like no other, and that it is one of the most intractable phenomena to date.

Perhaps more important than these issues is the fact that it is not yet clear what our explanatory theories of consciousness need to involve, what they need to explain, or what form they need to take. I suspect that this problem, itself, is the main reason why the phenomenon appears to be so difficult to explain. For instance, there appears to be a variety of influential works within the literature on consciousness which have influenced

---

1 See Levine (1983) for more discussion on what is called the explanatory gap. This is, arguably, the most important and influential argument for there being an epistemic gap of this sort within the literature. While other examples of epistemic gaps with regard to consciousness can be given, Levine’s (1983) paper captures the essence of the worry in a precise way.

2 See Balog (2008) for discussion on the ‘gridlock’, where arguments and concepts regarding consciousness seem just as plausible, and we do not seem equipped to be able to determine which one is more accurate. I prefer the term ‘standoff’, however, as ‘gridlock’ falsely implies that no movement is able to be made.
how we are expected to approach the phenomenon, and have inflated what is required from an explanation. While these works are primarily philosophical, they have had a significant impact on the sciences which study consciousness, as the field is incredibly interdisciplinary. This chapter will serve to characterize these influential works under what I am (for convenience) calling the inflated approach (hereby IA) to consciousness and briefly explicate what I take to be its most problematic facets. Chiefly, each facet of the inflated approach either leads our explanatory inquiry down a garden path, or expects our explanations of consciousness to involve much more than what a potentially successful theory of consciousness needs to involve. Thus, they make our job of explaining consciousness much more difficult than it needs to be. So, the purpose of this chapter is to outline the IA (and its features) before I engage in the task of problematizing each facet throughout subsequent chapters.

The structure of this chapter will look like so: In §1.1, I will discuss how the literature surrounding consciousness seems unstable, and yet, I will attempt to pinpoint what is taken by most to be the explanandum (phenomenon in need of explanation) of consciousness and why it seems like such a mystery. Then, §1.2 will introduce and characterize the IA to consciousness and discuss its various facets. Following this, §1.3 will summarize, identify some positive aspects of the IA that I will accept throughout the thesis, and lay out my plan of attack in a bit more detail. And finally, §1.4 will make some concluding remarks.

1.1. Conceptualizing Consciousness
Consciousness, as a term, seems confused, unstable, and anything but precise. Even though Chalmers (1996) may be correct in stating that “there is nothing we know about more directly than consciousness”, Block (1995) argues that ‘consciousness’ is a hybrid concept, or a mongrel concept, and Churchland (1983) predicts that certain terms like ‘consciousness’ will “lose their integrity and fall apart” much like other antiquated terms that have since fallen to the wayside in light of modern empirical and conceptual advances. Even though consciousness seems to be what we are most intimately acquainted with, its theoretical use as a term and as a concept may be challenging to pin down or get right, and might even be entirely misplaced. Why are there such differing views about consciousness and how we seem to understand it as a concept? Perhaps we do not yet possess a stable conceptual grasp of consciousness, as there seems to be differing concepts of it which stand separate from one another, at odds, and potentially refer to different phenomena.

Block (1995), for example, has identified two distinct phenomena related to consciousness, and distinguishes them from one another. Those two phenomena, however, are often conflated, or one is sometimes eliminated entirely (in favor of the other) by specific philosophers or scientists. Namely, the distinction is between \textit{phenomenal consciousness}, and \textit{access consciousness}. The former has to do with Nagel’s (1974) notion of “what it is like”, where “it” serves as a placeholder for a particular subjective character or quality of a given experience; whereas the latter has to do, roughly, with the availability of information within cognitive systems. The primary and most troubling problem is that when researchers discuss consciousness, they often seem to conflate them and misidentify the phenomenon at hand. While Block (1995) has
expressed his own concerns related to this sort of conflation, Chalmers does the same, remarking that:

The ambiguity of the term consciousness is often exploited by both philosophers and scientists writing on the subject. It is common to see a paper on consciousness begin with an invocation of the mystery of consciousness, noting the strange intangibility and ineffability of subjectivity, and worrying that so far we have no theory of the phenomenon … In the second half of the paper, the tone becomes more optimistic, and the author’s own theory of consciousness is outlined. Upon examination, this theory turns out to be a theory of one of the more straightforward phenomena of reportability, of introspective access, or whatever. At the close, the author declares that consciousness has turned out to be tractable after all, but the reader is left feeling like the victim of a bait-and-switch. (Chalmers 1995, 202)

Rosenthal (2009, 160) adds that a similar thing is done when considering his own distinct types of consciousness, as “failing to distinguish creature, transitive, and state consciousness can also have important consequences for theories of consciousness”. And thus, there is often a worry that researchers—in discussing, explaining, or theorizing about consciousness—conflate distinct concepts of consciousness or misidentify the phenomenon entirely.
Black (2017) identifies Block as a supporter of what he calls the conflation view, and Rosenthal as a supporter of the ambiguity view. The former is the view that there is one term (“consciousness”) which is used by people to talk about the same sort of phenomenon, but that it is then conflated with a variety of underlying concepts. The latter, too, worries about conflation; though defenders of the ambiguity view “aren’t committed to the view that there is an entrenched usage of the term that systematically conflates different concepts of consciousness” (Black 2017, 31). Rather, Black holds that they are concerned with the usage of the term being rendered ambiguous by the sheer number of concepts with which it is associated.

Importantly, through Black (2017), we can spot an important problem in the concern that Block (1995) and Chalmers (1995) have previously raised with regard to feeling like the victim of a bait and switch when researchers conflate and exploit the ambiguity of the term ‘consciousness’. Rather than researchers conflating concepts of consciousness, identifying a different explanandum, and thus, missing the apparent target of consciousness, Black (2017) argues that what researchers are doing is using different concepts as explanans for (roughly) the same explanandum that the vast majority of researchers (including Block and Chalmers) share. Thus, when a researcher appeals to access consciousness after noting that they will explain phenomenal consciousness, they are not merely conflating the two concepts and missing the target of explanation. Instead, they are using access consciousness as a way to explain the phenomenon of phenomenal consciousness—which is often taken to be the target explanandum and the mystery at the heart of the phenomenon. Importantly, “charity of interpretation, I want to suggest, favours this [the above] alternative interpretation” about using different concepts of
consciousness as explanations for other problems or phenomena (Black 2017, 40). When appealing to access consciousness, integrated information, global access, or whatever else it may be, “none of this is reason to think that they [scientists and researchers] are confused about the fact that the explanans of their theories … is conceptually distinct from the explanandum, consciousness. This is to be expected: H2O is identical to water, but the concepts of H2O and water are logically distinct” (Black 2017, 40). In other words: researchers, when talking about consciousness, will typically use the same sorts of examples. This is a good indication that they are using the term with the same extension, and so talking about the same thing (but, of course, there may be exceptions.) Thus, when invoking different concepts and phenomena (other than the target explanandum, consciousness, itself) we should—in the interest of being charitable—not confuse this as misinterpreting the target explanandum, as Block and Chalmers have done.

Furthermore, while Block (1995) has done a lot of work to defend the view that varying concepts of consciousness are often conflated, Black (2017) rejects it. Rather than accept any view that holds that there are distinct concepts of consciousness which are often conflated (be it the conflation view or the ambiguity view), Black observes that there are commonalities (some of which I will show below) regarding the usage of the theoretical term from a vast cast of researchers in both philosophy and the sciences. Because of this, there seems to be a good understanding of the extension of the term (i.e., what sorts of concepts it ranges over). Thus, Black (2017) holds the view that the term ‘consciousness’ is a \textit{cluster term}.\footnote{Black (2017, 46) explains that cluster terms are those “that apply to what they do in virtue of criss-crossing similarities between the members of their extensions — what Wittgenstein calls ‘family resemblances’.”} Out of Black’s (2017, 45) large list of features, two...
stand out as the least controversial and most important, and of which I will be accepting, throughout, as primary features of consciousness:

(1) **Awareness**: Conscious experience is what one is immediately aware of from the first-person perspective when one is awake, aware, and properly responsive to stimuli.

(2) **Subjectivity**: Conscious experience has subjective, phenomenal character.

While Black’s list is much more expansive and includes a number of additional features, many of such features are trivial, unclear, or peripheral to the concerns of the IA and of this project. It is the conjunction of these two features (1-2) that seem to be the target explanandum for most researchers—including proponents of the IA (Nagel 1974, Levine 1983, and Chalmers 1995, most prominently).

Black’s view is attractive for the study of consciousness, as the interdisciplinary and empirical study of the phenomenon is incredibly young, and we are not yet in a place to determine which features of the term are necessary or sufficient for consciousness. It is thus important to not take anything related to consciousness for granted, or hold (without reason) that a particular feature is (or is not) a part of the phenomenon at all. However, Black’s view is simply one of observation—it tracks the common ground that is found between most researchers across philosophy and the sciences.
Within philosophy, mainly, there has been an inflated approach (the IA) to consciousness; one which is particularly interested in the two uncontroversial features listed above, from Black. However, the IA demands that these features of the phenomenon be investigated and understood in a certain way. First, it holds that our target of inquiry is to explain what it is like to experience certain experiences. Second, it holds that consciousness cannot be investigated and understood in specific ways—namely, through many of our current scientific methods—and thus requires a radically different scientific approach with an expanded ontology. And third, it seems to emphasize metaphysical explanations which relate to the mind-body problem in a way which targets the substructure of consciousness and its place in nature. To explain, I will now focus on the IA, and outline these facets in some detail. However, in the interest of avoiding redundancy, in this chapter, I am merely introducing the ideas and arguments that will appear in more detail in further chapters. In each chapter that deals with its respective facet, I will explicate the details of the facet in greater detail, and provide my arguments against the features within them that I find problematic.

1.2.0. The Inflated Approach

1.2.1. First Facet: We Must Explain What Experiences are Like

Nagel, in an extremely influential (1974) paper titled “What Is It Like to Be a Bat?” attempts to define consciousness, and identify the necessary and sufficient conditions for an organism to be conscious. Precisely, an organism has conscious mental states “if and only if there is something that it is like to be that organism—something it is
like for the organism” (Nagel 1974, 436). While there may be something it is like to be a bat, there is not anything it is like to be a rock. If this is true, then bats possess conscious mental states while rocks do not. The trick, for Nagel, is to identify what it is like to have the conscious mental states of a bat—to identify the subjective, qualitative character of experience that bats have. Thus, if we are not understanding what it is like to have a particular subjective character of experience, then we are not understanding more about consciousness. For Nagel, to explain anything else is to make a category mistake. And thus, Nagel (1974) begins his paper with a critique of the earlier reductive identity theories that aimed to identify consciousness with particular brain states. About them, Nagel (1974, 435) claims that their “discussions of the problem [regarding consciousness] give it little attention or get it obviously wrong”. This is because Nagel’s competitors at the time were attempting to explain consciousness in a purely objective way, whereas Nagel believes the phenomenon to be one which can only be known subjectively. Thus, when it comes to knowing what we need to know about consciousness (what it is like to be X), the problem becomes much more challenging from a third person point of view.

Perhaps most noteworthy is the fact that Nagel has given rise to a locution (‘what it is like’) which has been accepted as a term that seems to get at the heart of what we mean by consciousness. In addition, the locution seems to point toward what it is about consciousness that requires explanation. However, there is a fundamental problem with Nagel’s approach which I aim to attack within the second chapter of this thesis: Namely,
it is the view that our inquiry into consciousness must begin with, or include, an understanding of what it is like to experience particular experiences. Questions about what a particular experience is like to experience are satisfied by answers which only provide description—not explanation—and thus such questions, and their answers, are not all that productive in terms of explaining the phenomenon. I will argue that even if we understood everything about what it is like to have the experiences of a bat, knowledge of such facts is underwhelming in terms of being informative and explanatory. And thus, Nagel’s line of inquiry and expectations for what our theories must set out to answer sends us down a garden path.

1.2.2. Second Facet: The Extra Ingredient

Chalmers (1995, 201), too, identifies this ‘what-it-is-likeness’ as the key feature of the hard problem—in that we are attempting to explain consciousness, where “in this central sense of consciousness, an organism is conscious if there is something it is like to be that organism, and a mental state is conscious if there is something it is like to be in that state”. Apart from this, it is important to get a grasp of what the ‘mystery of consciousness’ seems to be, as it is the characteristic feature of the IA’s explanandum. Primarily, the mystery of consciousness has to do with how and why phenomenal experience happens at all.⁵ Chalmers (1995, 201) explains that “when we think and perceive, there is a whir of information-processing, but there is also a subjective aspect

---

⁵ This notion can be seen in Levine (1983), and heavily inspires Chalmers’s (1995) work.
… This subjective aspect is experience”. Here are some examples of what Chalmers takes to be conscious experiences:

When we see, for example, we experience visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought. What unites all of these states is that there is something it is like to be in them. All of them are states of experience. (Chalmers 1995, 201)

The proponents of the IA typically use the terms ‘consciousness’ and ‘experience’ to be co-extensions—in that they are referring to the same phenomenon of subjective mental qualities. Chalmers, in particular, explicitly states that a method of avoiding confusion is to “reserve the term ‘consciousness’ for the phenomena of experience” (Chalmers 1995, 202).

Furthermore, Chalmers (1995) makes a clear distinction between explananada: the easy problems of consciousness are argued to be distinct from the hard problem. The former problems are “vulnerable to explanation in terms of computational or neural mechanisms … [and] in each case, an appropriate cognitive or neurophysiological model can clearly do the explanatory work” (Chalmers 1995, 201). Easy problems include: the
ability to report mental states, a system’s access to its own internal states, attentional mechanisms, and more (Chalmers 1995, 200). Importantly, answers to some of the easy problems might shed light on how we can solve the hard problem, but the hard problem itself is entirely distinct from the easy problems and cannot be wholly solved by way of (i) solving the easy problems, and (ii) via the same methods used to solve the easy problems. Argumentation for this can be seen through various a priori thought experiments contra physicalism. Chalmers (2002a) illustrates his famous conceivability argument with zombies: beings that are physical, functional, and cognitive duplicates of us, but who lack consciousness. Additional arguments of the sort stem from Jackson’s (1982) knowledge argument, and Levine’s (1983) explanatory gap. I will discuss the conceivability argument, other thought experiments, and the view that consciousness is a distinct problem from the easy problems all in more detail within the third chapter of this thesis.

The hard problem, as Chalmers famously describes it, “is the problem of experience” (Chalmers 1995, 201). Fundamentally, it is the question of how and why “physical processing give[s] rise to a rich inner life at all?” (Chalmers 1995, 201; see also Levine 1983). Because such an answer to the hard problem is argued to be so distinct from any answer that can be given via the easy problems, Chalmers (1995, 200) advocates that we confront the phenomenon directly, in a way which is separate from the easy problems. The implied hope is to be able to pinpoint and explain consciousness, without the risk of simply explaining some other phenomenon which relates to consciousness, but is merely an easy problem.
Via zombies (and other arguments), Chalmers (1995) argues that we can never capture consciousness as some combination of cognitive activities, and thus, consciousness has to be an ‘extra ingredient’ over and above cognition (and the easy problems, more generally). Thus, even when all of the easy problems have been solved, the hard problem of consciousness will apparently remain a separate task that is left over. To solve it, there must then be radically different explanans involved (Chalmers 1995), which is not featured in anything related to the easy problems. Identifying such an ingredient will be to identify a potential explanation for the hard problem alone. For proponents of the IA, what is most important here is that experience is more than any other thing, and made wholly distinct. Nagel (1974), Levine (1983), and Chalmers (1995) (amongst others) all share the same concern, wherein any other cognitive, neurophysiological, biological, or physical event can (theoretically or conceivably) happen absent of consciousness. And thus, consciousness is not wholly constituted by any of the additional phenomena associated within these domains.

