Neutral Monism Beyond Russell

By

Michael P. Schon

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Philosophy)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: May 5th, 2023

The dissertation is approved by the following members of the Final Oral Committee:

Alan Sidelle, Professor, Philosophy
Lawrence Shapiro, Berent Enç Professor, Philosophy
Farid Masrour, Associate Professor, Philosophy
Martha Gibson, Faculty, Philosophy
Robert Streiffer, Professor, Philosophy and Medical History & Bioethics

For Grandpa,

Who always told me “Study hard and maybe you’ll be as smart as me one day.”

Maybe. I’ll keep trying.

Acknowledgements

I would like to thank Alan Sidelle for reading numerous drafts of every chapter of the dissertation and giving me in-depth and valuable feedback on every part of it. Without his comments, there would be far more mistakes and confused ideas than there are in this final version. All mistakes and confusions that remain are surely because I haven't understood some comment or other. And without his encouragement, I probably would not have finished writing the thing. I'd also like to thank Farid Masrour and Larry Shapiro for valuable comments on earlier drafts of several different chapters throughout the dissertation. Thanks also to Martha Gibson and Rob Streiffer for agreeing to be on the oral defense committee even though I am sure they have better things to be doing. And to Martha Gibson and Dennis Stampe, thanks for letting me watch the farm while you were away. It has been a writing sanctuary for me throughout the years. Thanks to Paula Gottlieb for always taking the time to ask how my own work was going when I was her TA. She didn't have to ask, and it made me feel like my own work was valuable to her even though she doesn't work on my area of philosophy.

And to my fellow graduate students, alas, ten years is far too short a time to study among such excellent and admirable philosophers. I don't know half of you half as well as I should like, and I like less than half of you half as well as you deserve (to adapt a phrase from a certain Hobbit). Your camaraderie has kept me sane during some difficult times, and your hard work has inspired me to keep plugging away even when I didn't feel like it. I'd like to especially thank Greg Nirshberg for his friendship over the years, and for suffering through endless discussions about this stuff even when there might have been better things going on, like eating tacos and drinking margaritas. Dylan Beschoner and Hubert Marciniec have also provided interesting and valuable conversations every time we talk. Even when it might not seem directly relevant, their insights have influenced how I think about the mind-body problem in important ways. Thanks also to everyone who has ever asked about my dissertation, even just in passing. It's nice to think that somebody might have an interest in what I think and what I work on.

And last, but not least, I'd like to thank all of my family and non-philosopher friends for reminding me that there's more to a good life than writing a dissertation and encouraging me to

pursue all of my other interests even while I struggle with my writing. It might mean that it takes me longer to write, but the writing is better for it in the end. And more importantly, so am I.

Table of Contents

Outline: A Roadmap of the Dissertation v

Chapter 1: The Beginning 1

1. A Question 1

2. Philosophical Developments 8

3. The Puzzle 13

Chapter 2: The Impasse 19

1. The Causal Argument 19

2. The Epistemic Argument 28

3. A Brief Note on Parsimony 41

4. Proposing a Way Forward 42

Chapter 3: Against the Russellians 45

1. The Best of Both Worlds 45

2. The Basics of Russellian Monism 47

3. Varieties of Russellian Monism 52

4. Problems with Any Formulation of Russellian Monism 60

Chapter 4: Toward a Non-Russellian Neutral Monism 66

1. A Path Forward 66

2. Epistemic Neutral Monism 68

3. The Puzzle – Final Version 79

Chapter 5: Conventionalist Neutral Monism 81

1. Revisiting the Neutral Monist's Dilemma 81

2. The Problems with Strong Necessities 83

3. Conventional Properties 91

4. Not a Dualism 98

5. Not a Materialism nor an Idealism 106

Chapter 6: The Answer 111

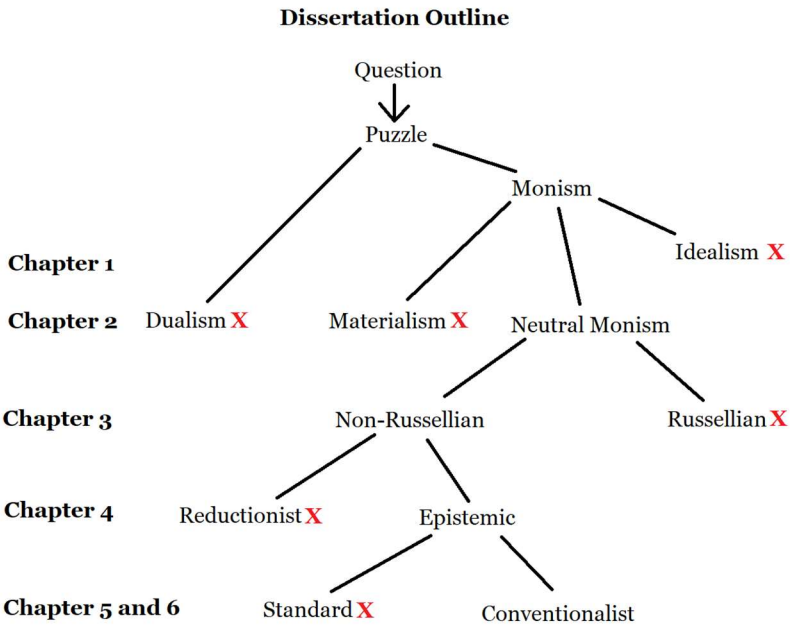
1. Conventionalist Neutral…Monism? 111

2. Re-examining the Epistemic and Causal Arguments 115

3. Concluding Remarks 121

References 124

Outline: A Roadmap of the Dissertation



The dissertation begins with a question: Are the mind and brain the same or different?

I spend Chapter 1 showing how this simple question leads to a philosophical puzzle known as the mind-body problem. By way of explaining the puzzle, I show that there are two demands that a satisfactory solution to the puzzle must meet – what I call the Ontological and Explanatory Demands. I also explain why ontological idealism is not an attractive solution to the problem. Along the way, I refine the starting question into something more familiar to current discussions of the mind-body problem: Are phenomenal properties fundamentally the same or fundamentally different from physical properties?

Chapter 2 develops the impasse that contemporary philosophers find themselves in by discussing the Causal Argument against dualism and the Epistemic Argument against materialism. Neither materialism nor dualism seem capable of meeting both demands from Chapter 1. Thus, the rest of the dissertation is an exploration of a fourth, non-standard solution – neutral monism – that hopes to meet these demands.

Chapter 3 shows how the most popular way to develop neutral monism, called Russellian monism, does not work. The basic problem is that most ways to develop this view falls to a dilemma such that it is either incoherent or really just one of the more standard solutions to the puzzle in new dressing. The final two options – Reductionist Russellian Monism and Epistemic Russellian Monism – have the capacity to accomplish the goal of neutral monism, but the uniquely Russellian aspects of the views do nothing to help accomplish this. As such, I leave Russellian monism behind and focus on non-Russellian views.

Chapter 4 argues that reductionist neutral monism cannot work. Since it relies on reductionism, it faces a different dilemma. Either the phenomenal can be reduced or it cannot. If it cannot, then obviously

reductionist neutral monism is false. If it can, then there is no motivation for supposing that it reduces to some neutral stuff rather than reducing to the physical. So, the only possibility is to attempt to develop an epistemic neutral monism. I spend the majority of the chapter explaining what this style of solution might look like, paying close attention to very similar types of answers known as *a posteriori* physicalism and the powerful qualities view. The result is that epistemic neutral monism as standardly employed also can't be made to work, but I suggest one final possibility that relies on a conventionalist metaphysics.

Chapter 5 is where I develop an account of conventionalist neutral monism and explain how it avoids the common dilemma raised against Russellian monism. Here I show how a conventionalist neutral monism is both coherent and substantially different from the other solutions to the mind-body problem.

Chapter 6 relates the solution developed in Chapter 5 to the question raised at the beginning of the dissertation and shows how a conventionalist neutral monist ought to answer the question. This answer does two things. First, it helps explain what goes wrong with the standard understanding of the Causal Argument and the Epistemic Argument that creates the appearance of an impasse. Second, it shows how conventionalist neutral monism successfully meets the demands developed in Chapter 1.

Chapter 1: The Beginning

> "Criticism of an inadequate theory that does not deal with the deeper convictions lying behind it will generally result only in another equally inadequate theory."
>
> - David Ray Griffin, *Unsnarling the World-Knot* (1998, p. 23)

1. A Question

This investigation begins with a question: is the mind the same as the brain? This is a question that any inquisitive person in modern society will be familiar with, and yet it is one for which a satisfactory answer continues to elude philosophers and scientists. The philosophical and scientific developments of the past century or two have refined our understanding of what the right question is and where people tend to look for a satisfactory answer, but nevertheless, every answer that has thus far been given leads to a philosophical puzzle. Solving this puzzle in a satisfactory way is the goal of this dissertation. In this chapter, I'll show how the philosophical developments lead to the version of my starting question that is currently the focus of investigation in Anglo-American philosophical circles. I'll then show how this question, and the most popular unsatisfactory answers, help show us what is required of a good answer and thus creates the current Anglo-American version of the puzzle known as the mind-body problem. Finally, I'll use the constraints and the puzzle to show the direction I intend to take the rest of the dissertation.

For well-trained philosophers, it might seem obvious why the question of whether the mind is the same as the brain is such a long-lasting puzzle, but I think it's worth it to state the "obvious" in order to see how the puzzle begins. The way the question has been posed above is vague and inadequate. But by exploring some answers to it, we can refine it into a question that will be lead to the important philosophical debate. In order to illustrate this, let's begin with a simple answer and say, "Of course they're different! One just has to look, and the difference couldn't be more obvious." Many people will even state something like this when hearing the question for the first time. But this misses the point of the question. We don't want to know if they *appear* different, but rather we want to know if they *are* different. When we look, we see the appearance. And they do certainly appear different but whether they are *really* different is the question. Of course, there are many instances where it appears as if two things are different, but they turn out to be the same. For instance, despite appearances, water and ice are the same, the Sun is the same kind of thing as a star, and (arguably) Bruce Banner and the Hulk are the same. And if we were to ask whether ice and water are the same, we can see that there are certain apparent differences, but nevertheless these don't necessarily settle the answer. What we want to know is whether ice and water share something fundamental in

common. So, to avoid this kind of spurious answer based purely on appearances, a better starting point is the following question: Is the mind fundamentally the same as the brain? Even the addition of this one word has clarified what we intended to be asking to begin with. It avoids the simplistic answer that misses the point of the question above. When we ask whether the mind is the same as the brain, we want to know whether they are the same at a deeper level, a fundamental level. That is what we mean when we say that we want to know whether they are *really* different, rather than merely appearing to be different.

Nevertheless, one might insist that, in this case, the differences are too great to reconcile; the mind really is different from the brain. This was the answer of René Descartes in the 17th century and is thought of as the traditional answer.[1] This answer has come to be known as mind-body dualism. Dualism has the benefit of providing a really simple explanation for why the mind and brain appear different. They appear different because they *are* different. Unlike water and ice, there is no underlying chemical structure to be discovered that might make this apparent difference go away. The problem with this answer is that it's not clear how the mind might interact with physical objects, or more poignantly, with the body. Someone might liken this problem to the mixing of oil and water. Try as hard as you like and they will never mix, they are simply two different kinds of liquid. However, the problem with the mind and body is of a deeper concern for the dualist. They say that the mind and body are fundamentally different in all respects. Oil and water, though different enough to be unmixable (without an emulsifier), are not so fundamentally different. After all, oil floats on water. When the water goes up so does the oil, and if one were to put oil under the water, the water would be moved by the oil as it travelled up to float on top of it. But in saying that the mind and body are truly different, the dualist maintains that they are so dissimilar that the mind is not even made up of physical objects. If it were, then the obvious candidate for what makes it up would be the brain and then

[1] Starting the story with Descartes follows the way undergraduate philosophy courses in the United States are often taught, but it simplifies things in two important ways. The first thing to note is that the puzzle has been around much, much longer. Moreover, the dualist solution that Descartes offered was not new when he wrote it, not even to him. He was well-aware of scholars before him who had offered up substantially the same solution. Nonetheless, the way in which he presented the solution and the thoughts that it provoked did help to transform the problem into the one that current philosophers recognize. The second way in which the main text oversimplifies the situation is in the suggestion that there is one traditional answer. Perhaps this is true of European philosophy in the modern era, but it is far from true that dualism is the traditional answer in all philosophical cultures. For instance, the solution that I will eventually arrive at is a close relative of the traditional Madhyamika Buddhist answer which is not to be found much (if at all) in the sphere of Western philosophical thought. I wasn't aware of this until very recently, though, so the way in which I arrive at my answer differs substantially from the Madhyamika path to their answer. As of yet, it's unclear to me how similar my position ends up being to the Madhyamika position, but it's clear that there are significant parallels. Another example of a traditional answer that is not dualist is found closer to home in current scientific and philosophical thought. The standard has become a materialist solution against which many recent arguments have been leveled.

the mind and brain wouldn't be so different after all. Thus, the mind, being fundamentally different from matter (and therefore the brain) does not react to matter in ways that can be described by the laws of physics. So, one is left to wonder how they interact, if at all. This seems to split the world in two side-by-side worlds which goes against both everyday experience as well as our most comprehensive scientific theories of the world. If appearance is to be our guide, then the appearance of pain causing me to wince should tell us that the mind has *something* in common with the physical world. The result is that this simple answer leads very quickly into some questions that are far from easy to answer.