To telegraph what I take to be problematic features of what Chalmers is arguing, and which I aim to take up within this thesis: First, the view that there exists an ‘extra ingredient’ left over which serves as an explanans for consciousness. Second, that an explanation for consciousness cannot be found via the easy problems or via the same methods used to answer them, and that we ought to confront consciousness *directly* (separately from cognition, function, or anything related to the easy problems). I will argue that the justification for the ‘extra ingredient’ is dubious, and that, through Occam’s Razor, we have no good reason (at present) to accept a radically inflated ontology and methodology.
1.2.3. Third Facet: Emphasizing Metaphysical Explanations

While the first facet involves a particular line of inquiry and a demand with regard to what our explanations ought to do, and the second facet has to do with an inflated ontology and methodology, this third facet has to do with an inflated sense of ontological commitment within our explanatory theories, what they must involve, and an overlapping of philosophical problems. About the hard problem, Chalmers (1995, 201) says that “if any problem qualifies as the problem of consciousness, it is this one”: the problem of how and why conscious experiences arise at all. However, certain theories of consciousness, which Chalmers (2002b) calls *metaphysical theories of consciousness*, focus too strongly on solving the mind-body problem in their attempts to explain consciousness. This leads them to be rather uninformative in terms of how consciousness is caused or how it functions, and commits them to specific ontological conclusions regarding the nature of consciousness and its micro-constituents, which renders them trapped in a peripheral philosophical standoff about the mind-body problem.

Thus, while many theories have attempted to give a convincing answer to the mind-body problem in order to solve the hard problem (and/or other problems related to consciousness), I will advocate for the notion that we may leave our explanations of consciousness ontologically uncommitted to any positions related to the mind-body problem. This has the benefit of being maximally informative about what needs to be explained about consciousness, while avoiding a major, and peripheral, philosophical issue (the mind-body problem) which diverts our explanatory progress toward the
phenomenon of consciousness. To do this, I advocate for a semantic view of theories, wherein we can remain agnostic about what sorts of properties or entities our explanatory language seems to commit us to. We may do this by emphasizing the models that our explanations create, and by using undefined variables in place of the nature of consciousness and its micro-constituents, all while remaining just as explanatorily informative about problems specifically related to consciousness. With this approach, we may make explanatory progress on consciousness while absent a solution to the mind-body problem.

1.3. Summary & Plan

Various aspects of the IA seem to have been accepted by a wide range of researchers, and the facets of the IA are just as widely discussed now within the literature as they were when they were first published. For example, the need to know what particular experiences are like still seems a fundamental problem to investigate if we wish to understand consciousness, the hard problem and the identification of an extra ingredient remain as pertinent and problematic as ever, and metaphysical explanations regarding the nature of consciousness remain all the rage within specific circles of research that are working to explain consciousness.

Importantly, much needs to be said about which features I wish to accept from the IA, and which ones I will refute: First (in chapter 2), I accept that the primary target of explanation seems to be phenomenal consciousness—the notion of there being subjective experiences of which there is something it is like to undergo them. However, I will deny
that our explanations of consciousness require a description of what particular experiences are like to experience, and (as Nagel (1974) argues) that this is the target of our explanation. Second (in chapter 3), I do not wish to wholly deny the existence of a hard problem. There does indeed remain a problem of how our physical systems become conscious. However, in the chapters to come, I will argue that it is not clear that an extra ingredient (outside of the easy problems) is what is required to answer it, and that the hard problem does not require a direct (or distinct) approach which differs from any other cognitive approach. And thus, I will eventually advocate for an indirect approach via cognitive methods and functions. And finally (in chapter 4), I accept that metaphysical explanations may do some significant work for the mind-body problem and perhaps the hard problem. However, in the interest of making progress toward an explanation of consciousness (and not getting stuck on overlapping or conflated philosophical problems), we may avoid overly committing ourselves to a particular ontological status regarding the micro-constituents of consciousness and its metaphysical nature in our explanatory theories and models of the phenomenon.

Critically, the IA is not a pretheoretical way of approaching consciousness. Apart from doing the pretheoretical work of identifying an explanandum, it does the theoretical work of (as I have described here) attempting to conceptualize consciousness and posits that consciousness has a particular nature, that it must be explored in a particular way, and that our explanations of it must involve particular details. And thus, it is a theoretical approach to the phenomenon; one which requires sufficient justification for adopting. If it lacks the justification, or if its reasoning seems dubious, then there is (i) no good reason to accept the proposed approach, and (ii) no good reason to accept the arguments which it
poses against any other approach's success (unless, of course, other independent arguments surface). The IA’s influence over the state of the literature is mostly due to its intuitive pull. Regardless of argumentation, the facets of the IA seem to be easily understood and widely accepted by common sense intuition regarding what we understand the mystery of consciousness to be.

1.4. Conclusion

Here, I have done the work of characterizing the target of this thesis: what I am calling the inflated approach to consciousness. I have listed the three primary facets (that our theories of consciousness must answer the question about what an experience is like to experience; that consciousness is an ‘extra ingredient’ over and above cognitive phenomena; and that our explanatory theories must commit us to a particular metaphysical nature for consciousness), and briefly outlined their implications on the study of consciousness. The problem is that each of these facets unnecessarily inflate what our explanatory theories of consciousness must achieve by either leading us down a garden path, or expecting that our explanations of the phenomenon involve details that are peripheral to what potentially successful theories of consciousness need to explain. In subsequent chapters, I will outline each facet in more detail, demonstrate the defense(s) of such facets, and ultimately tackle them with the hope of deflating what is required from our explanatory theories of consciousness.
CHAPTER 2
Describing What It Is Like vs. Explaining Consciousness

Philosophy seems—at least in one respect—to be in the business of asking questions. When looking to inquire about a particular phenomenon, the questions we pose about it are not only incredibly important to the way we shape our investigation, but they determine which sorts of answers might come about through our theorizing. With this comes the responsibility of asking the relevant sorts of questions that prepare us to give a desired kind of account of the particular phenomenon in question. Additionally, our intuitions about a given phenomenon often play a part in what sort of account of it might be required to satisfy our expectations (Bengtsson 2003)—and certain questions might not allow for such a desired account to be developed through its line of inquiry. When it comes to giving an explanatory account of something, for instance, certain questions might not prepare us to do so in the required ways.

It seems to be the case that one of the most commonly accepted questions pertaining to the study of consciousness is misdirecting our inquiry when it comes to developing an explanatory theory of the phenomenon. Stoljar (2016, 1161) is correct in saying that “whenever someone wants to introduce the notion of consciousness they think requires explanation, they talk about what it is like”. The question originally posed by Nagel (1974) of ‘What is it like?’ has been taken to be one of the fundamental starting points to giving an account of consciousness, and since its inception, the literature surrounding the phenomenon has become rife with talk of what it is like (hereby WIL). However, Nagel’s question does not prepare us to explain the phenomenon of
consciousness in the way that we might expect. Rather, an answer to the question of WIL will provide us only with *description*—not *explanation*—as a description is all that is required to satisfy a WIL question.

The project within this chapter is predominantly negative: I will argue that giving an account of WIL is not the same as giving an explanatory account of consciousness, and thus, the question that Nagel poses is not a question which is apt to prepare us to explain the phenomenon in the way that is required. An important distinction is needed: While we may be able to *explain* WIL-ness, itself (the phenomenon where distinct subjective experiences of which there is something it is like to be in them occur), I argue that questions pertaining to what it is like to have such experiences are questions which demand only *description*. In other words, we may explain *that* there are particular experiences of which there is something it is like, how they are caused, and how they function (WIL-ness); while answers about *what* it is like to have given experiences only provide description and provide no explanation (WIL). From Nagel is the demand that we begin our explanatory work on consciousness through the latter—understanding what it is like to have experiences. Thus, a crucial note: When I use the term ‘WIL’ throughout, I am specifically using the ‘it’ that is embedded into the ‘WIL’ acronym as a placeholder for any given experience (what a particular experience is like to experience), and not referring to WIL-ness—the overall phenomenon of subjective conscious mental states of which there is something it is like to experience them (unless stated otherwise).

Importantly, adopting the question of WIL to have a particular experience as the starting point to explain consciousness is misdirecting our explanatory inquiry, as an answer to such a question will only require description, and nothing related to an
explanation (be it compositional, causal, functional, etc.). So, starting with WIL in this way will tempt us to look for answers which do not provide any explanatory details. None of this is due to a fault in our ability to give an explanation of consciousness, but rather, it is the fault of the question and inquiry into WIL as that which has been taken to be the fundamental starting point for most attempts at explaining consciousness.\(^6\)

Importantly, an exhaustive explanation of consciousness might still be given through other means. Thus, we need not feel hopeless about developing an explanation of consciousness when considering the failure of Nagel’s question, as it turns out to be peripheral to the enterprise of explaining consciousness. This leads to the positive note: I will very briefly (i) detail what sorts of accounts are required to develop an explanatory theory of consciousness, and (ii) allude to some questions which do prepare us to give an explanatory account of consciousness—in light of the failure of Nagel’s question to do so.

To telegraph the structure of the chapter: First, §2.1 will feature a brief reconstruction of Nagel’s (1974) definition of consciousness, and introduce how an inquiry into WIL has been widely accepted as the fundamental starting point to explain consciousness. Second, §2.2 will offer an explication of *description* and *explanation*, contrasting the two terms. Then, §2.3, the bulk of the chapter, will argue that the question which Nagel poses does not prepare us to generate answers that are relevant to developing an explanatory account of the phenomenon of consciousness. Thus,

---

\(^6\) This serves as clarification: not to confuse me for a mysterian with regard to consciousness—holding that consciousness cannot be explained or understood, and thus some important aspect of it will always be a mystery to us. (See McGinn (1989) for a paradigmatic example of mysterianism.) Instead, I argue that we can only describe what experiences are like, but this does not limit what we are capable of explaining, and required to explain, when it comes to consciousness.
understanding WIL cannot be the beginning of our inquiry into developing an explanatory theory of consciousness, as an account of WIL will collapse into mere description, and will not provide any relevant detail for our explanatory theories of consciousness. The following sections will attempt to foresee and deal with possible objections, and make concluding remarks.

2.1. The Question: What Is It Like?

In an attempt to identify what it is about the mind that makes it so intractable, Nagel (1974) identifies consciousness as the problematic phenomenon at the heart of our inquiry, and argues that the physical sciences seem unable to provide answers through their methods. While his opponents were identifying consciousness with various underlying brain states (see Smart (1959) for an example), Nagel (1974) identifies consciousness as what-it-is-like-ness (WIL-ness). Nagel (1974, 436) defines a thing (or organism) as having conscious mental states “if and only if there is something that it is like to be that organism—something it is like for the organism”. Thus, because it is both necessary and sufficient for there to be something that it is like to be X for X to be conscious, it is thought that an understanding of WIL to be X (WIL to have the experiences that X has) will entail an understanding of X’s consciousness. Nagel (1974), then, postulates that if we have the desire to understand bat consciousness, we must pose
the famous question: “What is it like to be a bat?”\textsuperscript{7} and holds that anything other than knowing WIL for a bat to be a bat is to not yet understand bat consciousness.\textsuperscript{8}

Using other terms, Nagel (1974) calls WIL-ness the \textit{subjective character of experience}. It appears, for instance, that we ourselves, as humans, are conscious—there is something it is like to be us, as we are experiencing a variety of things all the time (from the visual experiences of the words on the page before us, to the sounds of a song, the sight of vivid colours, the feeling of pains, etc.). However, we do not believe that there is anything it is like to be a rock, or that rocks have any experiences. In such a case where there is a negative answer to the question “Is there anything it is like to be X?” then X is not conscious.

Talk of WIL, and the character or quality of experience, has since become ubiquitous within the literature on consciousness.\textsuperscript{9} It is often the hope that understanding WIL is to understand what must be understood about consciousness through some sort of “analytic significance” (Lynch 2020, 256), and that the mystery of X’s consciousness can be resolved by understanding the subjective character of X’s experience—to understand what X’s experiences are like to have. Because of this, the question of WIL to X has been taken to be our fundamental \textit{explanandum} when it comes to consciousness. However, it is

\textsuperscript{7} An immediate problem with Nagel’s (1974) question is that it is ambiguous: Do we need to know WIL to \textit{be} X; or do we need to know WIL to \textit{have the sorts of experiences} that X might have? Here, I assume the latter.

\textsuperscript{8} The same goes for any form of consciousness. The bat is only used as an analogy. Additionally, Nagel (1974) has other concerns within his paper. Namely, about whether or not we can express WIL to experience X within ordinary language. However, such concerns are not relevant to the topic of this chapter, as we will see in more detail later in §3.

\textsuperscript{9} While all examples would be much too numerous to list here, some examples of influential works which have been inspired by Nagel’s (1974) paper include Jackson (1982), Levine (1983), McGinn (1989), and Chalmers (1995). In addition, talk of WIL and the character of experience runs much, much deeper than just these iconic papers.
dubious that an answer to Nagel’s question of WIL can provide explanatory detail, and, thus, whether it is the type of question which can give rise to an explanatory account of the sort that we seem to need with regard to the phenomenon of consciousness. Before arguing that an inquiry into WIL fails us in terms of preparing us to give the desired kind of account of the phenomenon, let us briefly turn (in §2) to an explication of what I take to be the difference between *description* and *explanation*.

### 2.2. Description vs. Explanation

Contemporary discourse on the nature of explanation has seen a rise since the works of Hempel (1942; 1965). However, there is less work being done on the distinction between *description* and *explanation*: what types of questions lead to either type of account (a descriptive or an explanatory account), and which of the accounts might lead to satisfactory answers to certain varieties of questions.

Here, I do not aim to get deep into the details about the particular nature of explanation, itself, in the sense that is being discussed in the contemporary philosophy of science. For the purposes of this chapter, I simply wish to contrast *description* and *explanation*, in a rough way—remaining somewhat neutral about their intricacies. However, I take there to be a certain degree of technicality in our use of the terms. Roughly speaking (and for the sake of the scope of this chapter), let us clarify the terms:

**DESCRIPTION**: To give an account of the characteristics, qualities, and/or properties of X.
EXPLANATION: To give a causal, functional, or other related kind of account of X, that involves a plurality of elements (i.e., A in relation to B), and where the explanans makes the explanandum more intelligible.

I take inspiration from Ellis (1956, 504), when he notes the difference between descriptive questions regarding “What is happening (here before my eyes)?” compared to explanatory questions regarding “What has happened (to bring this about)?”¹⁰ To answer the former, we will require a description in the sense that I have clarified above. Such a description can remain completely independent of anything else. Importantly, it can be answered in the form of “X is happening,” without relating to anything else (and the inclusion of anything else will be peripheral, or unnecessary). However, the latter question will require a type of explanation. It is not enough to propose a description of X, alone, as a way to answer the latter question. There seems to be the need for a causal story to be told, involving a plurality of elements within the explanation. Thus, while descriptions can remain independent, explanations seem to be about a relationship between two or more elements (i.e., how/why X comes from Y in a causal sense, or how/why X does Y in a functional sense).

2.3. The Failure of Nagel’s Question

¹⁰ Even though Ellis (1956) argues that explanations might be a form of description, he holds that there is a difference in types of descriptions, and thus, my characterization of description and explanation might differ from Ellis’ in a verbal and trivial way. What is important for this chapter is to distinguish between two types of accounts that we can give for a particular question or inquiry, and Ellis’ two questions here do well at pinpointing what it is that I am concerned with in this chapter.
Nagel makes two main contributions within his (1974) paper: The first, as we have seen in §1, is a way to identify and define consciousness. This notion of consciousness has become widely accepted as the primary starting point for understanding consciousness—that if we are to understand X’s consciousness, we must understand WIL to be X. The second contribution is an argument for how and why we might fail to give an account of WIL through objective and scientific means. Here, my contention is not with the second. Additionally, while WIL-ness may be a useful way to identify consciousness, my contention is with the notion that an explanation of consciousness can be given through an account of WIL. If our goal is to develop an explanatory theory of consciousness, then an inquiry into WIL will not garner any explanation. And therefore, an inquiry into WIL cannot productively serve as the beginning of our explanation of consciousness.

Previously, there have been critics of using the WIL locution as the ‘slogan’ for consciousness. Among the most notable is Hacker (2002), who argues that talk of WIL is nonsensical, and Snowdon (2010), who argues that the use of WIL as a slogan is either trivial, or false; due to the breadth of cases to which it can be applied (i.e., 'WIL to X', where X can be anything and everything), and due to the ability to talk about what X is like without consciousness being involved at all (i.e., WIL to travel to the big dipper

---

11 Notably, there are a variety of thinkers who hold that Nagel’s (1974) additional thesis (regarding the failure of the objective sciences to be able to capture the qualitative character of experience) is false (or at least dubious or unjustified). Amongst them, the arguments vary from Nagel being too hasty with his judgements about what science is capable of telling us (Akins 1993a); science actually being able to tell us a greater detail about what it is like to be a bat than what might have been expected (Akins 1993b); and the denial of the thesis that mental qualities are only knowable via private access of consciousness (Rosenthal 2010).
whilst asleep). However, such concerns are with regard to the semantics, syntax, or the use of WIL as a slogan. Not much has been said about whether or not the inquiry into WIL will aid in giving an explanation of consciousness, and whether or not WIL is even an *explanandum* to begin with.

To begin, let us take a look at WIL *questions*. It is important to briefly discuss what is meant by a WIL question, and what could be conveyed through an answer. First, the former: Even though the meaning of WIL statements and questions are often seen as vague, it is widely understood that WIL questions are not questions regarding resemblance or similarity which aim to get at whether or not X is *like* Y in a comparative sense (see Nagel 1974, note 6; Snowdon 2010; Farrell 2016). Instead, WIL questions seem to try to get at a particular *subjective character of experience*—a particular phenomenal or experiential state. And thus, there is the hope that if we want to understand a conscious experience, we can do well by understanding WIL to experience it—hence Nagel’s (1974) question about WIL to be a bat. As for what an *answer* conveys: Tye (2011) correctly seems to understand answers to questions of WIL as being statements which go like “it is like *this,*” where ’*this*’ points out the specific phenomenal state (and all of its qualities) in question. It seems, then, that WIL questions about X are *wh*-questions, associated with a *question word* like ‘what’ (Lynch 2020), and that answers to such questions, vaguely, are related to saying that X is like ’*this*’ (Tye 2011).

But what exactly does ’*this*’ denote? ’*This*’ can be given as a vague answer to almost any given question. However, what we are concerned with is what sorts of facts

---

12 An additional note about the triviality of the WIL locution: It does not aid us in understanding what consciousness is. Rather, it merely provides us with a way to describe the phenomenon in different terms.
might be provided by 'this' within answers to questions about WIL. To understand, compare these two questions:

(1) What is it like to X? (i.e., be a bat; be drunk; see red, etc.)

(2) How did the dinosaurs go extinct?

Let us analyze what sorts of answers might be expected in order to satisfy these questions. Both can be answered by saying “this” (“X is like this”; “this is how the dinosaurs went extinct”), but knowing what sort of facts are indicated by ‘this’ is what is important when it comes to finding out what sorts of answers satisfy the respective questions.

In answering (2), if ‘this’ denoted only a description of the characteristics, qualities, or properties of the asteroid that led to the events which caused the extinction, it would not be enough to satisfy our inquiry. If we replaced ‘this’ in our answer, and instead, said: “The asteroid, being [insert its characteristics, qualities, and/or properties here] is how the dinosaurs went extinct,” then we would be leaving out the very important explanatory facts that are needed to give a satisfactory answer to the question. Knowing what we know now (the general knowledge) about the dinosaur’s extinction, one might think that the above answer is acceptable. However, that is because we are believing that the relevant explanatory facts are implied. The details provided in the answer above are not enough to tell the story of how the dinosaurs went extinct to someone who does not already know the relevant causal details that one may believe are implied. The questioner must be told that the asteroid impacted into Earth, which caused
a series of events leading to the extinction of the dinosaurs. Importantly, more explanatory facts and details about the causal events are beneficial as an answer to (2); whereas additional descriptions are merely a bonus. While some descriptions might be helpful (or even required), we might do without a description of the qualities of certain items within the answer, whilst still offering an explanation that exhaustively answers the demands of the question (i.e., telling the causal story). What is critical is that an answer to (2) cannot be completely satisfactory if ‘this’ only involves an account of the characteristics, qualities, or properties of a singular item (i.e., the asteroid), giving an answer akin to a description. For a satisfactory answer to (2), what ‘this’ must denote is a series of events, told in such a way that a casual story is developed, thus being a form of explanation.

On the contrary, an answer to satisfy (1) would require the opposite of what is required to satisfy (2). While ‘this’, in an answer for (2), would involve and require explanatory facts and could omit some irrelevant descriptions, ‘this’, in an answer to (1), seems only to require description, while any and all explanatory facts seem totally peripheral to what is needed to satisfy the question. It is already understood what we are looking to get at when we ask (1)—an understanding of the qualitative character of the phenomenal experiences of X. Regardless of scope—whether or not X is a singular quale (sing. for qualia; the phenomenal and subjective character of an experience), or a collection of unified qualia—‘this’, in an answer to (1), must denote the characteristics, qualities, and/or properties of the phenomenal experience in order to give an answer which satisfies the demands of the question. If, on the other hand, ‘this’ were to denote an explanation of how the phenomenal experience came about in the first place, then it
would not answer the question, as (1) seeks a description of the *subjective character of experience* (or in other words: what that experience is like to experience). Interestingly, this only applies to questions about what experiences are like to experience (such as (1)) and not always to questions about WIL-ness—say, how and why certain brain functions allow for conscious experience. Thus, importance is placed on how our questions are posed, what they are about, and what sorts of details are required to provide a satisfactory answer.

To avoid the concern about capturing an account of WIL in language (Nagel 1974), suppose the answerer of (1) possessed special abilities, and the utterance of “this” in an answer to the question was accompanied by a magical instantiation of the subjective character of experience (which is the topic of the question) for the questioner. Even then, the questioner would not be given any new *explanatory* knowledge—they would not gain an understanding of any causal or functional (or other related) relationship that would amount to an explanation. Rather, the only things that would become apparent to them are the characteristics, qualities, and/or properties of the experience, which would amount to mere description. To put this another way: We know what it is like to be ourselves at every given moment. We know our own answer to (1), where X is ‘ourselves’ (even if we can or cannot put an answer into words). However, even with a full answer to WIL to be us, there seems to be something missing—namely, any explanation about consciousness (for instance, how it is caused and how it functions). The question about human consciousness (our consciousness) still needs to be explained, as there is a need to explain consciousness—to know how it functions and how it is caused at all. Chalmers (1995) identifies this as *the* question of consciousness. And thus,
if all the relevant knowledge embedded in a completely satisfactory answer to (1) need not possess any explanatory facts, then we cannot identify questions like (1) as the beginning, and answers to (1) as the end, of our explanatory inquiry into consciousness, as they are not at all questions which seek explanatory answers. Instead, they are questions which demand mere description.

It appears that questions about WIL prepare us to provide mere description. Despite the questions of WIL, themselves, failing to prepare us to develop explanations, how important is knowledge of WIL within our explanatory accounts of consciousness? And does an account of WIL to have a given experience require its own explanation? Put more simply: Can we explain everything about consciousness that needs to be explained without explaining or describing WIL (what particular experiences are like)? To explore this, let us springboard off of Jackson’s (1982) famous knowledge argument, as “recent interest in [Jackson’s knowledge argument] arises particularly from Thomas Nagel, ‘What is It Like to be a Bat?’” (Braddon-Mitchell & Jackson 2007, 152).13 Jackson (1982) famously illustrates a thought experiment in which there is a neuroscientist, named Mary, who has always been inside of a colourless room, and has never seen colour herself. Within the room, Mary gains knowledge of all the physical facts pertaining to colour vision. If, when released, Mary learns something new—namely, a fact about the qualitative character of experiencing colour—then physicalism is false, as Mary’s knowledge of all of the physical facts did not involve the facts about what it is like to actually experience colour.

13 Note that the thought experiment has since been abandoned by Jackson (2007).
While there has been a great deal of ink spilled on the knowledge argument, most of the discussion has been on whether or not there are indeed new facts which Mary learns upon release (for examples, see Conee 1994; Tye 2000). However, not much has been said about whether the facts or knowledge that Mary gains upon leaving the room (if there are any new facts to be learned at all) would even be relevant to Mary’s ability to explain conscious colour vision, or if they add anything to what Mary needs to explain. Imagine, for a moment, that Mary is in her colourless room and knows all the physical facts related to colour vision. While we have no current idea of what an exhaustive understanding of neuroscience would entail (Churchland 1988), we could be open to the notion (without begging any important questions) that if consciousness is a physical phenomenon, then Mary might (in principle) be able to construct a complete explanatory theory of how visual colour experiences are caused by physical systems, and how visual colour mechanisms function in the way that they do. To do both of these is to give individual explanatory theories of the WIL-ness of colour vision: a causal account, and a functional account. Other, related, explanatory accounts might also be given.

However, the next few pages will argue that even if there is genuinely something new to be learned (a new fact) upon being released from her room, there seems to be no good reason to conclude that Mary would need to incorporate an understanding of WIL to experience colour in her explanatory theory of colour vision. The gained knowledge of WIL to experience colour might very well lead to a eureka moment for Mary, making her exclaim “Oh, that’s what it’s like!” However, an addition of this sort of knowledge (regardless if Mary could capture WIL in language, or if her knowledge remains purely subjective), only provides description; capturing only the characteristics, qualities, or
properties of WIL to experience X. It provides Mary with no new explanatory details apart from those which she already knew while within her room. However, this need not worry us, as it is possible to construct an explanation of consciousness that does not involve description or explanation of WIL. In the process of arguing this, I will reply to a possible criticism of this general idea:

A critic might raise a concern regarding the general sort of explanation of consciousness that I have so far illustrated (explaining how phenomenal experience works and comes about at all), versus theories that explain how specific physical events bring about specific mental qualities. In doing the latter, it might be said that we are indeed explaining (or at least involving details about) what certain experiences are like to experience—because we must include an account of specific mental qualities. With this in mind, however, we must be careful not to presuppose a certain theoretical position with regard to the nature of mental qualities (Rosenthal 2010): Namely, that mental qualities are knowable and identifiable only through subjective, conscious states, of which there is something it is like to be in them.\(^\text{14}\) In fact, it seems possible (as I will argue) that we may identify specific mental states and explain them by using other features as our explanatory tools—all without knowing what such mental states are like to experience. Thus, the critic must not conflate two separate tasks when posing this concern about theories that target specific mental states: First is to (i) explain how specific physical states instantiate specific mental qualities (explain WIL-\textit{ness}).\(^\text{15}\) Second

---

\(^\text{14}\) This is the position of Nagel (1974) and his supporters. However, there has since been much talk about mental qualities \textit{sans} WIL-\textit{ness}. See Rosenthal (2010) for discussion.

\(^\text{15}\) For example, why does physical state P instantiate mental state M\(^\text{1}\), as opposed to M\(^\text{2}\)? Related to this is the problem of inverted qualia—potential cases where the same two physical states in different subjects instantiate different phenomenal states.
is to (ii) explain what particular mental qualities are actually like to experience, in a phenomenal sense (explain WIL).

As long as it is possible to answer (i) without providing an account off what specific mental qualities are like to experience, and show that (ii) either provides no relevant explanatory detail for our understanding of the phenomenon of consciousness, or that it does not lead to an additional explanatory inquiry, then the concern can be dispelled. Certain explanatory projects might show promise in their ability to complete these tasks. One option might be through representationalism—the view that consciousness can be explained through intentional content, and that all conscious states are of or about something. Thus, in explaining the link between specific physical states to specific mental qualities, we would be explaining the link between those physical states and their representational content—not through details about what they are like to experience. Clark (2001, Appendix II) demonstrates the motivation behind representationalism when he states that “if a difference in representational content can, indeed, always be found alongside every difference in phenomenal feel, what possible grounds could we have for insisting that there is something more to explain?”. While there may or may not be more to be accounted for, namely (ii) (WIL to have such representational mental qualities) such a task will only require additional description, after the explanatory work has already been completed. The implication is that the full explanatory work on consciousness (giving a causal and functional explanation of even the specific conscious mental qualities) may be completed without giving a description or explanation of what an experience is like to experience, and we may do so by explaining mental qualities without the mention of WIL to experience them. If the further question
(ii) remains left over, then giving an account of what such mental qualities are like to experience is only an additional task which would provide no additional explanatory detail for consciousness. It would lead by a question akin to (1) (“What is it like to X?”), and merely require a description of what such mental qualities are like to experience, as all of the surrounding work (about WIL-ness) has been completed through the inquiry into (i).

Another promising option might be to construct a specific type of quality space (QS) in order to taxonomize the mental qualities which a given mind can process. A basic example of this can be understood through our gustatory sense. Every taste has multiple dimensions: sweetness, sourness, bitterness, etc., and each taste can be mapped out onto a quality space given such parameters. While this is an example of a QS theory which attempts to provide a description of what experiences are like to experience, a different type of QS theory may successfully explain specific mental qualities in relation to perceptual functions in a way that does not mention what such mental qualities are like to experience. Thus, said QS theory would succeed at dealing with the critic.

Rosenthal (2010; 2015) takes this latter approach by describing and explaining “mental qualities not by appeal to what it’s like for us to be in conscious states that exhibit qualitative character, but instead by appeal to the role states with qualitative character play in perception”. In Rosenthal’s case, only an explanation of the functional roles of specific mental qualities in relation to perception is provided, in a way that offers an answer for what the critic is demanding, all while avoiding talk of what those mental qualities are like to experience. Importantly, Rosenthal (2015) rejects the notion that mental qualities are only identifiable subjectively, and through knowledge of what they
are like to experience. Rather, we may identify mental qualities and map them into a QS based on perceptual roles and functions, in a purely objective way that avoids talk of WIL. This has an advantage over the view that mental qualities can only be known by virtue of knowing WIL to experience them, because “we typically learn about the nature of things by appeal at least in part to causal ties they have with each other and with other phenomena. But subjective awareness of qualitative character [awareness of WIL] reveals nothing about the causal ties qualitative states have with other mental states and with behavior” (Rosenthal 2015). If our explanatory theories accept a similar concept of mental qualities to the sort described by Rosenthal (2015), they can remain robust, and aid in providing part of the explanation that is required—all while avoiding the misdirection of the inquiry into what experiences are like to experience; as an inquiry into WIL, alone, provides mere description.