On the other hand, we might begin with a different simple answer to our starting question and say, "Of course the mind and brain are the same! They may appear different, but they are the same at a deeper level." Many people also state this upon hearing the starting question. The difficulty here is in proving that they are the same while respecting the apparent differences. One model for this might be something like mass and energy. With mass and energy, Einstein's famous equation – $e=mc^2$ – demonstrates that mass and energy are the same at a deeper level than appearances. It also allows us to explain their superficial difference. A simple way to state the equivalence is that measuring mass (at rest) is just one form of measuring the energy that an object contains (along with its kinetic energy, potential energy, etc.). So, what we call "mass" is really just one form of energy. But the very large constant – "c^2" which is the square of the speed of light – makes energy appear to be something very, very different from mass. Another way to explain their sameness is that energy gives rise to the phenomenon of mass when condensed by the very large constant of c^2. Thus, while these are two different phenomena, they are fundamentally the same in that one is based wholly on the other. As we can see, the relevant notion of fundamental sameness here is not meant to imply identity but instead merely means a deep connection. Recent developments of the various mind sciences – psychology, neuroscience, cognitive science, etc. – have shown some deep connections between the mind and brain. This recent success has given hope that an explanation like Einstein's mass/energy explanation can be found. However, despite some scientists' confidence in the possibility of an explanation of the apparent difference, no one as yet has been able to find one. There is no simple "mind=brainc^2" equation or anything analogous. At least, not one that is so uncontroversial as $e=mc^2$.

There are two standard ways of approaching this issue, and they bear some resemblance to the above explanations of mass and energy. In a simplified way of explaining the equivalence, mass is simply one type of the more fundamental type, energy. Trying to show how the mind and brain can be the same

despite an apparent difference has led to two similar proposals. One or the other of the mind or the brain are proposed as fundamental or "real" while the other is illusory or derivative. The more common of the two is to say that the brain is the really existent thing in the world and that the mind is an illusion or somehow arises out of the more fundamental brain. This is mind-body materialism or physicalism.[2] The less common, but still standard, monist answer says the opposite of materialism.[3] That is, it states that the mind really exists while the brain is an illusion or somehow depends on the mind for its existence. This is mind-body ontological idealism or simply idealism. These two standard approaches show a similar way of attempting the required explanation of the sameness of these two apparently different things. They both state that one (either the mind or brain) fundamentally exists while the other is somehow derivative on this more fundamentally existing kind of thing. This is a promising beginning of an explanation of the sameness between the mind and brain because there really exists only one type of thing. The other type of thing is an illusion or arises out of it. If it works, this could explain their sameness while explaining away their apparent difference. Unfortunately, both types of solutions (idealism and materialism) quickly run into problems.

Idealism strikes many people (perhaps most) as quite a bizarre view as it runs counter to much of commonsense. Nevertheless, there are solid philosophical reasons that might be used in support of idealism. Amongst the most important of them is a worry that we may be incapable of knowing anything about the world outside of our own minds. Strangely enough, it is the very illusions, mirages, and dreams that make many laypersons think that material objects must be real that motivates the skepticism of ever being able to know anything outside of our minds with any certainty. In the same way that mirages turn out

---

[2] Some philosophers have used a distinction between materialism and physicalism. On this distinction, materialism refers specifically to views that operate under Newtonian physics in which the most fundamental stuff is matter and everything reduces in some way to matter. Physicalism, correspondingly, refers to the more recent version of this view but includes such things as electromagnetic fields and nuclear forces alongside matter. Surely, this is a useful distinction if one is primarily comparing or contrasting various historical figures or the views that they held. It might also be important in cases where energy is importantly contrasted with matter given the aforementioned reduction of mass (and thereby most understandings of matter) to energy, but for my purposes, the distinction is not so helpful. As will be seen in the next chapter, the standard argument against materialism applies just as well to physicalism. Thus, the (apparent) difficulties of understanding how the mind could be accounted for materially are equally difficulties for understanding how it could be accounted for physically. So, unless otherwise specified, I'll use the terms materialism and physicalism interchangeably.

[3] There is a third, non-standard monist answer called neutral monism. It is non-standard both in the sense that it is underexplored and in the sense that it is often not taught as a solution at all in courses that introduce the mind-body problem. Here, I want to limit my discussion to more standard solutions, and so I don't mention the idea of neutral monism. Don't worry, there will be plenty of discussion of neutral monism in the remainder of the text, especially in later chapters.

to not be there, *all* material objects might turn out not to be there. Thus, any apparent knowledge of material objects is unjustified and must be met with skepticism. Worries about this skepticism can be avoided if all that exists is in fact a product of our own minds. In essence, what is a very natural thing to say about mirages may very well be true of everything. It is *all* just in our heads. And since we know the things in our own heads, worries about how we could know other things can be avoided. Thus, by supposing the mind is the foundation upon which the world is built, it makes it possible to have certain knowledge about the world.[4] So, if one starts from a conviction about the unity of the world and combines it with a serious aversion to not being able to know what is in this unified world, idealism starts to become somewhat more of a natural way of thinking. Being all of one type of thing makes the unity of the world quite easy to come by and if that one type of thing is mental, then it seems quite easy to know about it, given the fact that it is the mind that does the knowing.

However, this leaves idealism the difficult job of explaining matter in a world that is otherwise entirely mental. Indeed, there is the troubling fact that material objects are accessible to more than just *my* mind or *your* mind, but they are generally accessible to everyone (given that they're in the right location, using the right faculties, etc.). Moreover, material objects (largely) remain stable even when no mind is concentrated upon them, and they even enter into causal interactions without the operation of any minds. For instance, if I were to leave this room and subsequently die, my coffee and the table it rests on would remain for you to find them despite no other mind being in contact with them. The coffee would grow cold but remain in the mug on the table. And since I would be dead, no mind would be there to be the foundation of all of this. Even more fantastical material events might happen that no one has ever predicted; an unpredicted comet might descend from the sky and destroy the coffee, the mug, the table, and everything else around it. If idealism is to be a satisfactory answer, it has to account for these aspects of matter that differ greatly from the mind and explain how they arise from the mind.

Perhaps the most famous idealist account to attempt this is Berkeley's. He posited that the ultimate mind, God, is what makes matter coherent and persistent in the case of our own lack of individual experiences. Nothing is ever outside the mind of God and so the coffee and the table it rests on would still

[4] According to my understanding of George Berkeley, this is the primary motivation for his ontological idealism, which has become the canonical version of the view. However, my brief presentation is extremely oversimplified and so you should not try to draw any conclusions about Berkeley's much more robust arguments and views from what is presented here.

have the mental foundation upon which idealism says everything rests. It is not *our* experiences which create the material world but God's. This does get idealism out of the hole, but its invocation of a supernatural entity makes it hard to swallow. Moreover, it undermines the initial reasons for pursuing this line of thought in the first place, avoiding serious skepticism about how we could know anything in the world. If one finds that the avoidance of skepticism is of central importance, then invoking God's mind as the foundation of the world does not help. There is nothing less knowable than the mind of God, even to firm theists. Insofar as there are idealist accounts of matter that avoid invoking supernatural entities, they face the same problem of how to explain the features of matter that differ from the mind.

While idealism faces the challenge of explaining how matter arises from the mind, materialism faces the opposite challenge. It must explain how and why it is that the mind arises from the material workings of the brain. Many scientists and philosophers have been driven by the success of the sciences, particularly the sciences associated with the mind and brain, to conclude that everything can be explained in terms of the matter that composes it. As such, the materialist answer has come down to the claim that the mind just is the brain, or at least that the mind arises out of the more fundamental material structures and functions of the brain. However, as I've already mentioned, giving such an explanation is quite difficult since the mind and physical objects appear to be so fundamentally different. For instance, at a first glance, the brain is extended in both time and space, while the mind, it would appear, exists in time but can be found nowhere in physical space. One can point at a brain, but not a mind. Of course, if materialism is right, then one *can* point at a mind by pointing at the corresponding brain. But we need some indication of why or how it is that the brain gives rise to the mind if this is the case.

One reason for the current popularity of materialism is due to recent trends in other sciences where non-material objects have been disposed of without any loss of explanatory power. For instance, materialists often point to a somewhat analogous controversy about living things in biology around the turn of the 19th century. At the time, it was proposed that some vital force – often referred to as the *élan vital* after Henri Bergson's (1907) *Creative Evolution* coined the term – gave creatures life and drove evolution, thus distinguishing between living things and non-living things. This theory of life is analogous to the brain and mind in that the mind is what distinguishes sentient creatures from non-sentient creatures. However, developments in genetics and organic chemistry have shown that the distinguishing factor between the living and non-living is merely an organization of various chemicals to become genes. Thus, if the *élan vital*

exists at all, it is really just the chemical combinations that make up genes. So, if we were to ask if life were fundamentally different from non-life, we can use the chemical organization of genes to prove that they are fundamentally equivalent despite some obvious differences. Living things just are a certain type of matter organized in a way to have certain functions. In the same way, materialists pose that we are on the way to showing that the mind is not fundamentally different from the matter of the brain; a sufficient understanding of the organizational features of the brain will give us the proof that we need.

There are two problems with this kind of inductive argument, though. First, where a few decades of research into what distinguishes life from non-life definitively disproved the *élan vital*, several more decades of much more advanced research into the nature of the brain have produced no such definitive results. This may simply be a difference in relative complexity. As many scientists would be quick to point out, the brain is much more complex and harder to study than living creatures. However, the second problem with this kind of argument gives us reason to not be so optimistic that more research into the brain will produce what is necessary to show that the mind and matter are the same. Namely, the situations we are in with respect to the mind and life are not so similar as materialism would require. David Chalmers convincingly argues in his (2010) book *The Character of Consciousness* that where life can be fully explained by explaining all of the functions that living creatures perform, the mind is quite different. Even if we could show how the brain does all of the things that we ordinarily think of as being done by the mind, there would still be something important left out, namely the experience that we have when doing the things that we do. These experiences are one of the defining features of the mind.

To be clear, this leaves a great many things for the brain to do; it causes me to raise my hand when I have a question, recoil from the hot stove, shout in pain when I run my shin into the coffee table, etc. The promise of the various mind sciences is that all of these things can be explained by understanding the organization of the brain, neurons, and the body's nervous system. If the mind were perfectly analogous to life, this would be all it would take to explain how the mind arises from matter. But explaining all of what the mind does would still fail to explain the corresponding feelings of all of those actions – the experiences of confusion, heat, or pain. Even if we can show how the brain produces all of the actions and performs all of the functions of the mind, we still don't know how these experiences arise or, indeed, *why* these experiences exist at all if the functions can be performed by the brain without any need of the experiences.

Just as idealism has trouble explaining the persistence and public availability of material objects, materialism struggles to explain the first-personal feelings and experiences of the mind.

This shows us a second way in which our initial question should be refined to make sure that we are asking about the relevant point of disagreement. Instead of merely asking about the mind and brain, we need to be clear that the relevant potential difference is between mental experiences and what happens in the brain or anything material. So, we can further refine the starting point to be: Are mental experiences fundamentally the same as certain things that happen in the brain? This certainly gets closer to asking the relevant question while avoiding some answers that aren't relevant to what we really want to know.

Asking the question in this way reveals that the general problem for monism is that our experiences and what happens in the brain appear to be two wholly different types of things. So, it is mysterious how one might give rise to the other even if general considerations make us think that they have to be fundamentally the same. To take up the example of mass and energy again, Einstein's equation shows us how the apparent difference between mass and energy comes about despite their fundamental equivalence. Multiplying or dividing by the speed of light squared puts energy and mass on two massively different scales such that we cannot help but see them as incredibly different under normal conditions. It is only with very finely tuned measuring instruments that we can see their equivalence. As such, it is impossible to understand this on the basis of instinct alone. In a way, monism's general difficulty is simply the mirror image of the problem that dualism runs into. Where dualism has difficulty explaining how these two fundamentally different things interact in one unified world, monism has difficulty explaining why the phenomenal and physical appear to be so fundamentally different.

2. Philosophical Developments

There are three crucial developments in philosophy over the past century or so that are important to discuss before we finish refining our initial question and move on to attempting to answer it. The first development is one concerning the exact features of mental experiences and the events in the brain that are central to our question. So, we need to discuss in more detail what are the features of mental experience and what happens in the brain[5] that are relevant to what we really want to know about. Second, though

[5] One recent development that I won't discuss is whether it is really the brain *alone* that might be the relevant part of the physical world. There has been a move in psychology and cognitive science to start discussing not just the brain but also include various things external to it, either including the body as with the research program called embodied cognition or going further and including things even outside the body as with the related idea called the extended mind.

idealism has had something of a resurgence recently, the new discussions fall short of providing new reasons for accepting the view. As such, we should discuss why idealism is not an attractive answer to explore. Third, we should likewise discuss why panpsychism is not also an attractive answer. Both of these answers have seen a recent resurgence, and they both have some philosophical merit to them, but in my opinion, we'd be better off spending our time exploring more promising answers to the question before turning to these. Obviously, the second and third developments are less important for refining the question. Instead, they are important to show why certain types of answers won't be explored before getting side-tracked on them.

We've already seen how the move from asking about minds and brains to asking about mental experiences and what happens in the brain is a helpful refinement of the question. It helps us to see just what it is about the mind that makes it seem so different from the brain. Similarly, it is important to focus on what it is about mental experiences that seems so different from what happens in the brain. There are many dimensions along which these things seem to differ, but it has been agreed that really only one dimension of difference captures the generally puzzling features discussed above.[6] The part of our experience that seems to be so different from what happens in the brain has come to be called the *phenomenal* properties of our experience. These properties are those parts of our experience that give our experiences the particular feelings of what it is like to have those experiences. For example, the phenomenal property of a painful experience, like stubbing your toe, is the painfulness that you feel upon stubbing your toe. In other words, it is simply what it is like to go through that experience. What seems different about what happens in the brain is that the brain is made up of various physical properties that act in a multitude of ways when different external and internal stimuli are present. For instance, when you stub your toe, not only do you have a feeling of pain, but there is a whole cascade of electro-chemicals passing along routes from the nerves in your toe to the neurons in your brain. For pain, it has become the norm to gloss whatever the physical happenings are in the brain as "c-fibers firing", supposing that the psychologists and

As will become apparent shortly, it doesn't really matter what part of the physical world we think might be the same as the mind. So, as interesting as these developments might be, they will not concern us here.