The misdirection is to demand that our theories of consciousness begin with an understanding of what a particular experience (or mental quality) is like to experience. This is because: questions pertaining to WIL demand mere description; and the above approaches show that it is possible to provide explanations of consciousness for the things which need explaining (the general problem of WIL-ness, as well as the specific problem of mental qualities in relation to their causal and functional states), all without the mention (through description or explanation) of what the particular experiences of which we are explaining are like to experience. In other words, we may explain bat

---

16 There are indeed other explanatory inquiries related to consciousness (which do provide us with genuine explanation) that overflow just these two issues. However, such problems will likely be associated with what Chalmers (1995) has called the easy problems, and are arguably not related with what any particular experience may be like to experience. Thus, they would not serve as counter examples. The two problems here relate more strongly to the unique problem of experience that is associated with the famous hard problem, and WIL.
consciousness—WIL-ness and its mental qualities, in a causal, functional, or even decompositional way—without explaining WIL to be a bat; as this inquiry only demands description, and no explanation, given that our explanatory work on consciousness can be done before we engage in this separate task of giving an account of WIL. The former (giving an account of WIL-ness) would require an exhaustive explanation of the relationships between the bat’s physical system and its experiences in causal, functional, and other related ways; whereas the latter (giving an account of WIL) would only require additional description, after all of the explanatory jobs have been completed through a theory akin to the ones mentioned above. Precisely, inquiries into WIL-ness seem susceptible to explanations involving relationships between the experiences and causal or functional states; whereas inquiries into WIL to experience such experiences remain a task for description, and do not relate to anything other than experience, alone, in a way which is not already explained through the former inquiries into WIL-ness; which themselves do not require mention of WIL to experience such experiences.

So, what about Mary? We might imagine that Mary (given full knowledge of all physical facts) might experience an “A-ha” moment when she experiences color for the first time. However, regardless of whether or not this “A-ha” is due to her learning a new fact, such a new fact (if there even is one) only requires a description to be given, and will not be a necessary component of her previous explanations about conscious colour vision, as Mary may have been able to exhaust the explanatory tasks for consciousness while within her room. She may, in theory, do this and deal with the critic by constructing a QS of colour vision, and understanding the extent of mental qualities that the human visual systems can discriminate based on our fine-grained perceptual roles and
functions—all from a third person perspective, and without breaking down or mentioning what such mental qualities are like to experience. Importantly, the full extent of the explanatory work regarding consciousness and WIL-ness may be done without mention of what experiences are actually like, and thus an answer to the question of WIL to experience such mental qualities would only require details which are not relevant to our explanations of how consciousness is caused and how it works.\textsuperscript{17} The additional task of understanding WIL is not one for the explanation of consciousness, as it is possible that all of the explanatory work for consciousness could be done—involving both a general theory of how any mental state becomes a conscious mental state and how and why specific mental qualities are associated with, or come about from, specific physical states—without the need to inquire into WIL. If the question of “Well, what is X experience actually \textit{like}?" is left over after the explanatory tasks above have been completed, then such a question is demanding only a description, as we have seen in the earlier portions of §3 when we compared questions and what answers satisfy them.

Thus, such knowledge of WIL to experience colour seems peripheral to the explanatory account. If we are to give an explanatory theory of consciousness, then an account of WIL is not required at the beginning, middle, or end of our explanation. This is because we may form general theories of how experience comes about at all, and very precise theories of specific mental qualities and their relation to specific physical or functional states, all without needing to explain or describe WIL to experience such

\textsuperscript{17} Whether this type of description is enough to allow the questioner to understand WIL is another discussion. What is important is that such an answer will only require description, and not an additional explanation of any sort, as the relevant explanatory inquiries (general and specific) about WIL-ness are exhausted through other means, unrelated to inquiries into WIL. If there is a problem with our ability to give an account of WIL, then it will be a problem for our ability to describe—not our ability to explain.
mental qualities. So far as what does and does not need to be explained, the success of any particular theory (like representationalism or QS theory) does not matter. What matters is that we may give an explanation along causal and functional lines without giving an explanation of WIL. As long as something akin to this is possible, in principle, we have good reason to reject WIL as a required *explanandum* of consciousness, as all of the explanatory work relevant to the phenomenon of consciousness can be conducted without mention of WIL to have any particular experience.

Even though Nagel’s question seems to ill-prepare us, and misdirect our priorities when it comes to developing an explanatory account of consciousness, there are a variety of other questions which prepare us to provide *explanatory* answers that are relevant to understanding consciousness, rather than *descriptions* of mental qualities. Or at least, their answers will impact the way in which we might develop an explanatory theory of the phenomenon. For instance, even though the *hard problem of consciousness* is heavily influenced by Nagel’s (1974) paper, the question which underlies the hard problem of “why and how [consciousness] arises,” is, itself, an explanatory question related to WIL-ness (Chalmers 1995). In answering such a question, we must tell a casual story about consciousness, and make it more intelligible for why and how it comes about at all.

Amongst other useful issues for developing an explanatory theory of the phenomenon of consciousness, another one of importance is the overflow problem: the problem about whether or not phenomenal consciousness is richer than (has a higher capacity than) cognitive access, and thus cannot be completely identified with or explained through cognitive access. An answer to the overflow problem, much like the hard problem, will help us understand what is required for a mind to be conscious, and perhaps how
consciousness is caused and functions (see Block (2007) and Cohen & Dennett (2011) on this problem).

2.4. Further Concerns & Objections

Even though I have already dealt with one major objection, this section will attempt to briefly deal with some other possible concerns:

*WIL is still important:* One might worry that I am advocating for the elimination of the entire enterprise of understanding WIL to X. However, this is simply not the case. Here, I have illustrated the difference between questions and answers pertaining to WIL, and those pertaining to explanations of consciousness. With this in mind, there seem to be two distinct tasks: The first is to know what specific experiential states are like to experience, and the second is to explain the causal basis, function, and other related types of explanations for consciousness. They are both important inquiries, but my argument here is meant to drive a wedge between questions and answers pertaining to a *description* of WIL, and those pertaining to an *explanation* of consciousness—not to dismiss the use of, and eliminate the inquiry into WIL to experience a given experience as a whole.

*Epiphenomenal:* Another concern might be that I am merely treating experience as an after-thought to explaining consciousness—leaving it as something epiphenomenal, and assuming that it plays no causal role. However, my contention is with taking the inquiry into WIL as the starting point of explaining consciousness. I acknowledge that we may need to appeal to certain experiences in order to explain how conscious beings are able to function in virtue of said experiences. A question related to this would be an
explanatory question involving conscious states of which there is something it is like to be in—and one which I think is of incredible importance. However, such an explanation would be about the function that said states provide—not a description (or explanation) of WIL to experience said states.

*Technical term:* One objection might be that we use WIL-talk in a technical way—that it is used “in a special technical sense” (Lewis 1995, 140). From this, one might hold that we, as philosophers or scientists, use the term in a way which demands (not merely description, but) explanation. However, there does not seem to be a good reason to accept that Nagel (1974), or anyone else, is using the phrase in a way which demands a different sort of answer from the sorts of answers that are expected when the phrase is used in any non-specialized way (Farrell 2016), and thus, we have no good reason to accept this objection. Questions about WIL to X seem just as simple as any other, and demand the same sorts of responses—in specialized contexts or not.

### 2.5. Conclusion

Here, I have illustrated the effects of Nagel’s (1974) famous question (“What is it like to be a bat?”) on the literature surrounding consciousness. I have argued that inquiries of this nature are misdirecting our explanatory priorities, and I have attempted to give reason to abandon inquiries into WIL as the starting point for explaining consciousness. This is because such inquiries demand answers which serve as *descriptions*, and demand no *explanatory* facts to be provided. Despite the failure of questions pertaining to WIL, an inquiry into WIL need not be explained within our
theories of consciousness, as we can exhaust the explanatory inquiries that are relevant to consciousness without including any mention of what experiences are like to experience. Thus, questions pertaining to WIL need not be the beginning of our inquiry into explaining consciousness. This is not to say that the inquiry into WIL is wholly useless. Rather, it is to identify that there are two distinct tasks to pursue: to describe *what it is like* to experience a given experience (WIL), and to explain consciousness (WIL-ness).
Arguably, it is not yet clear from which scientific domain(s) a successful explanation of consciousness will be born. However, the second facet of the inflated approach (IA) attempts to pinpoint the area of inquiry which has the ability to explain consciousness by rejecting certain approaches and bolstering possible explanans. With this, it identifies and proposes an explanandum that is isolated from any other cognitive feature. This is a typical tactic for philosophers and scientists when first approaching a phenomenon—identify an explanandum and delegate the work of explaining it to the various domains which might have the proper tools to be able to do so.

It turns out that the hard problem and the proposed ‘extra ingredient’ (components of the second facet of the IA) dramatically shift the scope of inquiry away from any cognitive, neurobiological, computational, and functional methods of explanation. Let us call those who reject the view that consciousness can be explained via such methods anti-cognitivists, and those who purport that consciousness may be (or sometimes more strongly will be) cognitivists. However, proponents of the IA have not sufficiently justified such a move toward anti-cognitivism, as the justification for the view that there is more to consciousness than cognition is dubious. Defenders of this anti-cognitive view utilize a variety of a priori arguments against the cognitivists. It is not clear, however, that the a priori arguments which they propose can shift the scope of inquiry one way or the other as they rely on a weak form of conceivability. Thus, if all that the anti-cognitivists have at their disposal are dubious a priori arguments which cannot seem to shift the debate one way or another, and if the cognitivists have stronger arguments of
their own in favor of cognitivism, then we seem to have good reason to accept at least a weaker form of cognitivism (that an explanation might potentially come about through cognitive methods) than that of an anti-cognitivist stance.

In this chapter, I aim to demonstrate just that: The a priori arguments against cognitivism do not succeed at demonstrating that an explanation of consciousness cannot be made through cognitive methods, as they rely upon a form of conceivability (prima facie conceivability) that is much too weak to guarantee soundness from the a priori arguments themselves. Therefore, we have no good reason to accept the anti-cognitivist stance that the IA proposes. Furthermore, I will advocate against the existence of an extra ingredient (a phenomenon and potential explanans of consciousness that is over and above cognition) in favor of a cognitivist approach, and motivate such an approach by discussing some explanatory benefits.

To do all of this, the structure of the chapter is as follows: First, §3.1 will introduce and dissect the hard problem and the extra ingredient. I will explore the notion that the former can be understood as a neutral question; whereas the latter is the proposed explanans which dilutes the neutrality of the hard problem, and embeds the anti-cognitivist claim in it. Second, §3.2 will explore the a priori arguments contra cognitivism, and argue that they each beg the question due to their inability to bridge the gap between a weaker form of conceivability to the stronger form of conceivability which would be required to demonstrate that the arguments are at all sound. §3.3 will then propose various arguments against the extra ingredient, and motivate cognitivism. Amongst the arguments that I will invoke, the first is the classic argument from analogy. While arguments of this sort have been advanced against the hard problem itself, I will
argue that they do not work in this way. More importantly, however, the arguments from analogy do in fact work contra the extra ingredient—which is the driving force of the anti-cognitivist claim. The second argument is one from Kim (2001) regarding causality between physical and non-physical (or cognitive and non-cognitive) things. And finally, the most compelling argument is one which invokes Occam’s Razor against the extra ingredient.

3.1. The Hard Problem & The Extra Ingredient

The hard problem has a rich history and stems from the works of Leibniz (specifically the mill argument; 1720/1925), Nagel (1974), and Levine (1983), but has most recently been developed by Chalmers’s (1995; 1996) work. Before the coinage of the term from Chalmers, the essence of the hard problem was expressed by earlier thinkers: For instance, Leibniz (1720/1925) asks his reader to imagine walking through a brain like one might walk through a mill, and asks that we look for a perception. Nowhere in the machine, he purports, would you find such a thing—apart from the inner workings of the machine, none of which is the perception itself. And thus, perception (or perceptual experience, more related to consciousness) is separate from the physical workings of the machine (the brain and body). Similarly, Nagel (1974), as we know, argues that there cannot be a linguistic or objective description of the qualitative character of experience, and thus, our scientific methods will systematically fail at telling us about what needs to be known about consciousness. And finally, Levine (1983) identifies an explanatory gap between whatever processes might underlie conscious
experience, and the actual experiences themselves. This, he suggests, is due to a felt contingency—that it does not seem to be the case that conscious experience can be necessarily identified (through rigid designators) with its underlying processes (whether they be neurobiological brain states or functional states).

Chalmers (1995;1996) builds from these positions, and identifies a distinction between what he calls the ‘easy problems’ and the ‘hard problem’. Easy problems are not easy, per se. Rather, they are only dubbed as such because they are “directly susceptible to the standard methods of cognitive science, whereby a phenomenon is explained in terms of computational or neural mechanisms,” and thus have a clear and established line of investigation for solving them (Chalmers 1995, 202). Roughly, some of the easy problems are (Chalmers 1995, 202):

• the ability to discriminate, categorize, and react to environmental stimuli;
• the integration of information by a cognitive system;
• the reportability of mental states;
• the ability of a system to access its own internal states;
• the focus of attention;
• the deliberate control of behavior;
• the difference between wakefulness and sleep.

Importantly, even if all of the easy problems above are exhaustively solved, Chalmers argues that consciousness will still remain just as mysterious as ever. While answers to the easy problems might shed light on the phenomenon of consciousness, its nature will
remain a mystery and it remains unexplained unless we approach it directly; separately from the easy problems and the methods used to solve them.

Thus, there appears to be a distinct problem about consciousness which is separate from the easy problems. Much like Leibniz’ mill, the only features of the mind which may be solvable once inside the machine (the brain and body) are the easy problems. Keeping with the analogy: nothing about the inside of the brain and how its mechanistic parts function, Chalmers argues, can give insight into consciousness. This is because consciousness has everything to do with phenomenal experience and it has a “subjective character of experience” which may not be accessed through the easy problems and the methods used to solve them (Nagel 1974). Intuitively, it seems as if it is logically possible for all of the cognitive processing related to the easy problems to take place without any conscious experience, and thus, the problem of conscious experience is a separate issue. In contrast, then, with the easy problems, Chalmers’s hard problem is phrased like so:

It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. (Chalmers 1995, 203; my emphasis.)

---

18 See Levine (1983) for a comprehensive argument of this sort.
So, in other words, why and how does physical processing give rise to phenomenal experience? Let us analyze the hard problem further, and see how Chalmers separates its potential solution from the methods used to solve the easy problems.

If the hard problem is to be understood in the most neutral sense, then at most, it takes for granted the widely accepted notion that conscious experience seems to have some sort of physical basis: Without a brain, body, or whatever other physical basis might be required to have conscious experience, one would not have any conscious experience. However, apart from this assumption (which I will not contest), the hard problem seems like a neutral question about how and why conscious experience comes about through physical processing. In its most neutral interpretation, it is not yet disconnected from the easy problems and the methods used to solve them. Chalmers advances an additional thesis which makes this step clear—the extra ingredient. Therefore, on its own, the question within the hard problem (of how and why conscious experience arises at all) is indeed an explanandum. Yes, it makes minor assumptions, but it is fundamentally a question to be answered and to be given an explanation for. The neutrality of the explanandum gets taken away when Chalmers begins to argue against the methods of the easy problems as a way to solve it, and instead, motivates the only possible explanans via an ‘extra ingredient’.