[6] Herbert Feigl's (1967) essay *The "Mental" and the "Physical"* does a great job explaining why many dimensions of difference do not really capture what is interesting for the philosophical puzzle at hand. In general, these other dimensions of difference can either be found in both experiences and what happens in the brain (e.g. the quantitative vs qualitative differences) or else build in certain assumptions about what counts as experience and what happens in the brain such that a solution to the problem we are discussing is built in by definition (e.g. purposive vs. mechanical differences). I won't re-hash what has already been discussed there.

neuroscientists can fill us in on the details about what is going on. It is these physical properties of what happens in the brain that we want to know whether or not they are fundamentally the same as the phenomenal properties of our experiences. They happen at the same time, have the same causes and effects, and vary isomorphically with respect to important qualitative and quantitative aspects of our experiences, so it's natural to wonder whether they are fundamentally the same or whether they are as different as they at first appear. So, we can refine our starting question yet again to "Are phenomenal properties of experience fundamentally the same as the physical properties that happen in the brain during such experiences?"

One final clarification of the differences between phenomenal and physical properties will give us the question that current Anglo-American philosophers have in mind when they attempt to answer this question. This clarification involves giving a more precise definition of what we mean by phenomenal properties and physical properties in order to make sure we know what we are talking about before proceeding to attempt an answer. The exact definition of these is not particularly important for the rest of the dissertation, so a rough idea will suffice. Nevertheless, we should be clearer than we have been thus far. As far as I see it, the basic distinction between phenomenal and physical properties is that phenomenal properties are subjective while physical properties are objective. This makes sense of the apparent difference between what it is like for you to stub your toe – which is something only you can confirm or describe – and what is going on in your brain at the same time – which is something that anyone in the right location and with a suitable apparatus can confirm or describe. With this in mind, we can say that phenomenal properties are best understood as those properties which only one individual can confirm exist or describe in any detail while the physical properties are those properties which can be confirmed to exist by many people and described equally well by all of them, assuming they have a suitable position (and possibly observational apparatus) to observe them.[7] Thus, we have all that we need to pose the question the current Anglo-American philosophers mean to ask when they say, "Is the mind the same as the brain?" The more precise question, and the question that I will attempt to answer, is this: Are phenomenal properties –

[7] I take it that this agrees with such statements by people like Strawson (1994, 2003) who say that the appropriate distinction is between mental and non-mental or experiential and non-experiential rather than mental and physical. There, I gather that the offense of using "physical" to contrast with mental is that physical is taken to mean "the items of current physics" which is, he thinks (and I agree), too limited in scope. My only disagreement with him on this is that we cannot merely leave the contrast with one positive understanding and one negative understanding. That leads us to think that we know less than we do about the physical; we know at least this much, anything that counts as physical is intersubjectively confirmable.

i.e. the properties of experience that can be confirmed and communicated by only one subject – fundamentally the same or different from physical properties – i.e. properties of experience that can be confirmed by multiple subjects and/or from multiple perspectives?

Before moving on, I would like to explain why I think two historically important answers to this question are dead-ends, and thus why I will not pursue them here. First, the non-materialist answer that I mentioned earlier, idealism, has an insurmountable problem. This problem is the central problem of idealism, there just isn't an adequate story about how physical properties could arise from phenomenal properties. Since they can be confirmed and described by many subjects, physical properties are both persistent over time and invariant across observers. However, phenomenal properties are exactly the opposite of this. In fact, one instantiation of a phenomenal property is available to only one person (and plausibly enough, only at one time). The solution to this problem offered by George Berkeley invokes God, who has a subjective experience of everything. So, when physical properties are stable over time and invariant across observers, they are that way because each physical property is a single phenomenal property of God's experience of the world. This solution does provide an explanation that seems to work. But as I mentioned in Section 1, the invocation of God here causes more epistemological concerns and does not really solve those that it purports to. While God might provide intersubjectively available truth conditions for statements about seemingly materials objects, those objects don't seem to be genuinely intersubjectively available. They are only available to those who have access to God's mind, which seems to only be God himself. And even if these concerns can somehow be avoided, this solution is not one that I think we should accept.

We should not accept Berkeley's solution because of his invoking God as a central part of the explanation. This is not because God does not exist or even because there isn't enough evidence to prove his existence. Rather, the reason that God should not be invoked is because the question we are asking, even the naïve starting question we began with, is posed in a context that has taken on board a naturalistic presupposition. A question has a naturalistic presupposition within a context when an answer to it that invokes God or any other supernatural being would rule it out as not a genuine answer. Questions in some contexts do not have naturalistic presuppositions. These questions tend to have religious or spiritual significance. For instance, cosmological questions or questions about the afterlife are often posed in contexts where there is no such presupposition. These questions invite us to think about whether a God was

there at the beginning of the universe or whether a God will be there to greet us upon our death. So, answers to these questions that invoke God as a central part of their explanations are accepted as potentially good answers. On the other hand, there are questions that do have this naturalistic presupposition in most (if not all) contexts. These tend to be questions with practical, or scientific, or mathematical significance. For instance, the questions "Who stole the cookies from the cookie jar?", or "Why is it foggy today?", or even "Why does the square root of 5 not equal 2.5?" all have this naturalistic presupposition. In most contexts, answers to these kinds of questions are expected to avoid any reference to God or any other supernatural entities. If such entities are invoked, we will be right to be unsatisfied with the answer. The context within which I am discussing the question "Are phenomenal properties fundamentally the same or different from physical properties?" is of scientific rather than religious significance, and thus has a naturalistic presupposition.

To be clear, this is not necessarily invariant across time and context. For instance, though I think cosmological questions tend to not have any naturalistic presupposition, scientists, particularly astrophysicists, often ask these kinds of questions in contexts where the naturalistic presupposition is taken seriously. This is how we get scientific kinds of cosmological theories like the Big Bang. It's an open question just how different these kinds of theories end up being to cosmological theories that don't take on the naturalistic presupposition, but this is just a question of how well such naturalistic theories come to respecting the presupposition that they intend to respect. So, by saying that the question at the center of the mind-body problem has a naturalistic presupposition, I don't mean to say that everyone has always treated it this way. Berkeley certainly didn't. But many of idealism's critics clearly have, and even people who currently defend some form of idealism tend to take on this naturalistic presupposition as they see that Berkeley's invocation of God as a central feature in his theory made his answer to the question inadequate. The problem that I've seen is that none of the alternative versions of idealism seem to have an adequate way around this issue.[8] Thus, where I discuss idealism in the rest of this dissertation, I will be following the majority of current Anglo-American philosophy in discussing it as a view to be avoided.

Second, an important theory about the mind known as panpsychism is often treated as offering a solution to the mind-body problem and, as such, as answering the question at the heart of it. Panpsychism

[8] Though see (Farris & Göcke, 2021; J. Foster, 1993; J. A. Foster, 1982; Pelczar, 2019; Ramm, 2021; Tse, 2022) for some who would most likely disagree with this assessment.

is the view that everything has phenomenal properties of experience, even things that we ordinarily don't think are the types of things that have any experiences at all, like rocks or electrons. The main problem with this theory is that it simply doesn't do anything to answer the question of whether phenomenal properties are fundamentally the same or different from physical properties. The proposal is not meant to directly answer the question, but my understanding of such theories is that they are meant to show how the phenomenal properties of humans are not as distinct from everything else in the world as we antecedently think that they are. If there are phenomenal properties everywhere for any object, then I suppose one might think that the puzzle over the differences between phenomenal properties that we experience and their physical correlates does become less grand in a certain sort of way. The specific human phenomenal experiences are not unique among all of the other kinds of properties. However, even if it is correct that things like rocks and electrons have phenomenal properties (however different they may be to ours), this does nothing to help answer the question of whether these properties are fundamentally the same or different from their physical correlates. Take the example of a rock with phenomenal properties of its experience. We can still wonder whether these are fundamentally the same or different from the physical properties of the rock that are studied by geologists and any other rock enthusiasts. We may determine that they are fundamentally different, which would make us dualists, and we would then have to explain how such properties interact within the one world of the rock. So, we would face the same issues as the dualist who isn't a panpsychist. The same goes for a panpsychist who says that the phenomenal properties of the rock are fundamentally the same as its corresponding physical properties. They would need to give an account of the apparent difference. So, by itself, panpsychism offers no genuine answer to the question. It introduces only new puzzles, like what the phenomenal properties of things without subjects could even look like. Nevertheless, panpsychism has had something of a resurgence in Anglo-American philosophical circles in the past few decades. It has been put forward as a serious alternative to the standard options that I have discussed thus far. As such, I will discuss panpsychism more than idealism, but similar to idealism, I will mostly discuss it as a view to be avoided.

3. The Puzzle

Before attempting to answer the question from the previous section, it is important to discuss what a good answer would look like. The basic shape of a good answer can already be determined by looking at

the troubles with the naïve responses to the unrefined question in Section 1. The shape of our more refined question is unchanged from that starting question. As such, a similar set of considerations will shape the contours of a good answer. In this section, I will use the difficulties in Section 1 to develop two central demands on a satisfactory answer to the question. The real puzzle known as the mind-body problem comes from the attempt to answer our refined starting question in a way that satisfies these demands.

When discussing the flatfooted dualist position above, I used an analogy with oil and water to describe the position. The main issue with this analogy was that the situation with dualism is much more robust. Whereas oil and water are unmixable in normal contexts, they are fundamentally the same kinds of thing, made up of electrons and protons and such. Thus, they interact with one another. Dualism says that phenomenal properties are fundamentally different from physical properties. So, the picture we get is one where phenomenal and physical properties exist alongside one another but never interact. However, this goes against both our everyday understanding of the world as well as our best scientific understanding of it. The fundamental difference between phenomenal and physical properties needs to conform to our view of the world as one extent whole rather than two side-by-side but separable wholes that, for all intents and purposes do not interact with one another. This gives us the first demand on a good answer to our starting question, a demand that I will call the Ontological Demand.

Ontological Demand – The world is one unified whole and a satisfactory solution to the mind-body problem should reflect that

To explain any phenomenon in the natural world is to fit it into our understanding of the rest of the world, and being unable to connect things together is tantamount to admitting that you don't understand them, especially if there's only one thing (or type of thing, like phenomenal properties) left out. Imagine a world wherein 1,001 things exist and only those 1,001 things. Now, imagine that 1,000 of those interact with one another while 1 thing cannot interact with the others at all. In wondering how to explain everything in such a world, we could easily explain the 1,000 things by relating them to one another and showing how they all interact in various ways. But in trying to explain the one thing left, we could not do so. That one thing left would be totally unexplained. It would be right to wonder how it even fits in the world at all. Now suppose a philosopher comes along and proposes a solution that says that there are just two types of phenomena, the 1,000 and the 1. The 1,000 can all be related to one another, but the one is fundamentally different. The explanation of how it fits in the world is simply that it is different from the 1,000. But this is

no explanation at all! It's just a description of the problem. And without any explanation of the 1, the "solution" on offer is simply an admission of failure, an admission that there is no explanation for how that 1 thing fits in the world.

Of course, in this fictional world, there seems to be a genuinely unexplainable phenomenon. As I described it, the 1 thing really can't be related to any of the other 1,000 things, so it's right that it goes unexplained. But this fictional world is not our world, and we have no evidence to support the possibility that any of the natural phenomena in our world are entirely separated from the rest. Especially with respect to phenomenal and physical properties, they certainly seem to interact with one another – it seems like pain causes me to flinch and it's certainly the case that a punch to the face causes me pain. Although the question that gives rise to the mind-body problem is not *only* about the causal relation between phenomenal and physical properties, whatever answer we give must be able to place everything into the natural world which means finding some place in this causal network. Even if something is not itself a cause, there should be some way that it relates to those things around it, some way to understand what is going on. In the context of dualism which has special difficulty with this demand, the demand is generally rephrased as a demand to explain how phenomenal and physical properties interact with one another. As will be shown in the next chapter, this difficulty becomes especially acute in the current debate. Modern physics has been able to show how physical properties fit into the causal network and there simply doesn't seem to be any room for phenomenal properties. As such, dualists have had to get quite creative to try to meet this demand.

On the other hand, materialists' appeal to previous scientific and philosophical puzzles to explain apparent differences in such cases as matter and energy or chemicals and living organisms seem to come up short in the case of phenomenal and physical properties. In the case of matter and energy, the apparent difference can be explained away by the very large constant involved in the equation. But attempts at producing such an identity in the case of phenomenal and physical properties have yet to yield any uncontroversial results. And in the case of chemistry and living organisms, there are identifiable functions that living organisms perform and once we had chemical explanations of these, the apparent fundamental difference seems to have dissolved. Again, no such uncontroversial functional analysis seems forthcoming in the case of phenomenal and physical properties. What both of these cases show is that a satisfactory answer to our question must meet a further demand, what I call the Explanatory Demand.

Explanatory Demand – There is an undeniable appearance of fundamental dissimilarity between phenomenal and physical properties; a satisfactory solution to the mind-body problem should be able to explain this

In many ways, meeting this demand just *is* providing a solution to the mind-body problem. The problem is generated by first recognizing an appearance of dissimilarity and then wondering what the explanation of this dissimilarity is. So, failure to meet this demand is a failure to provide a genuine solution. In the context of monism, which has special difficulty with the Explanatory Demand, this generally gets rephrased as a demand to explain how the derivative thing in the world arises out of that which is fundamental (mind or matter depending on the preferred variation of monism). It's not always obvious how things arise even when they arise from the *same* thing, like when a couple produces offspring. The process of reproduction is complex, and its explanation is difficult. And this difficulty is for the production of something of *exactly the same kind*. For new things arising out of different types of things, it gets more difficult. In the mind-body problem, the appearance of not just differences but a *fundamental difference* between phenomenal and physical properties makes this an incredibly difficult demand to meet.

As Jaegwon Kim states in his book *Mind in a Physical World*, "If a whole system of phenomena that are prima facie not among the basic physical phenomena resists physical explanation, and especially if we don't even know where or how to begin, it would be time to reexamine one's physicalist commitments" (1998, p. 96). It was ultimately on the basis of not seeing a way for materialism to meet this demand that he abandoned the view in favor of a form of dualism. In essence, he abandoned a view that he had come to believe could not provide an answer at all in favor of a view that at least seems to provide *an* answer even if it is not fully satisfactory. His ultimate dissatisfaction with dualism is demonstrated by the fact that he insisted on calling his position "near enough" to physicalism in his 2005 *Physicalism, or something near enough*. Physicalism, he thinks, would provide a satisfactory answer if it could only provide an answer at all. The trouble, though, is that his "physicalism near enough" is a perfect example of a solution that meets the Explanatory Demand without an answer for the Ontological Demand, which reveals that this "solution" is also not merely an unsatisfactory solution but rather an admission of *failure to find any solution at all*. This is a perfect demonstration of the tension between these two demands. Meeting one makes it difficult to see how the other can be met. Thus, the mind-body problem just is the problem of providing an answer

to the question of whether phenomenal and physical properties are fundamentally the same or different while meeting both the Ontological and the Explanatory Demands.

The current debate over the mind-body problem is largely centered around two arguments. These arguments are the *Causal Argument* against dualism and the *Epistemic Argument* against materialism.[9] Both of these arguments are quite strong; it is only when folks have turned to defending their preferred view against their opponents' argument that they have failed to give satisfactory answers. If, as the debate has largely been conducted, there are only two open possibilities, then merely providing a strong negative argument might be enough to show that the only possibility remaining must be correct. However, the waters get substantially muddier when both sides have strong negative arguments. I'll discuss these arguments in Chapter 2, focusing on how they create the current impasse that makes progress on the mind-body problem so difficult. As will become obvious, the central feature of each argument is showing in great detail how the opposition fails to meet one of the two demands discussed above.

After demonstrating the forces that create the impasse in the current debate, I will attempt to show a way through this impasse. The view that I intend to develop has been attempted before and is called *neutral monism*. From the "monism" in the name, it is obviously intended to be opposed to dualism. From the "neutral" in the name, it is also obviously intended to be different from both idealism and materialism as the fundamentally existent kind of property is intended to be neither phenomenal nor physical. The general strategy, then, must differ from either the materialist or idealist strategies discussed in the first section of this chapter. Rather than saying that one of either phenomenal or physical properties is real while the other is derivative, the idea is that somehow *both* the phenomenal *and* the physical are derivative. Thus, they share some deep connection and are fundamentally the same, but neither can be said to be the more "real". Though attempted in the past, it doesn't appear as if any formulations have been made to work thus far.

As we'll see in Chapter 3, much of this has to do with the fact that most formulations end up being merely variations on idealism, materialism, or dualism. That is, it is difficult to see how the physical and

[9] You might wonder about issues concerning parsimony here. I will discuss parsimony in the next chapter after discussing the Causal and Epistemic Arguments. As I'll argue there, any reasonable parsimony argument must presuppose that monism has already met these demands, particularly the Explanatory Demand. If it doesn't adequately explain the basic phenomenon under consideration, it doesn't matter how parsimonious or elegant a theory is. As such, arguments that primarily rely on parsimony will still have to answer the Epistemic Argument in an adequate way. But more on this in the next chapter.

phenomenal can *both* be derivative, and so previous formulations end up sneaking one or the other or both into the fundamental fabric of reality. The trick then, will be to show how both the phenomenal and the physical can be derivative without sneaking either of them into the fundamental properties. In Chapter 3, I'll discuss more thoroughly the failure of the main formulation of neutral monism, called *Rusellian monism*, with a focus on narrowing down possible ways to formulate neutral monism that avoid this problem.

Then in Chapter 4, I'll narrow the field down further to just one possible approach, which I'll call *epistemic neutral monism*. Since this approach has some similarities to *a posteriori* materialism and the powerful qualities view in the dispositions debate, I'll discuss these with a view toward explaining how to develop an epistemic neutral monism. It will turn out that the standard formulations also have not been made to work, and so I suggest one final possibility that relies on a conventionalist metaphysics about the mental and the physical. The view I seek to develop I call *conventionalist neutral monism* in order to distinguish it from the standard formulations of epistemic neutral monism.

Chapter 5 develops this conventionalist neutral monism by wading into the metaphysical weeds that need to be untangled to make the view coherent as well as show how this view does not fall back into being a variation of one of the more standard three solutions to the mind-body problem. Finally, in Chapter 6, I'll discuss some loose ends relating to how this view meets the two demands presented in this chapter and clarify how the view answers the question at the heart of the problem.

Chapter 2: The Impasse

> "I am supposed to explain to you my doubts? By laying stress on these it will appear that I want to pick holes in you everywhere. But things are not so bad, because I do not feel comfortable and at home in any of the "isms." It always seems to me as though such an ism were strong only so long as it nourishes itself on the weakness of its counter-ism; but if the latter is struck dead, and it is alone on an open field, then it also turns out to be unsteady on its feet. *So, away we go!*"
>
> - Albert Einstein, *The Collected Papers of Albert Einstein*, Vol. 8, Doc. 624 (1998)

1. The Causal Argument

In the current debate, philosophers have been struggling to resolve an impasse created by very strong arguments against materialism and dualism. Here, I'll review both of these arguments and show how philosophers have tried to respond to them. There are several responses which lead to variations of dualism and monism. As I'll show in this chapter, these responses are coherent, but they nonetheless fail to pave the way for much of a satisfactory answer to the mind-body problem. The result is that philosophers can recite responses to the arguments that they know will yield a coherent view but that fail to satisfy those who aren't predisposed toward the view. The issue seems to be that while each response can handle the explicitly stated premise that it is direct at, they all fail to deal with the underlying concern about the relevant demand – the Ontological Demand for dualism and the Explanatory Demand for materialism.

Given that the Ontological Demand and Explanatory Demand are faced only by one of either dualism or monism, a suspicious type might think that they are merely the result of the opposing side trying to stack the deck in their favor. If you are such a suspicious type, you might think that you don't have to meet these demands. Rather, you can explain why the demand simply doesn't apply to you. As we'll see below, this is a route that many have tried. And although the better of these attempts do provide valid arguments and coherent positions, the resulting "solution" to the mind-body problem is always a little empty. Either the same issues with the demand are regenerated in a slightly different guise or one simply feels that the "solution" has somehow missed the point of the whole endeavor, thus showing that the demands really must be met in order to provide a satisfactory solution. The point of the chapter, therefore, is not to defend any particular argument or any response to these arguments. Instead, the point is to demonstrate what the standard responses are and to show how one might be led by a dissatisfaction with these to look for a different kind of answer altogether. Nevertheless, a satisfactory answer needs to do more than respond to relevant criticisms that might be leveled at it. It must also meet both of the demands

mentioned in the previous chapter. I'll begin with the Causal Argument against dualism before moving on to the Epistemic Argument against materialism.[10]

Here is the Causal Argument in its canonical form:

1) All physical states have a sufficient (micro)physical cause.
2) If all physical states have a sufficient (micro)physical cause, then phenomenal properties have no causal effect on the physical.
3) If phenomenal properties have no causal effect on the physical, then phenomenal properties either are not real or depend on physical properties for their existence, and dualism is false.
4) Dualism is false.

The first premise is largely uncontroversial since it basically states our commonplace belief that every physical event was brought about by a physical cause. This is true even if indeterministic physics is true. Of course, this only applies in theory since practically speaking scientists will be limited in their capacity to trace these causes. Nonetheless, if some physical state exists, then *something* was sufficient to bring it about. And the premise states *that* thing was also physical. The main justification for believing that the cause is always physical is inductive – all discovered causes have turned out to be physical – and so premise one is not definitively proven. For instance, the source of life has turned out to be various chemical configurations, many mental functions, like responses to stimuli or reporting how one is feeling, have turned out to be the result of physical properties in the brain, and even the planets which were once thought to be divine and distinct from physical objects on Earth have turned out to obey the same physical laws as things here.

However, if this premise turns out to be false or even if we merely consider giving it up, it has severe consequences. Moreover, rejecting it seems unmotivated. Whatever we think about the completeness of physical causes, it doesn't seem that motions we make with our bodies (or other kinds of physical events

[10] Both of these arguments are indebted to Chalmers's "Consciousness and its Place in Nature" (2002). However, while I agree with his general analysis of each premise and much of the resulting views, his analysis there doesn't recognize any sort of underlying problem at the root of each position's failure to provide an adequate solution to the mind-body problem. His goal there seems to be categorizing the various views in the debate. Here, I intend to not only present the variations of dualism and materialism by rehearsing the arguments that have produced them but also show that their problems have an underlying source – a failure to meet the demands presented in Chapter 1. After developing my own solution to the problem, I'll return to these arguments in Chapter 6 in order to explain how they are sound but nonetheless do not imply the problems that they are standardly assumed to imply.

associated with our experiences of the world) are good candidates for physical events without physical causes. Indeed, we have at least a rough idea of how the brain and nervous system – physical things – cause almost all of our bodily actions. So, we have little reason to doubt the first premise in cases that matter to the mind-body problem.

Nonetheless, some dualists[11] have attempted to undermine the Causal Argument by arguing that this premise is false. Denying this premise requires arguing that some physical effects have (at least partly) *phenomenal* causes that can't be explained by the associated physical events. This view is called *interactionist dualism*.[12] Historically, it is the most common type of dualism. However, as I stated above, this is generally taken to go against modern physics. It should nonetheless be admitted that there is room in current physical theories for new natural forces to be discovered; something akin to magnetic fields could be discovered, say "phenomenal fields". There are also those that argue that the existence of phenomenal causes is compatible with some interpretations of collapse theories of quantum mechanics.[13]

Physical → Phenomenal → Physical
Interactionist Dualism

I have to admit that I am not qualified to judge the relevant physical theories and so I cannot say definitely whether or not this compatibility is true or whether there is space for "phenomenal fields" to exist side by side with physical fields. Nonetheless, even if we do accept that these arguments are right, this is not evidence *for* the existence of such causes. Moreover, cognitive science gives no direct evidence for the existence of such causes. Perhaps more importantly, it's not clear that any of these interpretations of quantum mechanics *require* the existence of phenomenal causes rather than merely not ruling their existence out. This is not nearly enough to establish good reason to deny the truth of premise one, especially since there are other interpretations that *do* rule out the existence of phenomenal causes.

You might be tempted to respond to this situation with the pithy aphorism that "absence of evidence is not evidence of absence", but I'd caution against applying this phrase in the current context. It is true that

[11] See (Averill & Keating, 1981; Cucu, 2020; Cucu & Pitts, 2019; Lycan, 2009; Okon & Sebastián, 2020; Pitts, 2019) for recent discussion of interactionist dualism and its prospects.

[12] This corresponds to Type-D Dualism on Chalmers's classification scheme.

[13] There is an interesting question about whether this should really count as denying the truth of premise 1. After all, this kind of response would still end up saying that physical theories tell us all there is to know about the world. The problem is in the way this response says that physics can explain phenomenal properties. It says that physics will be able to explain phenomenal properties by treating them on par with other physical properties. So, the answer ends up saying that phenomenal and physical properties are fundamentally different, but they can be incorporated into one theory. This sounds more like a kind of interactionist dualism to me rather than a kind of physicalism.

absence of evidence is not evidence of absence, *sometimes*. But sometimes absence of evidence *is* evidence of absence. For instance, if you look around the room for something that is invisible to the naked eye, i.e. a virus, then the absence of any visual evidence does not provide any evidence that it is not there. There or not, you simply can't see it. However, if you look around the room for the keys that you remember leaving on the counter and you do not find them, then the absence of evidence is very good evidence that your keys are absent from the room, or at least all the places in the room that you looked. In the first scenario, we are using the wrong instrument to gather evidence – our eyes rather than a microscope – but in the second instance, we are using the correct instrument and still find no evidence. Thus, the absence of evidence is evidence of absence.

The situation is the same for phenomenal causes of physical states. We have looked for non-physical causes of physical states of affairs and every time turned up empty. It certainly could be the case that we simply haven't been using the right instruments, but this is highly unlikely. We've used the best instruments suggested by any relevant hypothesis and have come up empty of phenomenal causes every time. On the contrary, psychologists and neuroscientists always find that the causes of our bodily actions are states of the brain and central nervous system – physical causes – even when a phenomenal state is also in the vicinity. For instance, although it is quite natural to say that the pain in my arm caused me to pull it back, the scientific identification of the cause is in the brain and central nervous system, not the pain itself. It is, of course, possible that we have yet to use the correct apparatus despite all of our trying, but then it is on the dualist to explain what kind of instrument would give us evidence of non-physical causes where all of our other experiments have failed, and then to use the instrument to gather the appropriate evidence.