3.2.0. Anti-Cognitive Arguments

3.2.1. The Form of The Arguments
The *a priori* arguments contra cognitivism (and against physicalism about the mind more broadly) take on a specific structure. Precisely, many of the arguments operate by first “establishing an epistemic gap … [and] proceed by inferring an ontological gap” (Chalmers 2002b). It is because of these gaps that cognitivism is purported to fail. This is primarily because if such gaps exist, then we either have no way of identifying if we are indeed studying consciousness, or instead, some other physical or cognitive phenomenon which we are mistaking it for (through an epistemic gap), or that we are in fact studying something other than consciousness entirely (through an ontological gap), and thus, making a category mistake. Stronger anti-cognitivist arguments take the latter form, and infer an ontological gap between cognition and consciousness.

Each argument takes the form of a thought experiment, and aims to yield understanding or knowledge about consciousness through conceivability, possibility, and/or necessity—all in a strictly *a priori* way. Therefore, it is incredibly important to understand how conceivability works, and what forms of conceivability are required for such arguments to be successful.\(^{19}\) It is thought that conceivability, possibility, and necessity are a good guide to *a priori* knowledge through thought experiments. However, Chalmers (2002a) illustrates that there are various types of conceivability.\(^{20}\) Perhaps most important are *prima facie conceivability* and *ideal conceivability*. First, take S to loosely

---

\(^{19}\) While there are very compelling, and earlier, works related to this (see Brook 1975), I will be operating with Chalmers’s (2002a) understanding of conceivability, as it directly relates to how the extra ingredient is being justified.

\(^{20}\) Chalmers has done extensive work on conceivability and possibility in a variety of papers (see Chalmers (2002a) for a thorough discussion, and (2009) for an applied example of the conceivability of zombies). For sake of simplicity, the two most important types of conceivability here are prima facie conceivability and ideal conceivability.
represent some statement. When it comes to prima facie conceivability, “S is prima facie conceivable for a subject when S is conceivable for that subject on first appearances” (Chalmers 2002a). With this, prima facie conceivability is not thought to be a good indicator of possibility, as one can easily be mistaken about their prima facie intuition (Chalmers 2002a). On the other hand, something becomes ideally conceivable “when S is conceivable on ideal rational reflection” (Chalmers 2002a). In part, one must know all of the relevant facts of a given matter—that which mere prima facie conceivability might not include, but that ideal reflection and conceivability would. An example of moving from prima facie conceivability to ideal conceivability would be “any mathematical statement M whose truth-value is currently unknown, but which will later be proved to be true. Here ~M is prima facie conceivable … but it is not ideally conceivable, as ideal reflection will rule out ~M a priori” once all of the relevant facts are known about the matter (Chalmers 2002a).

Considering that prima facie conceivability is not a good indicator of possibility and thus does not lend strength to the following a priori arguments, the strength of the arguments then relies upon ideal conceivability. However, as I will argue, each a priori argument contra cognitivism begs the question. This is because we lack the relevant information needed to ideally conceive of the scenarios within the thought experiments. Since this is the case, the anti-cognitivist conclusions of the thought experiments cannot be guaranteed. Similarly, as I will demonstrate, we cannot ideally conceive of the counter thought experiments that act against the original thought experiments and generate opposite conclusions, either. Thus, our current a priori arguments for or against cognitivism cannot serve as knock-down arguments against their opposition or guarantee
the truth of their own positions. At most, we can motivate anti-cognitivist or cognitivist approaches via other means.

I will now demonstrate this through an analysis of these influential *a priori* arguments contra cognitivism and argue that they are not successful. And as promised, I will demonstrate that their counter thought experiments (thought experiments which follow the same argumentation, yet lead to an opposing conclusion) are just as conceivable (in a prima facie way). However, we cannot yet determine which conclusion is ideally conceivable. Then, in §3, I will provide some independent reasons to dismiss the extra ingredient and motivate a cognitive approach.

3.2.2. *A Priori Thought Experiments*

Now, to reconstruct and analyze some well known contemporary thought experiments contra cognitivist approaches to consciousness. First, the *Conceivability Argument* (Chalmers 2009) takes $P$ to be “the conjunction of all microphysical truths about the universe, specifying the fundamental features of every fundamental microphysical entity in the language of microphysics” (Chalmers 2009). It takes $Q$ to be “an arbitrary phenomenal truth: perhaps the truth that someone is phenomenally conscious, or perhaps the truth that a certain individual … instantiates a certain phenomenal property” (Chalmers 2009). In its basic form, the argument goes as follows (Chalmers 2009):

(1) $P \& \sim Q$ is conceivable.
(2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is metaphysically possible.

(3) If $P \& \sim Q$ is metaphysically possible, materialism is false.

(4) Materialism is false.

In other words, the thought experiment urges us to imagine a philosophical zombie; a physical duplicate of oneself ($P$), say, yet it lacks phenomenal consciousness ($\sim Q$). If it is possible to conceive of such zombies, then it means that $P$ does not necessitate $Q$ in all possible worlds—$P$ and $Q$ are conceivably different things. According to the argument, if zombies are conceivable, then they are also metaphysically possible. The argument concludes that “if there is a metaphysically possible universe that is physically identical to ours but that lacks consciousness, then consciousness must be a further, nonphysical component of our universe” (Chalmers 2002b). This is because the thought experiment would succeed due to the notion that for physicalism to be true, it must be necessarily true.²¹

The purpose of this argument is not only to show that $P \& \sim Q$ is both conceivable and metaphysically possible, but it is meant to demonstrate that the nature of $Q$ is distinct from physical things. If the conceivability argument is successful, then “consciousness does not consist of anything behavioral, or cognitive, or … physical,” as zombies would be physically identical to us, behave like us, and cognitively function like we do—though they do all of this without being phenomenally conscious (Brook 2009). Thus, the conceivability argument aims to shed light on the nature of phenomenal consciousness, the physical world, cognition, and more.

²¹ However, this notion can be contested.
There is a great deal of literature revolving around the conceivability argument. While it has its supporters, many have criticized it for a variety of reasons. The majority of the critics tend to refute the second or third premise. Relatedly, there are other variations of the argument which are more complex, and involve premises which directly get at two-dimensional semantics (Chalmers 2009; Schroeter 2019); though I have omitted these additional premises from the reconstruction here, as they are not directly relevant or necessary for my argument. Specifically, I aim to argue that we do not have a good reason to accept the first premise—that $P & \sim Q$ is conceivable to begin with. What is critical to understand is that it begins with the assumption of ideal conceivability in the first premise.

For instance, the conceivability argument’s first premise is that “$P & \sim Q$ is conceivable”, and not merely in a prima facie way—otherwise the argument would not even get off the ground (Chalmers 2009). Rather, in order for the thought experiment to succeed, we must have ideal conceivability of $P & \sim Q$. If we agree that the conceivability argument is valid, then ideal conceivability would essentially illuminate the metaphysical nature of consciousness. However, what is critical is that without an understanding of the nature of consciousness, we cannot determine whether or not we can ideally conceive of $P & \sim Q$. If this is correct, then what is required for the conceivability argument to be successful is itself an understanding of the nature of consciousness as something which is either physical or not. And in this case, we must know whether or not it is in order to tackle the very first premise—otherwise, for all we know, our conceivability is only prima facie conceivability. This is because we require additional means of coming to ideal reflection, which the thought experiment itself does not possess. For the time being,
the CA will only appeal to those who already presuppose that physicalism is false and that zombies are ideally conceivable. However, there is no reason to assume that we can ideally conceive of zombies without some understanding of the relevant facts needed for ideal reflection—those being the facts about whether or not phenomenal consciousness is a part of the physical world. With this comes the unfortunate outcome of the conceivability argument being both question begging and entirely circular as a thought experiment. This is because it is a thought experiment which is designed to decipher the nature of consciousness through ideal conceivability, but in order to know if we can ideally conceive of zombies, we must first understand the nature of consciousness. Thus, as Braddon-Mitchell (2003) makes clear: “some intuitions about what is actual are driving some intuitions about what is possible. Thus it is a kind of circularity to argue from those intuitions about what is possible to claims about what is actual”.

To explain further, take $P$ for instance. Generally speaking, we can understand what this entails: it entails the full and complete microphysical reality. However, our understanding of $Q$ is not at all as robust. It is the metaphysical status of $Q$, itself, that the argument is concerned with when it comes to an account of consciousness. Thus, how can we decipher the metaphysical nature of consciousness via a priori argumentation which presupposes a certain status of phenomenal character ($Q$)? To explain, $P \& \neg Q$ is only conceivable if a certain metaphysical status regarding $Q$ is presupposed. If, for instance, we presuppose that $Q$ is a physical phenomenon, then it does not seem to be the

---

22 Arguably, the same case can be made for $P$, as we do not yet have a completed account of physics (or a full understanding of what is considered ‘physical’). Thus, there may also be a gap in our understanding of $P$. (Panpsychists might attempt to fit phenomenal properties within our conception of $P$.) However, what is important is that there is a much larger gap (even a conceptual gap) in our understanding of $Q$. 
case that $P \& \sim Q$ is conceivable at all.

Our ability to conceive of such worlds in which $P \& \sim Q$ is only achievable because of our current epistemic status regarding the nature of $Q$ in the actual world. If we decipher the nature of $Q$ in the actual world, then zombie worlds in which $P \& \sim Q$ hold might not be conceivable anymore (say, if we discover that the nature of $Q$ is purely physical). Thus, the ability to conceive of zombie worlds is only due to (i) our current ignorance with regards to the nature of $Q$ in our actual world, and (ii) our presuppositions regarding the nature of $Q$ as something which is not directly linked to $P$ through some sort of identity in the first place. Therefore, our ability to conceive of $P \& \sim Q$ does not shed meaningful light on the metaphysical status of consciousness as something that is either physical or nonphysical, as a certain metaphysical status of $Q$ needs to be presupposed in order to be able to conceive of zombie worlds at all.

Interestingly for the case of the conceivability argument, “what’s conceivable about some type of thing depends on what concept of that thing we’re operating with” (Rosenthal 2010). None of this entails that the conclusion of the conceivability argument is completely false. However, the fact that the conceivability of zombie worlds is problematic in this way gives us good reason to be skeptical about its conclusion and renders the argument dubious. Rosenthal points out that “work in so-called experimental philosophy has raised questions about the reliability of many ostensibly pretheoretic intuitions used to guide theorizing in philosophy, some relevant to issues about qualitative consciousness,” and that it “suggests in turn that the alleged intuitions are
often, or even always, induced by the adoption of a theoretical position;” in this case, dualism, before the conceivability of zombie worlds (Rosenthal 2010).

Similar to the conceivability argument, the *knowledge argument* (Jackson 1982), as we have already seen in the last chapter, involves Mary, a wonderful neuroscientist who has been trapped inside of a black and white room from the beginning of her life. While within the room, Mary learns all the physical facts about color vision. However, when she finally leaves the room, she will personally encounter the colour red, for example, for the very first time. The question is: “What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not?” (Jackson 1982). The strongest variation of the thought experiment can be captured like so (Nida-Rümelin & O Conaill 2019):

(1) Mary knows all the *physical facts* concerning human color vision before her release.

(2) But there are *some facts* about human color vision that Mary does not know before her release.

Therefore,

(3) There are *non-physical facts* concerning human color vision.

---

23 It is clear that not only dualists claim to be able to conceive of zombie worlds. Some physicalists, for instance, share this conceivability (Chalmers 2002b). Such physicalists share the same intuition regarding the conceivability of zombies with the dualists.

24 To be clear, it must again be noted that Jackson has since abandoned the knowledge argument, and no longer holds that it is successful (2007). However, it is still cited and remains an argument which dualists and anti-cognitivists use today.
Thus, much like the conceivability argument, an exhaustive account of the physical facts (involving facts about cognition and neurobiology) leaves something out. In this case, Jackson believes that Mary would learn something new once she left the room—the subjective character of experience (qualia) which the physical facts could not include. This way, Jackson is conceiving of a possibility in which Mary knows all the physical facts, but lacks particular knowledge about the phenomenal experience of color. With this, phenomenal consciousness might not be encapsulated within the realm of physical and cognitive facts. Thus, the knowledge argument aims to shed light on the nature of color vision, phenomenal consciousness, the physical, and cognition.

The knowledge argument seems to suffer the same fate as the conceivability argument. For the conceivability argument, one can simply deny premise (1), as thinking that zombies are not conceivable is just as intuitively plausible. As for Mary, Jackson (1982) conceives of a possibility in which Mary knows all the physical facts—yet those physical facts lack phenomenal facts. One can just as well deny the conjunction of premise (1-2), asserting that Mary does indeed have all the facts that are to be had. Notably, an argument of this sort has been given by P.S. Churchland (1989), as we lack the relevant information to know whether or not Mary will know everything that is to be known through the completeness of her knowledge of physical and neuroscientific facts. Thus, there is no way that we can come to ideal conceivability for the time being. Arguably, the only way to solve this would be to do more neuroscience, with the hopes of uncovering what Mary might know with her complete knowledge of it. Once we do so, we might discover that Mary in fact knew all of the relevant facts of colour vision while within her room after all. However, this could only be discovered through empirical
means—a continuation of neuroscience. Thus, the thought experiment itself cannot do much to determine if we can ideally conceive of whether or not Mary knows all of the relevant facts at the moment. Furthermore, once we do know whether or not she does (and once we have all of the relevant information to bring us to ideal reflection), then there will no longer be a need for the thought experiment. This works both ways—whether or not we will discover that neuroscience is successful in terms of incorporating all of the relevant facts about colour vision. If we discover that we have a completed account of neuroscience, while still lacking an exhaustive explanation of colour vision, then the thought experiment will also be answered in the same way: in that there is no need for the thought experiment at all in terms of offering an answer to physicalism. We would have solved the problem, either way, through empirical means, and not through ideal reflection via the thought experiment.

Furthermore, even without simply denying the first or second premise based on our inability to currently ideally conceive, it might not be the case that Mary’s knowledge of the physical world lacks any important facts about consciousness which she can then learn upon leaving the room. At most, Mary might learn an ability to discriminate the colour red from a different colour (see Lewis 1983 & 1988 for an objection of this sort against the knowledge argument). This would involve a particular type of know-how, rather than know-that, and not involve any new facts at all. Similarly, Mary might simply become acquainted with the same facts in a different way (see P.S. Churchland 1989 & Conee 1994 for examples of this objection).

One additional way to see how these arguments rely upon presupposition is through the discussion on what some are calling ‘parity arguments’ or ‘counter thought experiments’
Generally speaking, a counter thought experiment will often mimic a given thought experiment in form, but lead to an opposing, or entirely opposite conclusion. In some cases, they might not even follow the same form, but still lead to an opposing conclusion. They work by applying counter pressure on the original thought experiment by offering an opposite conclusion which is just as plausible as the conclusion of the original thought experiment. With this, parity arguments have “premises that are just as intuitively plausible, and it cannot be the case that both the traditional scenarios and the reverse-scenarios are all ideally conceivable,” as this would lead to contradiction (R. Brown 2010). J. Brown (2003) does much to discuss whether or not certain counter thought experiments are indeed successful at countering. He argues that we might employ a ‘Ratio Test’ to see which thought experiment is more probable—the original, or the counter. In this, we would “assign a probability to the phenomenon of a thought experiment, given the thought experiment set up”, and the original thought experiments which are most close to a 1:1 ratio of intuitive probability with their counter thought experiments are seen as the strongest countered cases (J. Brown 2003).