One quick attempt at a response might be to say that we are simply using the wrong *kind* of apparatus. All of the instruments that we have been using are *physical*, so of course we have found only physical causes. Phenomenal causes will be "invisible" to these instruments. But this is self-defeating. If *no* physical instrument can interact with phenomenal causes, then there can be no phenomenal effects on the physical, and therefore, physical states can only have physical causes. I suppose there is the possibility of non-physical causes like the supernatural or of spontaneous physical states that have no cause, so this line of reasoning doesn't *prove* the truth of the first premise. But it does prove what is crucial to the current discussion, that physical instruments can either find phenomenal causes if they exist or they cannot. If the first disjunct is true, then absence of evidence *is* evidence of absence, thus showing that we have strong

inductive support for the truth of premise one. If the second disjunct is true, then there can be no phenomenal causes of physical states, thus showing not only that the first premise is true but also that the second premise is true.

The situation has something of the air of a conspiracy theory. The inductive evidence for the truth of premise one is quite strong while the evidence against it is flimsy at best. But since the falsity of the premise is compatible with some theory or another, some people seem willing to ignore the fact that all of the positive evidence points in the opposite direction. Of course, if new evidence becomes available that weighs in on this issue, the situation might change, but that is no different from any scientific inquiry. So, we have no reason *not* to accept the truth of premise one and many reasons to accept that it *is* true. Even the most open-minded person would be best served by accepting the truth of premise one until some positive evidence calls it into question.

Although I think the above is a perfectly strong justification for provisionally accepting premise one and is the justification usually given for its acceptance[14], I think we can also say something even stronger. Anything that can be demonstrated to have a causal effect on the physical will *necessarily* be physical. To understand this claim, we have to remember the distinction between phenomenal and physical properties discussed in Chapter 1. There, I argued that the appropriate distinction is between the phenomenal as subjective properties of experience – where subjectivity is most properly understood as individually confirmable – and the physical as objective – where objectivity is best understood as intersubjectively confirmable. In order to prove that an intersubjectively available event is the effect of some other event, the cause must also be intersubjectively confirmable, and therefore the cause of any physical event must itself be physical on this understanding. The reason for this is that proving that an intersubjectively confirmable event is caused by something requires being able to show that this is the cause to other individuals. But being able to show something to other individuals just is intersubjective confirmability. Thus, any purported cause of a physical event will come to be understood as physical precisely at the time when we come to see that it is, indeed, a cause of the physical event. If this is right, then any amount of empirical research will only reveal more fine-grained *physical* causes of physical events because this is the only type of cause that can be demonstrated as the cause of a physical event. This fits well with the history of science, too. Things

[14] For example, see (Chalmers, 2002).

have come under the umbrella of the physical precisely at the time when we come to understand how they are causally efficacious of things that we already consider to be physical – i.e. of things that are already intersubjectively confirmable. To be clear, this does not mean that *all* events are necessarily physical. This is, after all, only one premise of an argument that is meant to show that dualism is false. It merely means that all physical events can only be demonstrated to have physical causes. If there are non-physical causes of such events, then we could not demonstrate what they are. The possibility remains of a variety of other types of events that are *not* physical and can therefore have any other type of cause. For instance, we have not ruled out the possibility of phenomenal events being caused by phenomenal causes or any other type of cause. Indeed, this is one type of response that the dualist might offer; however, it is more appropriately a response to premise three, and so I discuss it there.

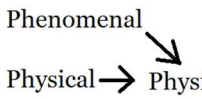
The second premise states that since physical states already have a sufficient physical cause, they do not *also* have an additional phenomenal cause. The justification for the second premise is simply that effects are not overdetermined in a systematic or regular way. Of course, this does allow for ordinary everyday "overdetermination" that says an event of a certain description might have two sufficient causes. For instance, it is not problematic to say that we won the game because Amy hit a homerun and because Charlie scored on an error from third base. It may be that either of these would be sufficient for us winning the game if we won the game 2-0 and thus our winning might be "overdetermined". But this is unproblematic for the purposes of the argument for one of two reasons. First, we might say that this is overdetermination, but it is not regular or systematic. There is no systematic overdetermination of winning games such that whenever there is one sufficient cause there is always another. In order to prove premise 2 false, there would have to be such systematic overdetermination.

Second, we might simply say that the state of affairs is underspecified. If we make the state of affairs more specific, then there will only be one sufficient cause. It is *not* overdetermined that we won the game 2-0, both Amy's run *and* Charlie's run are required for a sufficient cause of the score. Of course, merely stating their runs is not yet a full specification of the sufficient cause; we also have to mention the pitching and defense that went into the other team scoring no runs, all the outs that led to us *only* scoring two runs, and perhaps things like the weather and fans in order to have a fully specified sufficient cause. When all of these relevant details are specified and the score is sufficiently specified, then we'll have a full specification

of the sufficient cause of our winning 2-0.[15] Once this is known, then it is pretty clear that such ordinary occurrences only provide more evidence for the truth of the second premise.

The premise is also not intended to rule out instances like the following: the baseball broke the window and also the particles arranged baseball-wise broke the window. There are some legitimate metaphysical disputes about whether the baseball is identical to the particles arranged baseball-wise. However, in the context of the argument, if the case were to be given a phenomenal-physical corollary, then it would be obvious that the phenomenal state would be composed of the physical states. So, even if such overdetermination is systematic and regular, this line of response would not be any help to the dualist.

You could argue that there are *two* sufficient causes, one physical and one phenomenal, which would mean that by taking one away, the effect would still occur. This is a relatively rare view. So, I've never encountered an explicit name for it, but one might call this *overdeterministic dualism*.[16] The main problem with this line of response is again inductive. Overdetermination in every other area of science or the world comes in one of the two kinds mentioned above. It is either not systematic or it is due to some kind of composition. If overdeterministic dualism is plausible, it must be due to the kind of overdetermination that is regular and systematic. But this kind of overdetermination is always explained as either the result of an identity (the baseball is identical to the particles arranged baseball-wise) or it is explained by some kind of composition (the particles arranged baseball-wise compose the baseball). But composition is exactly the kind of necessary ontological dependence relation that dualism thinks does not obtain between the physical and the phenomenal.

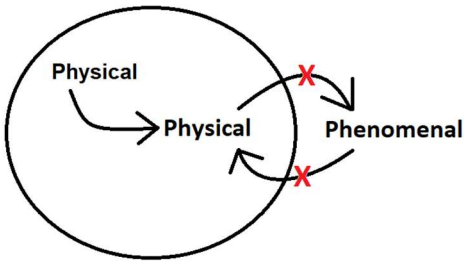


Overdeterminationist Dualism - each arrow represents a sufficient cause

[15] In (Lewis, 1986), David Lewis suggests that we shouldn't go too fine-grained in this way to address the problem of overdetermination. He argues that this would give rise to spurious cases where, for instance, one person *not* firing in a 22-person firing line would count as a cause of the victim's death because if he *had* fired, then the death would have been different death. But this is a case where we have illicitly assumed that part of the effect is more important than another. If we are a family member of the victim, we won't care that the death was one of 21 bullets rather than 22. The effect that we care about is the death, so the one effect shouldn't be treated as different from the other. On the other hand, if we are the firing line commander, we would absolutely care that it is a 21 bullet death rather than the intended 22 bullet death. The fact that one person didn't fire *does* make the one effect different from the other, not in that it is a different death, but in that it indicates that one person is either unwilling or incapable of doing their job. So, going more fine-grained is only problematic if we assume that one part of the effect is the only part of interest, but this changes based on one's perspective.

[16] Chalmers mentions this view as Type-O Dualism but only briefly discusses it as an underexplored view. I think there are good reasons for this as I discuss in the main text. But see (Mills, 1996) for a defense of overdetermination in this context.

The third premise states that since phenomenal effects are not causes of physical states, they either have no cause (and aren't real) or depend on physical properties for their existence. Thus, if phenomenal properties exist, they are really (or fundamentally) physical despite any appearance to the contrary. Combined with the first and second premise, this means that the physical world is causally closed. That is, if you were to map out the states of the universe linked by arrows with their causes and effects, and then were to draw a circle around all and only the physical states, there would be no causal arrows going into or out of the circle. The third premise is based on two related beliefs. First, there are no causal dead-ends; rather, all states have some causal effect on the world. This is an empirical version of the metaphysical claim "Alexander's Dictum", to exist is to have causal powers. Second, causes are of the same category as their effects – in other words, any type of state that is the cause of a different type of state might in turn be caused by that second type of state if the circumstances were suitably different. This second belief entails the first, so we might split the third premise into a stronger and weaker version if we were so inclined. I am not so inclined, so I'll leave that to you if you wish. In either event, both beliefs are true and so the third premise is justified.

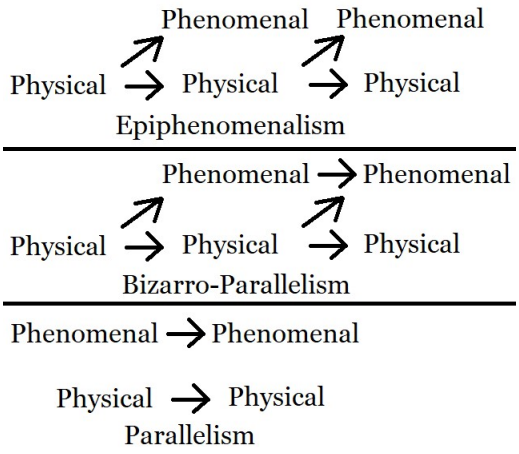


The Causal Closure of the Physical - The arrow from the physical to itself represents premise one. The lower X represents premise two. The upper X represents premise three.

William James makes a strong argument in support of the first belief when he states, "One may well refuse, until absolutely overpowered by the evidence, to believe that the world contains items which in no wise influence their neighbors; whose existence or non-existence need, so far as the remainder go, be taken into no account." (1879, p. 3). To continue in this line of thought, one might well wonder what evidence *could* exist if such items do not influence their neighbors in any way. Without *any* effects, there doesn't seem any possible way to gather any evidence of their existence at all. The second belief is, in many ways, simply a restatement of the Ontological Demand. We live in one unified world and so the causes and effects in the world must reflect that by being such that they can influence one another in suitable circumstances. To take up an example from earlier, it is not just that water keeps oil afloat, but that if the oil were to be under the water, it would displace it as it rises to float on top of the water. In any given instance, the causal direction goes only one way, but in a different situation, the causal direction can reverse. Thus, not only is there no evidence for the existence of phenomenal states that have no effects, it is also the

case that they couldn't exist if they couldn't have any effects on their physical causes given suitable circumstances.

There are two ways to respond to this. You might reject the first or the second belief that justifies the third premise. If you reject the first belief – that there are no causal dead-ends – then physical states might cause phenomenal states, but those states in turn cause nothing. A quicker but more jargon-filled way of saying this is that phenomenal properties are epiphenomenal, thus giving rise to the name of the view, *epiphenomenalism*.[17] If you reject the second belief – that causes are of the same category as their effects – then physical states might cause phenomenal states, but those phenomenal states then only cause other phenomenal states. I'm not sure that I've encountered any philosopher explicitly defending such a view. It has similarities with epiphenomenalism but is fundamentally a different view. I am tempted to call it *parallelism*, but that also refers to a different view, traditionally contrasted with interactionist dualism, where the phenomenal and physical have *no* interaction at all. Thus, the two run parallel to one another, never crossing paths. On the view that rejects the second justification for premise two, there is interaction but only in one direction like epiphenomenalism but then the phenomenal properties can affect other phenomenal properties, like parallelism. It's a bizarre view, so if it needs a name, I suppose *bizarro-parallelism* will do.



The problem with all of these views is best stated by David Chalmers when he says "the oddness of epiphenomenalism is exacerbated by the fact that the relationship between consciousness and reports about consciousness seems to be something of a lucky coincidence on the epiphenomenalist view" (2002, p. 264). Although he was speaking about just one of them, this problem applies to all three views mentioned in the previous paragraph. The issue is that someone can tell you about all sorts of phenomenal states that they're experiencing and each of them will be a true report. It certainly seems that these types of reports are caused by the corresponding phenomenal properties – it is the pain I experience that makes me say "Stop! That hurts" – but epiphenomenalism, parallelism, and bizarro-parallelism all say that's not the case. Instead, the

[17] This corresponds to Type-E Dualism on Chalmers's classification.

report of pain and the experience of pain are two independent events. In epiphenomenalism and bizarro-parallelism, it is true that the physical and phenomenal states have a common cause. But even so, it very well could have been that reports of phenomenal states might not have matched up with the right phenomenal experiences. Thus, it's merely a lucky coincidence that the pain I feel comes along with me saying that it hurts rather than me saying that, for instance, the ham is in the oven. This, of course, is a coherent view, but as I said above about epiphenomenalism, all these views lack any evidence in support of them and it seems impossible to go out and gather such evidence. Even if you could get such evidence, they are a deeply unsatisfactory view of the world. The reason is simply that they fail to meet the Ontological Demand. None of them explain how the mind and matter fit together in this world. On the contrary, they state that the phenomenal and physical don't really fit together at all. It's just an unfortunate coincidence that they match up with one another because it gives the appearance that they should fit together.

All of this discussion of the Causal Argument details the significant problems that dualism runs into. Without meeting the Ontological Demand, dualism is incapable of coming up with a satisfactory response to the Causal Argument and so one must conclude that dualism is false. At this point, we are left wondering what to do with these phenomenal experiences of the world. After all, the argument only shows the troubles for dualism, it does nothing yet to establish a solution to the mind-body problem. One potential response to this type of argument is to say that causation isn't the only way in which two events or properties might be related – thus showing how the Ontological Demand can be met by non-causal relations. Indeed, this is the tactic that materialism follows. Materialists attempt to show that phenomenal properties and physical properties are related by showing how the phenomenal properties non-causally depend on the physical properties and thereby show that phenomenal properties are either illusory or necessarily derived from physical properties. The argument to which we turn next, gives strong reason against this possibility.