R. Brown (2010) has done something similar, forming a taxonomy of counter thought experiments against both zombies and Mary to show that neither the original thought experiments or their counters can succeed a priori. First, an example of a counter thought experiment contra the conceivability argument—Zombies. Similar to the original, take \( NP \) to be the “totality of the nonphysical facts about me now” and \( Q \) to be “some qualitative fact about me, say that I am now seeing green” (R. Brown 2010, 50):

(1) \( NP \) and \( \sim Q \) is conceivable
(2) If (NP & ~Q) is conceivable, then (NP & ~Q) is possible
(3) If (NP & ~Q) is possible then dualism is false
(4) Therefore dualism is false

He explains that this goes *both ways*. Both the original thought experiment, and the counter thought experiment do not succeed. Although, only one of them can actually be *ideally* conceivable; the other would have only been prima facie conceivable. A similar counter thought experiment can be pressed against the knowledge argument—Mark:

Mark is the … Mary-like super-scientist who learns the completed nonphysical science without seeing red. When Mark is let out of his special black and white room and sees his first red ripe tomato there is no reason to think that he won’t learn what it is like to see red in exactly the same way that Mary did. This argument exactly parallels Jackson’s (1986) original formulation of the knowledge argument. Mark knows all of the nonphysical facts but yet learns something new when he sees red for the first time, therefore phenomenal facts cannot be deduced from the nonphysical facts. Whatever response the dualist gives to Mark can be given to Mary. (R. Brown 2010, 53)

Thus, the only way to determine which one is truly ideally conceivable would be through *a posteriori* means, as the *a priori* means are not sufficient due to both the original thought experiments and their counters being equally as plausible. Although, once we
have done so, there is no longer a need for the *a priori* method in the first place, as the problem within the thought experiment would have already been solved through *a posteriori* means (R. Brown 2010). However, there is an important addition to mention:

This can be also said for those who do not employ counter thought experiments, but deny the conceivability of a given thought experiment. However, they too, must be able to demonstrate that it is not conceivable on ideal reflection—which, too, requires an understanding about the nature of consciousness. Simply asserting that it is not conceivable is question begging in the same way as asserting that it *is* ideally conceivable.

If this is the case, then we seem to be in a standoff when it comes to these *a priori* arguments against physicalist and cognitivist approaches to consciousness. *A priori* arguments of this sort cannot seem to demonstrate whether or not cognitivism or anti-cognitivism is true. If not, then the cognitivist and the anti-cognitivist, respectively, require different arguments to motivate their approaches. For the anti-cognitivists, they must form an argument (that is not one of the above *a priori* arguments) to demonstrate that consciousness cannot be explained through physical phenomena or cognition and that there remains an extra ingredient over and above them which must be used as an explanans. However, anti-cognitivists do not possess substantive arguments beyond the *a priori* arguments of the sort discussed here. As for the cognitivists on the other hand, I believe that they possess good reason to tilt the scales in their direction and deny a potential extra ingredient. This will allow us to see that cognitive approaches are not only attractive based on their potential explanatory promise, but that they are the default
position that one ought to take when approaching the phenomenon of consciousness. I will now turn to arguments against the extra ingredient and in favor of cognitivism.

3.3. Denying Extra Ingredients & Motivating Cognitive Approaches

Here, I will consider some arguments for why we ought to be skeptical about an extra ingredient, and then I will motivate cognitive approaches as an alternative. First, it must be clear that none of the above mentioned a priori arguments have been successful at countering cognitivism. One might claim that even if they did, then they only serve as negative arguments contra cognitivism, and not positive arguments for the existence of an extra ingredient. However, if consciousness remains a genuine phenomenon to explain, then the existence of an extra ingredient is entailed by the arguments, as something is still required to explain the phenomenon and an extra ingredient would be required to act as an explanans. The only way to avoid a potential extra ingredient in the face of cognitivism’s falsity is to eliminate consciousness as an explanandum, outright. But, if cognitivism fails in light of the success of the above thought experiments, so do the surrounding means of explanation (including neurobiological features and computation, as well), and a currently unknown and ontologically distinct extra ingredient would remain in their wake. However, I will now give three arguments that serve to counter the extra ingredient and motivate cognitivism:

First, an argument from Occam’s Razor. Chiefly, Occam’s Razor suggests that any explanation which posits the least number of entities is usually the best explanation. This is typically because of the concern that we ought to only believe in (and include
within our explanations) entities that we have good reason to believe in, and thus, it is within our best interest to refrain from expanding our ontology to include entities or features which we lack sufficient justification to believe exist. Commonly, neither cognitivists nor anti-cognitivists would disagree that consciousness seems to come from a physical and representational basis. Additionally, cognitivists and anti-cognitivists alike agree that the phenomena related to cognition and the easy problems may potentially shed light on the hard problem and consciousness itself (though the anti-cognitivists argue that cognition and the easy problems will not exhaustively explain what needs to be explained about consciousness; see Chalmers 1995). Therefore, if both the cognitivist and anti-cognitivist agree on these terms, it is up to the anti-cognitivist to provide sufficient reason to accept the more radical position of adding an ontological and explanatory component to our approach, and the attempts through a priori arguments have so far been unsuccessful. In other words, the cognitivists currently hold the default position, as most anti-cognitivists agree with what the cognitivists ontologically posit. For now, it is wise to suspect that our cognitive approaches will continue to provide explanatory results for the mind until proven otherwise. It is thus up to the anti-cognitivists to provide sufficient justification for positing the existence of an extra ingredient for our explanations and ontology. Without proper justification, there is no reason to accept an extra ingredient over and above cognition. If we were to accept it without good reason, then there is no reason not to similarly accept an infinite amount of extra ingredients.

Second, there have been a variety of arguments from analogy which have been posed against the hard problem and against consciousness as a whole. P.S. Churchland writes:
The ‘spirits’ and ‘principles’ of alchemy, the ‘crystal spheres’ of pre-Galilean astronomy, ‘daemonic possession’ of Medieval medicine, ‘phlogiston’, ‘ether’, and ‘signatures’, are now nought but dry bones of an earlier intellectual ecology … a similar fate may befall concepts respected and revered in our own prevailing conception of how humans works, and the concept on which I mean to focus is consciousness. (P.S. Churchland 1983, 80)

And thus, consciousness as a whole is argued to be a concept which is ripe for elimination, in the same way that our earlier mentioned phenomena have been. However, Chalmers, I believe, is right to defend consciousness and the hard problem against such arguments (at least to a degree). He writes:

Experience is the most central and manifest aspect of our mental lives, and indeed is perhaps the key explanandum in the science of the mind. Because of this status as an explanandum, experience cannot be discarded like the vital spirit when a new theory comes along. Rather, it is the central fact that any theory of consciousness must explain. A theory that denies the phenomenon “solves” the problem by ducking the question. (Chalmers 1995, 206)
This is because, he argues, consciousness itself is an explanandum; it is a phenomenon that needs to be explained. Vital forces, on the other hand, were posited as a means to explain the phenomenon of life, and thus, it was used as an explanans. This way, vital forces were indeed eliminated as an explanans once a more successful explanatory theory came along. In this sense, I believe Chalmers is right to defend the hard problem against such arguments; though only if it is taken in its most neutral sense—as a question.

If the hard problem is taken in a purely neutral way, as a question concerning how and why consciousness realizes itself, then the hard problem remains a proper explanandum. However, the arguments from analogy maintain their strength against the extra ingredient (albeit admittedly not enough strength, themselves, to knock down anything entirely), as it is a proposed explanans for the hard problem. Furthermore, it is a proposed explanans which carries significant ontological commitments and poses a radical shift in our inquiry. And thus, much like the existence of vital forces, to posit an extra ingredient of the sort that Chalmers and anti-cognitivists wish to posit requires substantial argumentation and justification—which they have so far failed to provide, given that the \textit{a priori} arguments are dubious.

Third, a fairly simple argument regarding causality can be pressed against the anti-cognitivist. Since the anti-cognitivist typically agrees with the cognitivist that the mind has a physical and representational basis, and only disagrees about whether the mind is wholly physical, representational, and cognitive, then the anti-cognitivist must hold that the non-cognitive and/or non-physical mind can somehow causally interact with its physical basis. However, as Kim (2003) makes clear, causality makes sense within its own domain—not across multiple domains. We can make sense of causation, at least in
principle, within the physical world due to the spatio-temporality of physical entities. However, if consciousness is, itself, non-physical and non-cognitive, then there is no way to make sense of causation between consciousness and its physical basis. This is because “our idea of causation requires that the causally connected items be situated in a spacelike frame-work” (Kim 2003, 73). Without such a framework, it is not clear how a particular non-cognitive mental state would interact with a particular physical system, as opposed to any other physical system.

Apart from these three arguments, an additional problem also arises: If we truly require an extra ingredient that is distinct from the physical world, cognition, and/or function, then it seems as if our approach toward consciousness must be direct—meaning that we cannot explain consciousness through, or in virtue of, anything else like function or cognition. If this is the case, then it seems much more likely that an answer to the hard problem will only be circular, as consciousness cannot be explained by anything else (in an indirect way) other than by reference to itself and its distinct extra ingredient. Cohen and Dennett (2011) argue that if consciousness can be separated from function (and cognition) in this way, then the hard problem is rendered an impossible problem. This is because there would be no way for us to identify, through third person methods, whether a subject is conscious or not. And further, there might be no way for the subject, themselves, to know that they are conscious, as they might not even be cognitively aware of their own conscious experience.

Cognitive methods, however, have so far provided us with a lot of detail regarding function and representational mental states. For those that believe in representational states of any kind, it is not at all controversial to suggest that conscious
mental states are representational (though it is controversial to state that they are exhaustively representational), and thus, cognitive science might very well illuminate something significant about consciousness. So far, at least, we have no reason to think otherwise. Clark (2001) notes that, regardless of our ability to conceive of possible zombie worlds, “if, in the actual world, the links between the physical and the phenomenological facts can be this watertight, it is unclear why a full appreciation of the nature and origin of those links would not amount to a full understanding (for our actual-world purposes) of phenomenological consciousness itself”.

Importantly, cognitive approaches to consciousness are attractive in the sense that they typically approach the phenomenon via representation and function; two aspects which are uncontroversially related to consciousness. Additionally, cognitive approaches have the promise “of offering a reductive or non-circular explanation of phenomenal consciousness by explaining phenomenal consciousness in terms of cognitive states that are not themselves intrinsically phenomenally conscious” (Mandik 2018).

Unfortunately, as we have seen in the first chapter, anti-cognitivists typically hold that those who take a cognitive approach are misunderstanding the explanandum, and not explaining consciousness. More strongly, Chalmers (1996) has criticized cognitivists of “failing to take consciousness seriously” for this reason. However, it is not clear that cognitivists are misidentifying or misunderstanding the explanandum. Rather, they are simply using cognitive functions, representations, and so on, as explanans for phenomenal consciousness. The neutral version of the hard problem may very well be

---

25 Recall how Chalmers (1995) tells a story of someone who purports to explain consciousness, only to talk about cognitive functions in great detail. He argues that they have missed the point entirely.
solved on functional and cognitive grounds, and we have no reason yet to suspect that it will not. *A priori* arguments cannot seem to guarantee that it will or will not—so we can only sit back and find promise in particular approaches while we wait for the empirical work to be done. Unless there is good justification for why we ought to refocus our efforts through other explanatory means, the conservative and promising indirect approach through cognitive science and related fields seems to be the default position, and the anti-cognitivists have no good reason to reject it.

### 3.4. Conclusion

In this chapter, I have explicated the hard problem of consciousness in further detail; making it clear that in its most neutral sense, the hard problem is indeed an uncontroversial explanandum. However, the neutrality of the hard problem becomes tainted when anti-cognitivists dubiously posit the need for an extra ingredient as a potential explanans over and above cognition. Primarily through a mixture of intuition and question begging thought experiments, anti-cognitivists attack cognitivism through *a priori* argumentation. It is not clear that we can be sure that their arguments can succeed, given our current epistemic status regarding consciousness. Only through *a posteriori* methods can we shed light on the success or failure of said arguments. However, the anti-cognitivists lack a substantial method of inquiry beyond *a priori* conceptual analysis, and thus, we are left with cognitivism (and other related methods) as a reasonable (and so far undefeated) approach. Importantly, cognitivism has significant explanatory benefits and
appears to be an attractive approach which is grounded on uncontroversial foundations—accepting that consciousness at least stems from a physical and representational basis.
CHAPTER 4
Metaphysical Theories & Agnostic Explanations of Consciousness

Perhaps one of the core issues about our attempts to explain consciousness is not so much about the difficulties that the phenomenon itself might pose, but is more about how we aim to explain it. In earlier chapters, I have separated unproductive questions from productive ones (chapter two) and located a potential area from which our explanations can be fruitful (chapter three). However, one incredibly important problem remains: It appears that our theories of consciousness conflate the task of explaining consciousness with the task of solving other, persisting philosophical problems—namely, the mind-body problem. In order to begin explaining the phenomenon as productively as possible, we must work to figure out which explanatory tasks are important. By eliminating specific details, or learning that our explanatory theories can remain neutral about certain components, we can cease being stuck on persisting peripheral problems.

Important to philosophy is the mind-body problem: the problem of how the (subjective and apparently non-physical) mind relates to the (objective and physical) body or brain. To solve such a problem, we must give an account of the nature of mind and explain how it relates to other related phenomena (such as the body or brain). It turns out that a particular variety of theories about consciousness (sometimes called metaphysical theories of consciousness, which aim to target the hard problem,) either fully target the mind-body problem itself, or overemphasize and become stuck on metaphysical commitments to the mind-body problem. I believe this to be because problems and answers related to consciousness conflate problems about consciousness with the mind-body problem and lead us to expect our theories of consciousness to
involve an answer about the metaphysical nature of consciousness. Here, I will explore how the *semantic view of theories* might show promise at being able to productively allow us to provide answers to problems related specifically to consciousness (including the hard problem) without overly committing us to any answers about the mind-body problem, the micro-constituents of consciousness, or the metaphysical nature of consciousness. It does this by showing that we can formulate theories which remain agnostic about the existence of the properties which they seem to refer to (the micro-constituents of consciousness, or the properties which make up its nature), by using undefined variables within our theories, somewhat akin to mathematical formulae.

In this chapter, I will focus on the nature of explanation in theories of consciousness and will make a distinction between theories which target the mind-body problem and theories which target problems specifically related to consciousness. It turns out that many of the currently proposed answers to the hard problem situate themselves awkwardly between the two types of problems, as most of the answers tend to veer toward solving the mind-body problem. Importantly, such proposed answers are not as informative as suspected when it comes to problems specifically related to consciousness itself. This is because regardless of the metaphysics and constituents of mind, or with what we identify the mind as/with, higher-level explananda remain as independent explanatory inquiries which specifically relate to consciousness and are separate from the mind-body problem. Thus, the goal is to identify how we can make progress on the problems of consciousness (including the hard problem) in a way which avoids commitments to the mind-body problem.
To begin, §4.1 will first characterize and examine the types of theories that I take to be unfruitful; namely, what Chalmers (2002b) calls *metaphysical theories of consciousness*. Then, I will discuss exactly what it is that these types of theories aim to do and how they are problematic. Namely, they target the hard problem, but do so in a way which leaves problems specifically related to consciousness left over, and they divert our explanatory progress toward consciousness by entering into persisting debates about peripheral philosophical problems that are not required to be solved in order to explain consciousness. In §4.2, I will then motivate the notion that our explanations of consciousness, specifically, may be at a higher level of explanation than answers to the mind-body problem and that some explanations of consciousness may be ontologically neutral and uncommitted with regard to the micro-constituents of consciousness. This is because there are higher level questions about consciousness which remain left over even after the proposed metaphysical theories are deemed true, and which demand higher level explanations with their own relevant details (other than the details regarding the nature of consciousness and its constituents). Then, §4.3 will conclude.