2. The Epistemic Argument

The canonical form of the Epistemic Argument[18] is this:

1) All ontological dependence relations have a corresponding entailment relation.

[18] This canonical form of the argument differs from that presented by Chalmers (2002, p. 250). Although I don't think that I disagree with the gist of his analysis, the way that I argue differs, in some places substantially. Most obviously is the inclusion of premise one, which is something of an implicit premise in Chalmers' presentation and comes out in his discussion of Type-B materialism. I make it explicit in order to draw out some important metaphysical issues that are raised in the debate. Chalmers discusses these more explicitly in other places, especially when discussing strong necessities. See (Chalmers, 2014) for a relatively brief and explicit discussion of this premise.

2) Anything with an entailment relation can be proven on the basis of either *a priori* evidence or *a posteriori* evidence.
3) There is no *a priori* evidence that could prove an ontological dependence relation between the phenomenal and the physical.
4) There is no *a posteriori* evidence that could prove an ontological dependence relation between the phenomenal and the physical.
5) Therefore, there is no ontological dependence relation between the phenomenal and the physical, and so materialism is false.

The first premise is basically a different statement of the Explanatory Demand from Chapter 1. Any purported dependence between two things requires there to be an available explanation even if we do not have it, especially if the two things appear to be different. Materialism is the claim that the existence of phenomenal properties depend on physical properties. The idea of a dependence relation is meant to encompass several metaphysical relations – identity, reduction, supervenience, grounding etc. These are all relations involving metaphysical necessity.

The basic justification behind premise one is that an ontological dependence relation is no mere coincidence. In fact, when one thing necessarily goes along with another, it is the very opposite of a coincidence. But it is only coincidental relations that require no explanation and therefore no proof of their existence. Or rather, coincidental relations have no explanation, and the only proof of their existence is that they exist. In a real sense, a "coincidental relation" is no relation at all. It is merely the instantiation of two properties in similar enough times at similar enough places. For instance, while it is true that I was born on the same day as many other people, these are all merely coincidences (given that I have no twins). One might say that I bear some relation to all of these people, and certainly I am of the same age as they are, but this relation is nothing like an ontological dependence relation. None of our births depend in any way on the birth of the other, nor do they both depend on some third thing. There is no entailment between the fact that I was born on a certain day and someone else was born on the same day. There is only the fact that I *was* born on the same day. But this is not evidence of the fact, it is the fact itself.

On the other hand, even contingent relations require an entailment relation in order to exist. For instance, if I had a twin, it would be a contingent fact that I was born on the same day as a sibling. But me being a twin would be entailed by the fact that I was born on the same day as a sibling, even if as a matter

of scientific investigation, this fact might be unavailable to us. Of course, given that this would be a contingent relation, the proof that I had a twin would all depend on certain historical and biological facts remaining the same, but there would need to some entailment relation of any such contingent relation. For necessary relations, this entailment relation is all the more required. For example, the fact that Clark Kent was born on the same day as Superman is entailed by the fact that for any two things that are identical, they will have all and only the same properties. As such, given that Clark Kent has the property of having been born on such-and-such a day, Superman must also have that property. The fact that Clark Kent and Superman are identical explains the necessary relation between their respective birthdays.

It is precisely this type of explanation that is required to meet the Explanatory Demand. Since materialism claims that the relation between physical properties and phenomenal properties is one of ontological dependence, any materialism that relies on the denial of premise one will only be able to provide the barest veneer of a response to the Explanatory Demand. When asked to explain the relationship between the phenomenal and physical, such a materialism will respond that the phenomenal depends in some necessary way on the physical, but when asked to explain this relation, they will say that it is a brute necessity that has no explanation. This is, at best, a weak way of meeting the Explanatory Demand and, at worst, an expression that a solution to the mind-body problem need not explain the relationship between the mental and the physical – which is to say that a solution to the mind-body problem need not solve the problem at all!

Despite this obvious drawback to the position, it is worth exploring, as it leads to a metaphysical sticking point between dualism and materialism. To deny this premise, materialists have to give reasons for supposing that there are primitive dependence relations, or what Chalmers (1996, 2010, 2014) calls *strong necessities*. The key is to say that necessities are not (always) intelligible because the concepts we use to refer to the objects somehow obscure what the referent is. Materialists have two options here; they can say that primitive dependence relations are a common occurrence that happens with many, if not all, of our concepts, or they might say that there is something peculiar about the concepts used to refer to phenomenal properties.[19] Thus, although the premise would be true in general, the fact that we refer to this one property

[19] See (Block & Stalnaker, 1999; Papineau, 1993) for instances of the general strategy and (Levin, 2007; Loar, 1990, 1997; McLaughlin, 2001; Perry, 2001) for instances of the particular strategy, also called the *phenomenal concepts strategy*.

in two different ways makes it so that the dependence of the phenomenal on the physical properties is not intelligible to us.

The idea can be illustrated with an analogy. Think of a museum that has dim lighting to preserve all the artwork. If you were to view the art in the museum, you could definitely tell whether this painting is the same as that one. But now suppose that very painting was stolen by a colleague of yours and hung in their very well-lit office. Given the change in lighting, it might be very difficult to see whether the painting in your colleague's office is the same as the painting that you saw in the museum several months ago. Obviously, the painting is identical to itself, and so the properties it has when hanging in your office ontologically depend on the properties it has when at the museum. Nonetheless, the change in viewing conditions makes it impossible to know its identity upon looking at it. Analogously, the denial of premise one is saying that our "viewing conditions" (i.e. the concepts we use) of phenomenal properties obscure our ability to understand that phenomenal properties depend on physical properties. The general strategy claims that this is somewhat ubiquitous while the particular strategy claims that this is peculiar to phenomenal concepts. Strangely enough, the materialist view that denies premise one is often referred to as *a posteriori* materialism.[20] But this is a mistaken labeling. As will be discussed below, *a posteriori* materialism is a label better reserved for the view that the relation between phenomenal and physical properties can be understood by *a posteriori* investigation of the world, but the view that denies premise one claims that the relation is not, and need not be, intelligible *at all*. It claims that the dependence between the physical and phenomenal is a strong necessity. Thus, I prefer the label *strong necessitarian materialism* for this view.

The problem in common with both the general and the particular strategy of developing strong necessitarian materialism is that if phenomenal properties truly depend on the physical properties, then we should be able to perform some manipulation to discover this fact. By the materialist hypothesis, any change

[20] Or, in David Chalmers's taxonomy, Type-B materialism. As I argue in the main text, this type of materialism is really two drastically different views that have been confusingly lumped together under one heading. However, it is very rarely the case that you'll find philosophers attempt to deny premise one independently of an attempt to deny premise four. Obviously, you can't deny both at once for the relevant properties, but usually philosophers begin by attempting to deny premise four and are pushed by problems with their responses to deny premise one instead. Thus, I think the mistake in labeling has come about as the philosophers who deny premise one start off trying to deny premise four, which yields an *a posteriori* materialism, and only move to denying premise one after being forced to abandon their attempts to refute premise four. The label of *a posteriori* materialism has stuck with the view even though it is clearly a relevantly different position.

in a phenomenal property must be the result of a change in the corresponding physical property. So, we should be able to perform some manipulation of phenomenal properties and their "viewing conditions" such that we can discover whether there is any corresponding change in some physical properties. In the painting analogy, we can dim the lights in the office, move the painting back to its room in the museum, or otherwise manipulate the viewing conditions such that we can see that it is the same painting (suppose, for the sake of the example, that there is no forgery or copy of the painting). If there is truly only one painting, then it must be true that we can manipulate the conditions such that this becomes known. And if the one property type truly depends on the other property type, then even barring some problems with our concepts, we should be able to manipulate those concepts or the world such we can see that manipulation of the physical properties necessarily leads to a corollary change of the phenomenal properties. This is impossible to deny without denying the very basis of the materialist thesis, or at least without denying that the materialist thesis can be proven. There are some deep metaphysical issues that plague any materialism that attempts to deny premise one. However, this is not the end of the discussion on this point. As we'll see, it becomes of central relevance to the neutral monist view that I intend to explore. As such, more detailed discussion of this tricky issue will have to be postponed until later, specifically until Chapters 4 and 5.

Premise two is simply a statement of two ways that we might know or understand something. On the one hand, we may understand things by theoretical or semantic deduction from knowledge that we already possess. This is called *a priori* knowledge because it is knowledge that we are able to possess prior to needing any further empirical knowledge or experience of the world. On the other hand, our knowledge may require further empirical study of the rest of the world. This is called *a posteriori* knowledge because it is knowledge that we can gain only after some experience in the world. The second premise is nothing controversial, and to my knowledge there are no attempts to refute this argument by denying premise two. It is merely a statement that all knowledge is either gained without experience of the world, through *a priori* deduction or through experiencing the world by *a posteriori* discovery. Thus, the dualist attempts to disprove the purported metaphysical relation of materialism by epistemic means. This shouldn't be entirely surprising given the discussion in Chapter 1 of materialism's difficulty in meeting the Explanatory Demand.

The third premise states that there is no *a priori* evidence that could prove an ontological dependence relation between the phenomenal and physical. Put another way, this premise states that there exists an epistemic gap between the phenomenal truths and the physical truths. This means that there is a

difference in the way that we know, explain, or conceive physical and phenomenal truths such that phenomenal truths cannot be *a priori* deduced from the entire set of physical truths. In many ways this is simply a statement of the starting point of the mind-body problem – that the phenomenal and physical seem to be fundamentally different. However, although that starting point is largely uncontroversial, the statement of the premise goes further in that it states that the apparent fundamental difference between phenomenal and physical truths is a real epistemic difference, not merely an illusion of an epistemic difference brought about by our not being fully aware of the relationships between our concepts. There are three ways to justify this premise, stemming from three epistemic notions – knowledge, explanation, and conceivability.

The approach to the premise via knowledge can be illustrated by the famous thought experiment of the color scientist Mary to show that there is no *a priori* evidence that could prove an ontological dependence relation between the phenomenal and physical.[21] Briefly, the thought experiment goes like this. Imagine a scientist Mary who can only see in black and white. The mechanism for this is not that important, but the point is that she has never had the opportunity to see color. Growing up surrounded by people who can see color, she becomes interested in the subject and through great academic effort she becomes an expert in the field, making many breakthroughs in the study of color. It turns out that she is a brilliant person who can learn an immense amount and retain all this knowledge. Through her studies, she ends up learning *all* of the physical truths pertaining to color – truths about reflectance properties of materials, the correspondence of wavelengths of light to specific colors, optics, the eye, how the brain interprets color, everything. Then, the day comes that a doctor has developed a procedure to fix her colorblindness; she will finally be able to see the colors that she knows all physical truths about. When she opens her eyes after the procedure, she sees colors for the first time and learns something new. She learns what the colors look like; something that she could not have known even if she knew everything else about colors. Moreover, she learns that all of her physical knowledge of colors had left something out all along; she had never known not only what it was like for *her* to see colors but what it was like for *anyone* to see colors. Thus, the story goes, there must something different about our knowledge of physical truths and our knowledge of

[21] See (Jackson, 1982) for Mary's introduction into the literature. My presentation of it differs slightly to avoid fruitless digressions. This paper also gives a thorough defense of the knowledge argument against materialism. See also (Broad, 1925; Feigl, 1958) for precursors to the Mary thought experiment that have similar structure.

phenomenal truths; there is an epistemic gap such that if there is some ontological dependence between phenomenal and physical properties, it is not *a priori* deducible from the total set of physical facts.

The notion of explanation goes by a similar argument; we can even present it in light of the same thought experiment.[22] Suppose Mary, before her procedure, is asked to write a book to explain to us everything she knows about color. Her book would detail all of the structures of the eye, the brain, and various objects involved in color. Her book would also explain all of the various functions that color serves in the lives of animals and plants – finding mates and prey, avoiding predators, finding ripe fruit, attracting pollinators, etc. However, her book (which would presumably involve many volumes) would not explain everything. It would fail to explain what it is like to experience seeing color. It would have the explanation for the causes of color experiences but not why such causes give rise to this or that experience. The experience of seeing color is neither a structure nor a function, and since physical explanations can only involve structures and functions, they necessarily leave out phenomenal properties. Again, there is an epistemic gap between phenomenal and physical truths.

The final epistemic notion of conceivability builds on the previous two and takes the idea to its logical extreme. Noticing that our knowledge and explanations of physical and phenomenal truths can differ opens the idea that we can conceive of a state of affairs where all of the physical truths remain, but without any phenomenal truths. Such a state of affairs would resemble that of Mary's situation before her medical procedure. However, it would not only be one color scientist who can't see color, but it would rather be a situation where no one in the world would have any phenomenal experiences at all. Such beings have come to be known as *philosophical zombies*.[23] Outwardly, they would be indistinguishable with respect to all the physical facts from normal humans – they would have all the same physical structures, function the same way, act in all the same ways as you or I – but they would simply have no internal, phenomenal experience. They would report in exactly the same way as if they *did* have such phenomenal experiences, but they simply

[22] As far as I'm aware, this is not the standard presentation. Nonetheless, the substance of the justification remains the same as the standard presentation. See (Levine, 1983) for the standard presentation of the explanatory argument.

[23] Robert Kirk (Kirk, 1974; Kirk & Squires, 1974) is the first to introduce the zombie thought experiment to the literature. It is put to use most thoroughly in the modern iteration of the conceivability argument by (Chalmers, 1996, 2010), which is directly related to Descartes's formulation of the same style of argument against materialism. Of course, Chalmers's version is updated to reflect further nuances and recent developments in our understanding of these arguments, especially related to Kripke's *Naming and Necessity* (1980) and the fact that Kripke's arguments are directed solely at identity theories of materialism.

wouldn't have them. Such a world would be physically indistinguishable from ours, but phenomenally there would be *no* truths about what people in that world experience. In such a world, zombie-Mary would not learn anything new upon seeing colors for the first time, but she would nonetheless act in the same manner as if she were learning something new. However, since she (along with everyone else) would not have *any* experiences, there would be no new experience for her to learn from. The conceivability of such a world again shows that there is an epistemic gap between phenomenal and physical truths. We might have full epistemic access to *all* of the physical truths, but there is no *a priori* deduction that can use those physical truths to reveal to us any of the phenomenal truths.

To deny this premise, materialists must argue that despite appearances, there is no such epistemic gap. Such a position relies on the idea that the appearance of an epistemic gap is due to insufficient analysis of the cases that the dualist has presented to justify premise one or from an insufficient understanding of phenomenal truths. On further analysis, each case can be explained to show that such an epistemic gap is not present. There are various kinds of materialism that follow this strategy, but in general, the idea is that we can, in fact, deduce or otherwise determine the phenomenal truths from the physical truths without appeal to any further empirical knowledge of the world; it is enough to know the physical facts (along with some other minor facts, like indexical facts and the fact that those are all facts there are) to determine the phenomenal facts. Due to arguing that the phenomenal facts are *a priori* knowable from the physical facts, this is called *a priori* materialism.[24]

The basic rationale behind claiming that there is an *a priori* connection that we nonetheless are not aware of is that we have mischaracterized phenomenal properties in saying that they cannot be captured by the functions and structures that Mary knows. On the contrary, they are not any different from the functional or structural properties that we readily recognize as physical. It is just a matter of figuring out *which* functions and/or structures would count as phenomenal properties of experience, and so it is something of an illusion to think that we could not *a priori* deduce phenomenal truths from the physical truths. This illusion is driven by the fact that we do not yet know all the (relevant) physical truths and we make a mistake in thinking that an *a priori* deduction is impossible based on our inability to make such a deduction with our current physical knowledge. In order to show that a deduction is impossible, we must

[24] Under Chalmers's classification, this is Type-A materialism.

do more than simply show that we have not yet figured it out. *A priori* materialism claims that we are making this kind of mistake.

The idea that we are making this mistake is not suggested without its own justification. It is often argued that we have made this exact kind of mistake in the past. It was once thought that we could not *a priori* deduce the existence of life from merely physical facts, but this is fairly widely held to be false. We now recognize exactly what functions and structures of physico-chemical compounds are those that produce the phenomenon that we call life. *A priori* materialism argues that we have reason to suspect that an analogous mistake is being made with respect to consciousness in general, and phenomenal properties in particular. This view states that phenomenal talk can be reduced to the language of physical functions and structures. Thus, it is often called *reductionist materialism*, where the predominant view is called *functionalism* because all mental properties (including phenomenal properties) are functional states of the physical brain. As such, phenomenal properties are ontologically dependent on physical properties and this dependence is *a priori* knowable. Thus, we are mistaken in thinking that Mary would learn anything new by seeing color for the first time or that she would leave anything out of her explanation of color or that philosophical zombies are a genuine possibility.

The problem with this analogy with life and its *a priori* deduction from physico-chemical functions is that there is simply nothing to life beyond the functions of living beings whereas our own experience of the world reveals a subjective character of that experience that does not yield to the functional analysis that life does. What separates living entities from non-living entities are the functions that living entities perform that non-living entities do not, and these can be explained in terms of living entities being composed of various organic molecules in such and such formations. In the case of life, it wasn't that the concept of life had been insufficiently analyzed. Vitalists recognized that life would be explained by whatever plays the functional role(s) involved in being a living organism, but they didn't think that chemicals could ever play such roles.

But the case is different in the mind-body problem. There are many different properties that might separate the mental from the physical, but the relevant distinction under consideration is that of phenomenal properties as properties of individual experiences versus physical properties as intersubjectively confirmable properties. But phenomenal properties of experience aren't the functional roles that they in fact seem to play. Pain is just the property of experience that is painful even if it might also

serve the functional role of notifying us of bodily damage. Since pain isn't a functional property, there isn't a functional role that *needs* to be played by a physical property. So, either *a priori* materialism misunderstands the character of phenomenal properties or it is using a different distinction between the mental and the physical. If the former, then obviously the analogy with life can be disregarded. Life can be reduced to function in a way that phenomenal properties, properly understood, cannot be. If the latter, then either *a priori* materialism is simply talking about a different, though perhaps related, problem, or there needs to be an argument that the distinction between phenomenal and physical properties is in some way problematic. It is only this final strategy that would allow *a priori* materialism to appropriately respond to the premise at hand.

The main way to explore this strategy is to say that phenomenal properties do not exist at all and thus hold a kind of *eliminative materialism*.[25] Thus, the distinction between phenomenal and physical is illusory in that it is presented as a distinction between two types of properties in the world, but one property type, the phenomenal, doesn't exist. So, this isn't a genuine distinction, there are only physical properties. However, when we reflect on what the phenomenal properties are, this is tantamount to saying that there are no subjective properties of our experiences of the world, which is to say that there is no such thing as a subjective experience. It's hard to make sense of such a proposal given the obvious fact that the only interactions we have with the world are through our subjective experiences of it. Indeed, how could it be otherwise? It is no better if we try to say that these experiences are mere illusions because illusions are just subjective experiences with no corresponding objective referent. But the phenomenal properties just *are* the subjective experiences, whether or not they correspond to anything. So, there can't be any illusion about their existence; if they appear to exist, then they exist. We may, of course, be mistaken as to their *nature*, which is enough to open the door for materialism, but then we need to know what their nature is if not the

[25] There is need for some disambiguation on what view I am calling eliminative materialism here. First, early proponents of something called eliminativist materialism, like (Feyerabend, 1963; Quine, 1960; Rorty, 1965), vacillate between rejecting the existence of mental states and claiming that mental states just are brain states. The view called eliminativism is used to refer just to the former view nowadays, while the latter view is reductionist materialism. Second, many views called eliminativist materialism reject the existence of folk psychological mental states like belief and desire, but either don't comment on, or don't directly argue against, the existence of phenomenal properties. This includes prominent eliminativists, like (P. M. Churchland, 1981; P. S. Churchland, 1986; Stich, 1983). The Churchlands actually argue for functionalism about phenomenal properties elsewhere (P. M. Churchland & Churchland, 1981). So, although best known as eliminative materialists, the Churchlands of that 1981 paper count as reductive materialists on this account since I am only concerned with phenomenal properties or "qualia" as they call them (see pg. 129-130). On the other hand, what I am calling eliminative materialism is a view that argues that the phenomenal properties of our experiences do not exist at all. Such a view can potentially be seen in (Dennett, 1978, 1988). Related views can be seen in (Rey, 1982, 1986; Wilkes, 1988, 1995) where they argue against the existence of phenomenal consciousness itself.

nature that they appear to have. Moreover, to claim that phenomenal properties have a different nature from what they appear to have is to make the same claim as reductivist materialism. So, to argue for this would lead back to the problems discussed above.

If there is no *a priori* evidence that could prove an ontological dependence relation between the phenomenal and the physical and such relations must be provable if they exist, then the only avenue left for materialism is to show that there is some *a posteriori* evidence. Premise four claims that there is no such *a posteriori* evidence that could prove an ontological dependence relation between the phenomenal and the physical. As such, if premise four is true, along with the other three premises of the Epistemic Argument, then there is an ontological gap between phenomenal and physical properties and materialism (along with idealism) is false. This premise means that there is no amount of science that could discover that phenomenal properties depend on physical properties, and so there is no equation akin to $e=mc^2$ for the mind-body problem.

Of course, those who deny premise four must show how there could be such an equation, even if we don't have it yet. They must show that it is possible for an empirical investigation to uncover something about either phenomenal or physical properties (or property instantiations) such that we could explain how the one depends on the other despite their apparent differences. The standard ontological dependence relation for this view is identity. Thus, on this view, the phenomenal and the physical are identical and this identity is knowable *a posteriori*. The view is called *a posteriori* materialism.[26] Again, *a posteriori* materialism argues for this identity by analogy with scientific identities that have been discovered in the past. One common example is the identity between the common drinking substance that is found in lakes, rivers, and falls from the sky when it rains, namely water, and the chemical compound dihydrogen

[26] This corresponds to Chalmers's Type-B materialism, with the important difference mentioned above where I discuss strong necessitarian materialism. Chalmers also discusses a Type-C materialism, whereby the epistemic gap discussed in premise three can be closed in principle but is open due to our current state of knowledge. I agree with Chalmers that this type of response collapses into one of either Type-A or Type-B materialism. There does *seem* to be some room if one thinks, for instance, that the epistemic gap might be either *a priori* or *a posteriori* knowable. However, this is a misunderstanding of the operative idea of the existence of an epistemic gap. The existence of an epistemic gap is not some gap that exists and therefore we can figure out how to close it via *a priori* or *a posteriori* reasoning. Rather, saying that there exists an epistemic gap is simply to say that there is no *a priori* knowable connection between phenomenal and physical properties. As such, the concept of an "empirically closable gap" would mean that there is no *a priori* knowable connection, but there is an *a posteriori* discoverable connection. This is obviously just to admit that there is an epistemic gap, but no ontological gap. This view is *a posteriori* materialism or, in Chalmers's terms, Type-B materialism. The end result is that there are only two options here: either there is a gap, meaning that there is not an *a priori* knowable connection between phenomenal and physical properties or there is no gap, meaning there is an *a priori* knowable connection between phenomenal and physical properties.

monoxide, or H_2O. That these two things are identical is not obvious; it is not something that we could deduce *a priori* in the way that we could deduce the fact that two plus two is four. We had to engage in some significant empirical investigation to establish the identity.

In the same way, *a posteriori* materialists argue that we must engage in some very serious empirical investigation into phenomenal and physical properties in order to discover that they are identical. They say that premise four mistakenly draws the conclusion that no empirical investigation into the world could possibly prove an identity between phenomenal and physical properties from the idea that we currently cannot see the identity based on current empirical knowledge. Though related, this is a different line of argument from that pursued by the *a priori* materialist. For both types of materialism, dualism is trying to draw conclusions too early in the process or with too little evidence. The difference lies in the missing type of evidence that the two versions of materialism think is necessary to prove the dependence. For *a priori* materialism, the missing piece of the puzzle is a line of *a priori* reasoning, much like a proof in mathematics, that would make it clear that the phenomenal depends on the physical. For *a posteriori* materialism, the missing piece of the puzzle is empirical evidence that would reveal that what we have been calling phenomenal properties are none other than a type of physical property that we either have recognized by a different name or have yet to discover since our knowledge of the brain is so limited.

That H_2O and water are identical is an identity that only became known once chemistry had advanced far enough to realize that water has an underlying chemical structure that gives rise to the aquatic properties we know about water. Such an identity could not have been shown to exist without the empirical investigations required for the discovery of H_2O. However, once we discovered H_2O, we could see that water and H_2O are identical. A similar thing happened with genes and DNA, heat and mean kinetic energy, and a whole host of scientific discoveries. For *a posteriori* materialism, we are in a position like we were before the discovery of these underlying features of the world. In those situations, it would have been too early to say that no amount of scientific work could discover a physical identity. Indeed, in each case, a physical identity was discovered. Similarly, it is too early to say that no amount of scientific work (particularly on the brain) could discover an identity between phenomenal properties and physical properties (of the brain).

Again, the dualist response is to point to a crucial disanalogy in these cases where an *a posteriori* identity was discovered and the case we are in with respect to phenomenal properties. And, again, it has to do with how the relevant concepts work. In all of the cases of scientifically discovered identities, at least one

of the concepts involved have their reference fixed via something contingent to their referent. For instance, the concept 'water' refers to the watery stuff that is in lakes and rivers that we drink, etc. But none of these features are essential to water; they are just how we interact with it. As it turns out, being H_2O is what is essential for something to be water. On the other hand, the referent of the concept 'H_2O' is fixed by its essential property, being H_2O. If phenomenal concepts work like 'water', then there might be some physical property that we can discover that it would be identical to, thus allowing for the possibility of an *a posteriori* identity that we have yet to scientifically discover. However, it doesn't seem like phenomenal properties work like the concept 'water'. On the contrary, they seem to work just like the concept 'H_2O'. That is, they seem to refer via what is essential to their referent and that essence is, unlike H_2O, not necessarily physical. And if the essence isn't necessarily physical, then it can't support a necessary connection between the physical and the phenomenal. Take, for an example, the concept 'pain'. On the *a posteriori* materialist story, 'pain' refers to something physical (e.g. bodily damage) via a painful experience. But the natural response is that anything that is not painful is not a pain, and painful experiences don't need to be physical. So, if 'pain' refers to something physical, it refers via something that is not essential to pain, and so pain isn't necessarily physical.

So, it seems like the concept 'pain' functions more like the concept 'H_2O' in that it fixes its reference via the very thing that makes its referent that thing, i.e. it refers essentially. As such, there is no room to think that some discovery about the brain might show us the true underlying feature of pain that we were previously unaware of. We already know the essence of what the concept 'pain' refers to, the phenomenal experience of pain. Since they cannot be essential to the referent of 'pain', any functions or structures of the brain will be contingently related to the pain that we experience. Therefore, any physical properties of the brain cannot be identical to phenomenal properties. Even if the phenomenal and physical properties perfectly covary and scientists can show that pain is always had with such and such a brain state, the fact is that the referents of phenomenal concepts like 'pain' and physical concepts cannot be proven identical in this way.[27] With an *a priori* dependence and strong necessities already shown to have several problems

[27] This naturally leads *a posteriori* materialists to seek other ways in which phenomenal concepts are different from physical concepts and thus leads them to deny premise one which was discussed above. See footnote 20 for the oddity of continuing to call such a view *a posteriori* materialism.

given what was said about the first two premises, the Explanatory Demand cannot be met and thus there is no recourse but to conclude that phenomenal and physical properties are not the same.