4.1.0. *Metaphysical Theories of Consciousness*

Chalmers (2002b) does the work of taxonomizing a variety of theories of consciousness into ‘types’. While I do not believe that his taxonomy is by any means perfect (as, for instance, the segmented ‘types’ are often much too broad), Chalmers’s taxonomy does a good job at capturing the broad range of what he calls “metaphysical theories” which aim to respond to the hard problem and offer explanations of
phenomenal consciousness. Very briefly, Chalmers’s (2002b) ‘types’ can be broken down like so:

From the materialist/physicalist camp comes Type-A, B, and C. Type-A denies that there is both an ontological gap and an epistemic gap between the physical and the phenomenal. Proponents of Type-A hold that once we know certain things about the physical world, we know everything there is to know about the phenomenal. Some notable Type-A theories are eliminativism, illusionism, and analytic functionalism. Type-B, on the other hand, denies an ontological gap (much like the Type-A), holding that consciousness is wholly physical, but accepts that there is an epistemic gap. The most notable Type-B theories are the identity theories, which identify consciousness with particular brain states or functional states. The last of the physicalist ‘types’ is Type-C, which holds that there is likely no ontological gap, but that there is indeed a deep epistemic gap between the physical and the phenomenal. However, Chalmers (2002) notes that if physicalism is true, then Type-C is destined to collapse into either Type-A or Type-B. Often, it is only a temporary position, unless the particular proponent of Type-C is a mysterian.

On the other hand, non-physical or non-reductive, and other ‘types’ can be found through Type-D, E, and F. First, Type-D can be understood as interactionism—the theory that consciousness is ontologically distinct from the physical world, yet the physical world and consciousness can causally interact with one another from both the physical to the mental, and from the mental to the physical. Type-E, known as epiphenomenalism, is similar, yet holds that the causal interaction only goes one way—from the physical to the mental. In other words, physical events cause consciousness, but consciousness cannot
causally impact the physical world and physical behaviour. Last, Type-F encompasses *panpsychism* and *panprotopsychism*, which are essentially theories which identify consciousness with fundamental properties at the micro-physical level of reality, and hold that our conscious selves (as humans) are macro-conscious entities constituted by a variety of micro-conscious or proto-conscious entities.

Primarily, each theory attempts to identify what consciousness *is* and how it relates to the body, world, and/or mind (more broadly). Type-A typically identifies the phenomenon as behavioral states, functional states, representational states, illusory states, dispositional states, or theoretical errors; Type-B (while it can overlap with some of Type-A’s) typically identifies it with neurobiological states; Type-C remains somewhat agnostic with regard to what it can be identified with, but holds out for an answer in line with a physicalist ontology; Type-D and Type-E identifies it with non-physical and non-cognitive entities or properties; and Type-F identifies it as a fundamental microphysical property or as a collection of protophenomenal properties.

Before we understand what is problematic about these ‘types’ of theories (metaphysical theories), we must understand what they have in common. This also helps to simplify them even further for the purposes of this chapter. We must understand that “a solution to the hard problem would involve *an account of the relation between physical processes and consciousness*, explaining on the basis of natural principles *how and why it is that physical processes are associated with states of experience*” (Chalmers 2002b; my emphasis). It is precisely this that the following theories either try to *solve* or *dissolve*. The former tries to provide an answer to the hard problem and the latter denies the sort of
distinction or relationship between physical processing and consciousness that Chalmers motivates.

However, when the hard problem is phrased in this particular way and the potential solutions which Chalmers (2002) analyzes look like the various ‘types’ of theories above, it is either (i) not clear that the metaphysical theories being discussed are theories which do much more than identify what consciousness is (a physical property, non-physical property, etc.) or where consciousness is located (in the brain, at the micro-physical level, etc.), or simply (ii) make too many (perhaps unnecessary) commitments to the nature of consciousness and answers to problems which are peripheral to explaining consciousness, such as the mind-body problem. Thus, at the very least, the metaphysical theories in question seem to target the mind-body problem by identifying what consciousness is and how it relates to surrounding phenomena (such as the physical world, the brain, micro-physical entities, and so on). And further, most of the debate surrounding such theories (as we have seen in earlier chapters) has to do with whether or not they are correct in their judgments that consciousness can be identified as/with whatever particular thing they posit. So, to sum up, the proposed metaphysical theories for consciousness seem to target the nature of consciousness and how it relates to other surrounding phenomena (whatever phenomena a given metaphysical theory deems as being related to consciousness), and thus primarily target the mind-body problem. I will now go into further detail as to why this is problematic, and eventually, to how we can develop better explanatory theories of consciousness which avoid these problems.

4.1.1. The Problem with Metaphysical Theories of Consciousness
Chiefly, Chalmers (2002b) understands each of the ‘types’ of metaphysical theories to be positions regarding consciousness and the hard problem. Some of them purport to solve the hard problem, while others, instead, essentially aim to dissolve it. An example of the former would be through panprotopsychism—a variation of Type-F, wherein the constituents of consciousness are protophenomenal properties located in fundamental microphysical properties, and when they come together in specific ways, generate conscious experience. On the contrary, an example of the latter would be through illusionism, or on the opposite side of the spectrum, panpsychism itself. This is because illusionism aims to explain how phenomenal properties (at least how Chalmers conceptualizes them) do not arise from physical processing in a way which would require a hard problem at all. Instead, we are left with some sort of meta-problem of consciousness, where we must instead explain how we believe we are experiencing phenomenal mental states while we might not be (Chalmers 2018). Panpsychism, on the other hand, also dissolves the hard problem of consciousness by situating consciousness, itself, within microphysical properties. Thus, there is no sort of emergence that would require a hard problem. And instead, different problems seem to arise for the panpsychist as a replacement for the hard problem: Namely, the combination problem, where panpsychists (and potentially others; see Mendelovici 2019 for an argument for why the combination problem extends to all positions and not just panpsychists,) must explain how micro-conscious things combine or unite into the sort of macro-conscious things we call our own consciousness. And thus, the hard problem is dissolved, and restated in
another sense by shifting the focus from physical processing and consciousness to micro and macro consciousness.

It is this that is precisely the crux of the issue. Even if the above metaphysical theories are proven to be true, and consciousness is identified with brain states, functional states, micro-physical states, non-physical properties, or so on, then further questions about consciousness remain left over. Simply by identifying what consciousness is (as the above metaphysical theories attempt to do—by identifying it with specific substructures, properties, or entities), there will be a great deal about the function and/or causation of consciousness that is not explained. While there are some metaphysical theories that Chalmers (2002b) lists which might indeed solve the hard problem if they are deemed true, like panprotopsychism, our progress on consciousness, the hard problem, and further problems related to consciousness are often stunted due to disagreements about the nature of consciousness which lead to a standoff (as mentioned in the previous chapter). However, it is not clear that our explanations of consciousness need to engage with the nature of consciousness, as such problems are problems that are left to the mind-body problem—a problem which metaphysical theories cling too strongly to.

Thus, metaphysical theories bring about two problems: First, they place emphasis on the micro-constituents of consciousness and the nature of consciousness, thus targeting the mind-body problem, and leaving out a variety of higher-order problems specifically related to consciousness. This is because while the metaphysical theories may provide answers to the mind-body problem, they leave out a variety of the important causal and functional relationships that need explaining when it comes to the phenomenon of conscious experience. Second, in doing this, they divert our progress on
consciousness by unnecessarily confronting the mind-body problem and making ontological commitments to the metaphysical nature of consciousness. This leads to metaphysical standoffs in discussion which stunt the success of such theories due to the controversy of their posited nature of consciousness and their proposed micro-constituents (see Balog (2008) for discussion).

The goal in this chapter, then, is to provide a way to make further progress on consciousness and the problems which are directly related to it, without getting stuck on the same issues that the metaphysical theories do. A different view of theories—the semantic view of theories—can provide us with a way to develop explanatory theories of consciousness (and the hard problem) without committing us to a particular metaphysical nature of consciousness, any particular micro-constituents, or any answers to the mind-body problem, while all remaining just as robust in terms of what explanatory details are relevant. I will now begin to motivate this approach.

4.2.0. Agnostic Explanations of Consciousness

So far, we have taken a brief look at Chalmers’s (2002b) taxonomy of metaphysical theories of consciousness. I have described them in brief detail, outlined what they have in common, and outlined the problems that they leave us with. However, when it comes to developing explanatory theories which are as simple and yet as maximally informative about consciousness as they can be, it is important to understand at what level of abstraction our explanations can be located, and what sorts of details are relevant to the particular questions or problems we are faced with. In this section, I will
motivate the notion that our explanations of consciousness need not ontologically commit themselves to any specific lower-order properties. Rather, they can be best understood as more neutral models of consciousness which remain agnostic about the metaphysical nature of consciousness by utilizing undefined variables as placeholders for whatever potential constituents might be proven to make up consciousness. This way, we can make progress on problems related to consciousness by understanding that a variety of details (particularly ones related to answers to the mind-body problem) are not entirely relevant to explanations specific to consciousness.

4.2.1. Levels of Explanation and the Semantic View of Theories

As we have seen in the second chapter, particular questions seem to require particular answers which include details relevant to the initially proposed question. Additionally, for ordinary and pragmatic reasons, context seems to be an important feature for what sorts of answers are relevant to the proposed questions. In this section, let us compare a variety of explanations in order to see that different explanations concern themselves with varying degrees of detail, and that they are expected to reflect the level of abstraction at which their initial questions are concerned. The details that are relevant within a given explanation are based on both the demands of the question itself, and on the context of the question. To begin, Klein (2014) lists groups of explanations, with the goal of analyzing which ones are more successful within certain contexts:
1. (a) The square peg failed to pass through the hole because its cross section was longer than the diameter of the hole.

   (b) The peg failed to pass through the hole because [extremely long description of atomic movements].

2. (a) Klein got a ticket because he was driving over 60 mph.

   (b) Klein got a ticket because he was driving exactly 73 mph.

3. (a) Socrates died because he drank hemlock.

   (b) Socrates died because he guzzled hemlock.

4. (a) Esther ran because she was scared of the bee.

   (b) Esther ran because [complicated neural description].

Klein (2014, 209) points out that, for many, the first explanation within each group seems like a better explanation than the second—“even though the truth of the second sentence guarantees the truth of the first”. And as Klein makes clear about explanation (1a) through Hilary Putnam:

In this explanation certain relevant structural features of the situation are brought out. The geometrical features are brought out. It is relevant that a square one inch high is bigger than a circle one inch around. And the relationship between the size and shape of the peg and the size and the
shape of the holes is relevant. It is relevant that both the board and the peg are rigid under transportation. And nothing else is relevant. The same explanation will go in any world (whatever the microstructure) in which these higher-level structural features are present. In that sense this explanation is autonomous. (Putnam 1975, 296)

Importantly, Putnam understands (1a) to be a superior explanation because (i) it includes the minimal amount of information that is sufficient for an explanation given the context of the explanandum, and (ii) the relevant details within the explanation have to do with a specific level of explanation which is relevant to the level at which the question is concerned (in this case through a higher-order property: namely, the shape of the peg) (Klein 2014, 210). So, through Klein, we can understand that there are explanations which locate themselves at different levels of abstraction. Higher level explanations deal with much more abstract concepts, properties, relations, functions, and so on; whereas lower level explanations have to do with lower-order concepts, properties, relations, functions, and so on. In the case of (1a), it concerns itself with the relevant higher-order properties of being square and being circular, and not so much with the lower-order properties concerning specific atomic movements, and is thus a higher level explanation.

About (i)—the inclusion of the minimal amount of information that is sufficient for an explanation given the context of the explanandum—Klein (2014, 212) appeals to the Gricean maxim of Relevance, which “tells me to give only such information as is relevant to my hearer’s interests”. However, such a maxim is important not only for the specific interests of our assumed “hearer”. Rather, we can understand the explanandum,
or question itself, as demanding certain relevant bits of information involved within a proposed explanation. When we ask about how boats manage to stay afloat, for instance, we need not include irrelevant information about whether the boat must be made up of wood or metal. All that is relevant is to include the details related to buoyancy and perhaps the particular shape that a boat must have. Related to this, Klein writes:

My wife wants to know why I got another ticket; the fact that I broke the speed limit is sufficient to satisfy her interests, and the specific speed is (we assume) irrelevant to her interests in the conversation. Similarly, the Gricean maxims of Quantity and Quality should forbid me from giving (4b) as an explanation when the equally effective and much shorter description (4a) is available. (Klein 2014, 212)

However, there are additional reasons to take the Gricean maxims of Quantity and Quality seriously:

Indeed, to give (4b) would (on the assumption that I’m being cooperative) produce several false implicatures: that the extra detail is relevant, in the sense that counterfactuals involving small changes to Esther’s neural state would result in her calm or that I have good evidence in a particular case for the complicated neural process at which I have hinted. Neither of these is likely to be true. So to give (4b) would be misleading, in the sense that I would implicate something false to my listener. (Klein 2014, 212)
One might think that in applying this to consciousness, I am implying that consciousness, itself, is a higher-order and non-reducible property; or that our explanations about consciousness must involve such properties. However, this need not be the case, despite our potential explanations appearing to use higher-order terms. Thus, about (ii)—that the relevant details within the explanation have to do with a specific level of explanation which is relevant to the level at which the question is concerned—Klein (2014) believes that we can stay ontologically uncommitted and agnostic about the actual metaphysical existence of the proposed higher-order properties that we seem to refer to within our explanations. Sometimes, it is not actually relevant whether or not the higher-order concepts that we use within explanations are actual abstract objects and non-reductive properties within our world. Rather, they can be placeholders within theoretical models, and Klein (2014) warns us to not be a literalist about such predicates; that it is a poor practice from philosophers of mind to assume that when an explanation involves a term that appears as if it refers to a higher-order property, that the term thus attempts to refer to it as such, and that the explanation commits itself to the existence of such a property somewhere in the world. Hence, Klein’s term—literalist: one who takes the terms within our explanations as literally referring to real and specific properties or entities within our world.

It may not be the case that our terms within explanations actually posit the existence of specific properties or entities, even if they appear to do so. Explanations may be agnostic and ontologically uncommitted to any actual or specific higher-order properties to which they seem to refer. Instead, such terms can be understood as
undefined variables within specific contexts of the explanandum or question for sake of the Gricean maxims of Relevance, Quantity, and Quality. We can then do the additional work of mapping out our ontology, from the explanatory model to the actual world, after the more agnostic explanation has been given. As a reason for why this is the case, Klein (2014) tells a story about the Hodgkin-Huxley model of action potentials. Klein characterizes the Hodgkin-Huxley explanatory model like so:

Below the threshold membrane potential, $\frac{G_{Na}}{G_k} = 1$, and so small depolarizations result in offsetting Na and K currents. Above the threshold, $\frac{G_{Na}}{G_k} > 1$, which results in a net Na current with positive feedback. (Klein 2014, 215)

Klein (2014, 215) views this explanation as something akin to our previous (1a), as “it is arguably a better explanation than one that goes into the details of the opening of sodium channels and for the same reason: it gives us precisely the information needed to explain the threshold and no more”. Further, even if it included additional causally relevant details, such details are not relevant to the demands of the explanandum that Hodgkin and Huxley were trying to explain.

Now, as Klein suggests, the literalist about terms in our explanations may assume that $\frac{G_{Na}}{G_k}$ refers to a distinct higher-order property when it is used within the explanation. However, this would be a mistake for a few reasons. Most relevant is that “the mathematical form of the explanans is important: the mathematical properties of ratios can be used, along with the mathematical properties of facts about $G_{Na}$ and $G_k$, to
explain further facts about the shape of the action potential. Treating $G_{Na}/G_{k}$ as a single property … obscures this explanatorily useful relationship” (Klein 2014, 216).

Thus, instead, we might better understand explanations through a semantic view of theories. The semantic view of theories sheds our emphasis on the particular language within explanations, and instead emphasizes the models which they create. This tracks well to how many theories operate, because “many disciplines present models of some target phenomenon and then reason about them. This is most obvious in a field like cognitive psychology … one is presented with a model mechanism and an assertion that this is what the brain does—that is, an assertion that the brain is isomorphic to the model in some respect” (Klein 2014, 219). And additionally, our explanatory models of this sort are, themselves, multiply realizable—which is not to say that this entails that there are actual properties within the models that are also multiply realizable, of course, as the semantic view emphasizes the models themselves and not the language which may or may not postulate the existence of particular properties. Rather, they are multiply realizable in the sense that a multiplicity of angles can be taken in developing the same sorts of models of the target phenomenon, using different terms that belong to various disciplines. For example, Klein (2014, 219) points out that “Hodgkin and Huxley could be thought of as specifying a state-space for neural processes. Later work on the molecular configuration of sodium and potassium channels described the same state-space using the language of molecular biology. Same models, same theory, completely different language”.

4.2.2. Agnosticism about the Metaphysics of Consciousness
Klein’s work might be beneficial for a variety of reasons—including, mainly, the ability for our explanatory theories to be agnostic about higher-order properties—but how does his work, and the semantic view of theories, have anything to do with the hard problem, the mind-body problem, and our explanations of consciousness? Well, while I have shown (in the previous chapter) the lack of justification required to adopt anti-cognitivist and dualist views about the nature of consciousness, and on the other hand, espoused cognitive, physicalist, and functionalist positions, I admitted that more empirical work must be done before we can guarantee the truth of any of the metaphysical positions. This is because, as I have said, we do not seem to be in a place where we can deem any of the options true via a priori methods.

However, this fact need not halt our progress toward explaining aspects of consciousness. We need not stand idly by while we wait for the truth to be given about the metaphysical nature of consciousness. This is for two reasons: First, because there remains left-over, independent explananda, even after we answer the mind-body problem by identifying what consciousness is and how it relates to the world. Chiefly, we must then understand how consciousness functions and how it is caused. One question (amongst many) might be about how systems (whatever they may be) come together to formulate unified experiences—which might be understood as a form of the hard problem or the combination problem. Another might be about how systems process information in order to generate conscious representational states which they can attend to. Regardless, there are a variety of questions about consciousness that remain even after we provide an answer to the mind-body problem. Second, similar to how Klein has shown that many of
our explanations can be understood as models which are agnostic about their internal higher-order terms potentially referring to actual properties, we have good reason to think that they can work the same way with lower-order properties—i.e. the types of lower-order properties which are involved within answers about the metaphysical nature of consciousness (its micro-constituents, substructure, etc.). And thus, we may make explanatory headway with the phenomenon of consciousness, even in the absence of answers to the more fundamental mind-body problem. This is because, through a semantic view of theories, we can emphasize the models that they create and remain agnostic about which particular properties or entities might underlie the concepts mentioned within our explanations, or if such properties exist at all. Let me explain:

Questions about boats floating, chairs holding up weight, computers and how they work, and why cars move forward when we step on the pedal, all remain even after we talk about what boats, chairs, computers, and cars are made of. This is because, despite boats, chairs, computers, and cars being made up of wood, metal, atoms, or bits, these left-over questions are situated at a higher degree of abstraction than questions such as “What is the fundamental nature of X?” Left-over questions might have to do with the functions or causal aspects of the things that they are about, and thus require specific explanatory details which are situated at a higher degree of abstraction. The degree of abstraction is a higher level one, as the lower level facts about the micro-constituents of boats and chairs fall outside of what is relevant within explanations about their functions and causal aspects. Again, all that must be involved within an explanation about how boats float, or how chairs hold up as much weight as they do, are higher-order facts about
buoyancy and shape, or about structure and geometry; not about the micro-constituents of such shapes or structures. Even with these cases, the Gricean maxims apply.

Similarly, inspired by the discussion surrounding Putnam’s (1973) “Meaning and Reference” paper, water may be understood in a variety of ways, regardless of what it is constituted by. Further, we may still provide explanations regarding the function of water, despite our ignorance about its constitution (whether it be H₂O or XYZ). We might even know that it is caused by X amount of components, without knowing exactly what those components are. Thus, through a semantic view of theories, our explanations (about certain features) of water might have been able to make progress while remaining agnostic about what water is constituted by.

This is because, as Klein (2014) suggested through the Gricean maxims and the semantic view of theories, our explanatory theories can serve as models of processes and mechanisms which involve functional and causal explanations, all while only including the details that are relevant to the demands of the question. Further, they may do so while staying agnostic about the specific details which are not relevant, and instead, either (i) leave the irrelevant details out entirely, or (ii) use placeholders within our models for the details that are not as important to the context of the question. To apply this to consciousness more specifically, we may (in theory) develop an explanatory model for how X’s combine to form conscious experience, where X is truly an undetermined variable (which can be something physical, fundamental and micro-physical, proto-phenomenal, etc.), so long as our explanation demonstrates that it is isomorphic between aspects of the model and parts of the world.
In the case that panpsychists attempt this, and replace X with a term which seems to refer to some actual property (like microphysical properties), our semantic view of theories still holds for lower-order properties (such as the micro-constituents of consciousness), just as well as it might for higher-order properties. The emphasis is not so much on the literal existence of such properties, as it is the models surrounding them. It may turn out that in attempting this, the panpsychists are deemed somewhat successful at developing a model of consciousness wherein the X is replaced with panpsychists terminology. However, it may also be the case that the panprotopsychists, physicalists, or whoever else, might develop an identical model while only using different terms to replace their X’s—similar to how Hodgkin and Huxley’s state-space used the language of neuroscience, only to have the same explanatory model recreated with the language of molecular biology. Regardless of whether or not such state-spaces, themselves, are multiply realizable, or if it is only the models that are abstract enough to be multiply realizable through inter-theoretical language, it makes little difference in terms of pragmatic concerns—an explanans is being provided in one way or another. We need not fuss over the underlying metaphysical structures of things if such details are not relevant to the demands of particular questions, and/or if our explanatory models can remain agnostic about such metaphysical structures while still achieving explanatory goodness.

Proponents of the multiple realizability of conscious mental states should see this as being an attractive view. Multiple realizability of both the explanations (as we have seen through Klein’s explanation of the Hodgkin and Huxley explanation) and the multiple realizability of mental states, themselves, form a much more elegant and attractive outcome. If mental states, themselves, are not multiply realizable (and only our
explanations are, through a semantic view of theories), then our explanations which use undetermined variables as placeholders are potentially non-exhaustive when it comes to the full nature of the mind. This is because we might develop explanations of consciousness through a semantic view of theories without an answer to the mind-body problem. And, once the mind-body problem has been solved, the X’s (the undefined variables) within our models can be filled in with further detail related to the mind-body problem. However, if mental states are multiply realizable, then our agnostic models of consciousness actually do all of the work that is required. The placeholders within the models can remain placeholders, as they refer to a variety of different things.

The most notable example of an argument in favor of the multiple realizability of mental states comes from Putnam (1967). Putnam observes that there are a great deal of creatures who seem capable of experiencing pain, but notes that our psycho-physical identity theories (which identified mental states with specific neuro-biological brain states) cannot accurately map themselves across the wide range of creatures who experience pain yet possess different types of physical bodies and brains than our own. In theory, there may be silicon robots or aliens with even more radically different physical make-ups, who are still able to experience pain. And thus, when we talk about pain, we may very well be talking about a higher-order concept—perhaps a functional state. Importantly, it is not yet clear if conscious mental states are indeed multiply realizable in this way. However, regardless of whether or not they are, the semantic view of theories shows promise that our explanations, themselves, may be. And if so, they can remain agnostic about the existence of the properties that their terms and concepts may or may
not refer to. In doing this, the explanatory terms could either be left out entirely if they are not relevant, or used as undefined variables to be filled in at a later point, if at all.

To sum up: Our explanatory theories of consciousness can safely do without ontologically committing themselves to any particular nature of consciousness, or committing to any particular micro-constituents of consciousness, regardless of whether or not consciousness itself is multiply realizable. This is because, through a semantic view of theories, our explanatory models of consciousness may effectively utilize terms as undefined variables and remain agnostic about what properties they might refer to, if at all. Rather, they can focus on the models which serve to explain function and causation, as opposed to the metaphysical properties or entities which can be identified with, or relate to, the mind.

4.3. Conclusion

Suppose that we wake tomorrow morning to find that there is a sudden scientific and philosophical consensus that X metaphysical theory has solved the mind-body problem by explaining what consciousness is identified with, constituted by, and how it relates to the world. Whether X theory be panpsychism, physicalism, or so on, there remains a variety of questions and explananda related to consciousness at a much higher-level than that of the fundamental nature of the phenomenon. Some, for example, might be about how independent mental states come together into a unified conscious experience, in their own right; or how representational states allow us to understand the
external world around us through perception. Whatever consciousness turns out to be within nature, we must also understand how it functions and how it is caused.

Here, I have motivated the semantic view of theories, wherein our language is not taken literally within our explanations, but instead, it is our models that are emphasized. In adopting this view of theories, we can make headway on many of the problems related to consciousness without committing ourselves to any particular metaphysical properties which might underlie consciousness, and thus, we do not risk getting stuck on persisting philosophical problems like the mind-body problem. We may, in the process, remain agnostic about whether or not our terms and concepts within our theoretical models actually correspond to precise ontological properties, or which specific ontological properties they may refer to. This way, we may benefit from an interdisciplinary and inter-theoretical approach to building explanatory models of consciousness, as each discipline is welcome to include their own disciplinary language without experiencing competing ontologies between them.
CHAPTER 5
Conclusion

Before embarking on the task of explaining a particular phenomenon, it is beneficial to first understand what is required from an explanation of it. For the phenomenon of consciousness, it turns out that incredibly influential works have inflated what is apparently required from our explanatory theories of it. I have listed a few of the most problematic expectations of what our theories must involve and, for the sake of convenience, classified them under what I call the Inflated Approach. Throughout this project, however, I have attempted to deflate what is required from our explanatory theories of consciousness in an attempt to deconstruct each facet of the Inflated Approach. To do this, it is first important to understand what details are relevant within our explanations of consciousness and how we can approach the phenomenon.

First, I have analyzed questions which have been prompted by Nagel (1974). Such questions have to do with what particular experiences are like to experience. The hope was that if consciousness simply is defined as mental states of which there is something it is like to experience, then through an understanding of what experiences are like to experience, we can come to understand what must be understood about consciousness. However, I have argued that answers to such questions (about what an experience is like to experience) prepare us to provide mere description, and lack any explanatory power which would be relevant for our explanations of consciousness. This is because in giving an account of what an experience is like to experience, we are required to provide a description of the characteristics and qualities of a given experience—not to provide any explanatory details which make the function or cause of consciousness any more intelligible. Furthermore, our explanations of consciousness need not incorporate any
description or explanation of what experiences are like to experience, as we can do all of
the explanatory work without the inclusion of such facts. And thus, such questions about
what experiences are like to experience remain as a distinct task from explaining
consciousness.

Second, I have analyzed the hard problem and the extra ingredient as a proposed
explanans. I have concluded that the hard problem, itself, is a neutral explanandum of
how and why conscious experience arises at all, and that it is a genuine explanandum
related to consciousness. However, I further analyze our means of answering it, and
conclude that anti-cognitivist approaches to consciousness which are motivated by the
proposed extra ingredient are dubiously justified. This is because they are only advocated
for by a priori thought experiments which beg the question. Interestingly, even the
counter thought experiments (thought experiments which mimic the form of the original
thought experiments, but which generate opposite conclusions) are just as plausible, and
thus, a priori thought experiments are not a fruitful line of inquiry when it comes to
understanding consciousness. However, I also motivate cognitive approaches to
consciousness by arguing against the existence of a potential extra ingredient and indicate
that cognitive approaches are the default position until a more promising approach can be
justified. In the absence of good justification for anti-cognitivist approaches, and through
the explanatory benefits of cognitive approaches, we have good reason to bet on the
success of cognitive approaches to consciousness.

And finally, I have attempted to shed another peripheral target of explanation: the
mind-body problem. I argued that a variety of theories which purport to explain
consciousness and provide answers to the hard problem miss their target, and instead,
overly commit themselves to answers to the mind-body problem by giving accounts of
the metaphysical nature of consciousness, its micro-constituents, and what it can be
identified as/with within the world (be it brain states, functional states, microphysical
states, or so on). When our theories of consciousness do this, they fail to answer a variety
of left-over questions that are specifically related to consciousness at more abstract
levels, but they also leave themselves susceptible to attack—rendering them stuck in
what some call a metaphysical standoff which hinders our explanatory progress (Balog
2008). However, I have motivated the semantic view of theories, wherein our explanatory
theories can operate akin to mathematical models by viewing the terms and concepts used
in explanations as undefined variables. This allows us to remain agnostic about the
particular substructure and nature of consciousness, while allowing us to continue to
make progress toward explaining consciousness by emphasizing the models which are
featured within explanatory theories, as opposed to the terms within them and whether or
not they commit us to any ontological properties.

Chiefly, it turns out, the task of explaining consciousness need not be as tall of an
order as we might have suspected. The tasks of explaining what experiences are like to
experience, identifying a potential extra ingredient as a possible explanans, and
incorporating answers about the persistent mind-body problem, are all either irrelevant,
unjustified, or peripheral tasks when it comes to explaining what needs to be explained
about consciousness—how it functions and how it is caused.

I believe what is most important about this project is that it has reevaluated how
we approach consciousness—a terribly confused and convoluted phenomenon about
which we seem to be able to make very little explanatory progress due to stalemates. By
starting again at ground zero and critically breaking down what our explanations of the phenomenon must involve, we may begin anew, and approach the phenomenon with much more promise. Further research may add on to this project or work from it. The former, I suspect, might do further research regarding additional explanatory inquiries which continue to lead consciousness researchers down the garden path. In addition, as I have emphasized what our explanatory theories need not involve, further research might also evaluate what it is that our explanations of consciousness need to involve, in light of the work that I have done here. As for the latter, further research might advance from this project by attempting to make headway toward consciousness by understanding more of the features that our explanations of consciousness must involve and attempt to provide answers about those features. Importantly, researchers may be confident that they are not “failing to take consciousness seriously” when dismissing peripheral explanatory inquiries which the proponents of the inflated approach believe we must address within our explanations of consciousness. Rather, in identifying exactly what our explanandum is, and what our expectations from an explanation of consciousness are, we can then begin to really take consciousness seriously and begin to explain it, given that we have started to understand what is required in terms of answers to our proposed inquiries.
Works Cited