Even if we are willing to accept this argument, dualism still has difficulty meeting the Ontological Demand. If we suppose that dualism is only claiming that there is an ontological difference between the phenomenal and the physical, then this puts it directly up against the Ontological Demand. The Epistemic Argument shows that they are different, but it doesn't show how they nonetheless fit together. Again, we can see the difficult tension between the Explanatory Demand and the Ontological Demand. Once we can see how the phenomenal and physical fit together, then it's hard to see how they could still be ontologically different. But once we've shown that they are ontologically different, it becomes difficult to see how they can fit together. By giving such a strong argument against materialism, dualism lacks the resources to meet both demands. There would need to be some law of nature to explain how phenomenal and physical properties fit together, but this doesn't seem to be forthcoming. Indeed, it seems impossible for there to be such a law of nature given the causal closure of the physical that the Causal Argument has forcefully argued.

3. A Brief Note on Parsimony

All of this leaves us in quite a mess, and you would be right to wonder if parsimony can help resolve the situation at all. The idea is that the simpler the theory, the more likely it is to be true. Since materialism is the simpler theory, we should take it to be more likely and (provisionally) conclude that dualism is false. This note will be necessarily brief, but I don't think that parsimony, independent of meeting the Explanatory Demand can help resolve the problem, even provisionally. At the end of the discussion of the Causal Argument, I said that if we accept the results of that argument on face value, we still don't know why or how phenomenal experiences exist at all if the functions can be performed by the brain without any need for these experiences. The easiest way to understand parsimony arguments is that they claim that we ought to get rid of theoretical posits that are not needed in our explanation. Thus, we ought to get rid of phenomenal properties because we can explain all of the things that we do without needing them. However, the experiences are not theoretical posits. Rather, they are the phenomena that we are trying to explain. Far from explaining phenomenal properties, a parsimony argument like this simply insists that we shouldn't worry about explaining them. But it isn't a particularly helpful use of parsimony to say that we should just stop worrying about the things that we wanted to explain when we can explain *other* things without using

them. In essence, this doubles down on materialism's apparent ability to meet the Ontological Demand while not providing anything of use for materialism in trying to meet the Explanatory Demand. But it is precisely materialism's inability to meet the Explanatory Demand that is its main flaw. As such, insofar as parsimony might help resolve this impasse, it will only do so if both dualism and materialism can be shown to meet the demands in Chapter 1. We face a situation far from that, and so parsimony is no help to us.

4. Proposing a Way Forward

In my own experience, when I meditate sufficiently on one of either the Causal Argument or the Epistemic Argument, I am drawn to conclude that it must be right. However, upon meditating on the other argument, I am equally drawn to conclude this other argument must be right. I find that I am not alone in this experience. Anyone who, seriously and without prejudice, engages with and understands the Epistemic Argument is transformed into an anti-materialist (at least for the moment). But upon understanding and engaging with the Causal Argument, that same person is transformed into an anti-dualist (again, at least for the moment). This doesn't always happen immediately; it sometimes takes many months or years struggling with the argument to have this experience. Surely, many never get to this feeling at all. In the current context, it often seems much easier to take one of the "party lines" outlined above and insist that the opposing argument can be defeated. I don't mean to say that this is entirely disingenuous on anyone's part.[28] It is true that most people are pulled more strongly by one argument than by the other, but rather than being convinced that the weaker argument has been proven wrong, it seems that they feel that the weaker argument *must* be wrong. It is the weaker of the two, after all, and given the fact that the current debate presents only two options, it is natural to conclude that the weaker option somehow misses something, even if we don't know quite *what* it misses. So, most go about trying to figure out what is wrong with the weaker argument, but even while there are logically coherent responses to the arguments, these responses fail to get to the heart of the problem – dualism fails to meet the Ontological Demand and materialism fails to meet the Explanatory Demand. There is, of course, always the option of "biting the

[28] Although I'm sure that Griffin (1998) is right when he says that many people are driven by non-philosophical concerns about what they assume must be true in order for something else to be true (or false) that they really want to be. For instance, if one really wants the afterlife to exist, one might be driven to insist on the truth of dualism beyond what would be prudent on the basis of reason alone. Similarly, if one thinks that materialism is the only way to solve the mind-body problem while respecting science, then one might be pushed toward materialism. I hope that such "fearful" and "wishful" thinking is not the main driver of most philosophical argumentation on the issue, but it surely operates in the background in an important way.

bullet" – accepting a view despite the consequences (that are well-known and widely regarded to be negative on the whole). But even that phrase reveals that this is something we'd rather avoid. It's only something we do once we've exhausted all options and are tired of thinking about the damn problem. Fortunately (or unfortunately, depending on who you ask), many philosophers are rather tireless, and so refuse to bite the bullet. The result is that philosophers have reached for increasingly more ludicrous attempts to show what is missing, but to no avail.

The most recent major trend has been to try out panpsychism (again), the view that all things are conscious, even standard non-mental objects like rocks and tables. Indeed, the view has a long history as an alternative to standard forms of dualism or materialism or idealism.[29] One of the most influential formulations was developed by Leibniz, who seems to have been uninterested in making his view clearly fit into any of these more standard categories. This gives panpsychism something of an air of being a different view. However, as I argued in the previous chapter, given the question of whether phenomenal properties are fundamentally the same or different from physical properties, panpsychism will end up fitting one of the three standard options. One way to formulate panpsychism more explicitly shows it as a form of dualism and it again fails to adequately meet the Ontological Demand for the same reasons as other forms of dualism. Other formulations are kinds of idealism and thus fail to meet the Explanatory Demand. As something of a recognition of the inadequacies of panpsychism as standardly conceived, David Chalmers has proposed a variation that he calls panprotopsychism. This is the view that everything has fundamental properties that are not themselves phenomenal but rather are proto-phenomenal – such that they give rise to phenomenal properties when properly arranged. As Amy Kind (2015) argues, this view ends up being a variation of materialism for all intents and purposes. In particular, it suffers the corresponding materialist difficulty of meeting the Explanatory Demand. Beyond this regurgitation of the same problems in new wrapping, panpsychism has another problem. Once sufficiently developed, the best argument for the view seems to be that it hasn't been proven incoherent (Goff, 2017, p. 253).

Perhaps panpsychism is a coherent view and that should certainly count in its favor, but in finding an adequate solution to the mind-body problem, I think we can do better than mere coherence. I think we

[29] See Phillip Goff's *Consciousness and Fundamental Reality* (2017) for the most thorough recent attempt along these lines. There, he goes even further to defend "cosmopsychism", the view that the cosmos is the fundamentally existent object and it is conscious.

can find a coherent view that meets the Ontological Demand and the Explanatory Demand at the same time. It will involve first opening the door to other possibilities beyond the standard trio of dualism, materialism, and idealism. Panpsychism seems like it is an attempt at this, but since it doesn't directly treat the question of whether phenomenal and physical properties are the same, it inevitably misses the mark. On the other hand, the view that I'll defend in what follows, namely neutral monism, more explicitly tries to answer the question of whether the phenomenal and physical are the same. It does run into the same problem that panpsychism does though; as many opponents of the view have argued, it is really just a disguised version of idealism, materialism, or dualism. Either that or it is incoherent. In the remainder of the text, I'll be attempting to develop a coherent neutral monism in a way that avoids becoming one of the three standard views while meeting the demands from Chapter 1. In the final analysis, perhaps my own attempt will fall in the same category as just one more ludicrous attempt to find a solution, but if that's the case, I'll consider myself in good company. I think it's worth seriously considering whether we are seeing all the options even if the option we consider ends up being more ludicrous than anything else on offer. Even if the view ends up being incoherent on a final analysis, I think there is something to be gained by the attempt. Given the impasse presented in the above discussion, this seems to be the only way to make progress.

Chapter 3: Against the Russellians

> "The stuff of which the world of our experience is composed is, in my belief, neither mind nor matter, but something more primitive than either. Both mind and matter seem to be composite, and the stuff of which they are composed lies in a sense between the two, in a sense above them both, like a common ancestor."
>
> - Bertrand Russell, *Analysis of Mind* (1921, pp. 10–11)

1. The Best of Both Worlds

In recent decades, some philosophers have sought to resolve the impasse of the previous chapter by developing alternatives to dualism and materialism as solutions to the mind-body problem, seeking a solution that somehow fits "in between" them. Such a solution is supposed to offer all the benefits of both dualism and materialism while avoiding their corresponding problems. The general label for such solutions is *neutral monism*. This term was coined by Bertrand Russell in his *Analysis of Mind* (1921), the basics of which are in the epigraph above. There are various ways of developing a neutral monism, but the general idea – captured well by the name – is that there is only one type of fundamental stuff and that one type is neutral between the mental and the physical types. This should be distinguished from dualism which claims that the fundamental stuff comes in two distinct types. Here, I'll be focused on critiquing one way in which neutral monism has been developed – Russellian monism – paying particular attention to the way in which it tries to specify how the fundamental stuff of the world is neutral and what this could possibly mean.

Many seem to use the terms 'neutral monism' and 'Russellian monism' interchangeably. However, the general project of finding an "in between" solution to the mind-body problem does not necessarily commit one to the Russellian version of such a view. For instance, Stoljar (2015) distinguishes what he calls 'Nagelian monism' as a different type of neutral monism, though in Stoljar (2014), he refers to this view as a type of Russellian monism – however, even there, he includes a warning that it may *not* be a kind of Russellian monism. I use the term 'neutral monism' for any view that is sufficiently "in between" dualism and materialism in the appropriate way, whatever form this view ends up taking. What makes a view sufficiently in between materialism and dualism is that it satisfies the principal demands that drive philosophers toward each of those views. So, any view that can explain the relationship between the phenomenal and the physical while maintaining that the fundamental property is neutral between these two is what I call neutral monism. And any such view that can meet both demands from Chapter 1 will have provided a solution to the mind-body problem that captures the best of both materialism and dualism while avoiding the main objections to each. On the other hand, I will not use 'Russellian monism' to refer only to

Russell's version of this view; his is merely one variant of Russellian monism. Russell's general solution has been further developed since he proposed, and ultimately abandoned, the view. As such, my critique will be directed at this general subtype of neutral monism, the details of which will be explained below.[30]

To summarize the points from the previous chapters, the principal demand behind materialism is the Ontological Demand which results in the Causal Argument. This basically states that everything in the world can, in principle, enter into causal relations with anything else without systematic overdetermination. Materialists take this consideration and argue that since physics is causally closed, there is no room in the causal structure of the world for anything non-physical. The principal demand behind dualism is the Explanatory Demand which results in the Epistemic Argument. This states that any two things will not be separated by an epistemic gap if the existence of one necessitates the existence of the other – i.e. if we know everything about one thing that necessitates another, then we can in principle derive everything about the second thing. Dualists take this and argue that since there is such an epistemic gap between the phenomenal and the physical, the appropriate necessitation relation cannot hold, and they must be distinct.

I'll argue that Russellian monism will fail to satisfy its primary goal of resolving the mind-body problem by developing a neutral monist solution. Others have made similar arguments, but each of these has been deficient by either only considering particular interpretations of Russellian monism or overgeneralizing the conclusion such that it ends up being an attack on any form of neutral monism.[31] In contrast, I'll be arguing that there is a common problem amongst the different variations of Russellian monism such that it provides no hope for a truly neutral monist solution to the mind-body problem while showing that there is nonetheless some hope for developing a neutral monist solution.

The structure of the chapter will be as follows. In the next section, I'll explain the basics of Russellian monism in order to show what would make any solution to the mind-body problem a version of Russellian monism. In the third section, I'll present and discuss the main dilemma that any Russellian monist must face with an eye toward expanding on variations the view might take. It will turn out that only

[30] A few examples of this view can be found in (T. A. Alter & Nagasawa, 2015; Banks, 2014; Chalmers, 1996, 2013; Feigl, 1967; Goff, 2017; Griffin, 1998; Heil, 2020; Lockwood, 1989; Maxwell, 1979; Stoljar, 2001; Strawson, 1994). See (Stubenberg, 2018; Wishon, 2015) for detailed accounts of the differences between Russell's view and what goes under the guise of Russellian monism these days.

[31] See (Ayer, 1971; Stace, 1944) for examples of the former deficiency and (Cutter, 2019; Kind, 2015; Pautz, 2015) for examples of the latter.

some interpretations of the view fall to this dilemma. Other formulations of Russellian monism can avoid this dilemma, though they run into other difficulties. In the fourth section, I'll draw out an isomorphism between a central debate about the nature of dispositions and the mind-body problem such that even if Russellian monism is interpreted in one of the ways that avoids the general dilemma presented in the previous section, it should not be viewed as a neutral monist solution to the mind-body problem. Finally, I will discuss further implications of the argument for the mind-body problem generally and neutral monism specifically.

2. The Basics of Russellian Monism

The fundamental insight of Russellian monism – if, indeed, it is an insight – is outlined in Bertrand Russell's *Analysis of Matter* where he argues that "the aim of physics, consciously or unconsciously, has always been to discover what we may call the causal skeleton of the world" (1927, p. 391). This causal skeleton is merely the structure of how various events are causally related to one another. What physics does *not* tell us is what the intrinsic nature of these events is. If physics studies an event, it will only discover a more minute causal structure. But nothing of the intrinsic nature of any event will be given by the physical study of it. To this is added the speculative idea that our conscious experience of an event is the intrinsic nature of the brain state which corresponds to that event. For example, a neuroscientist may study the brain of someone drinking a cup of coffee and say that it produces such and such a chemical response in the brain. But if we break down what these chemical structures and their receptors in the brain are, we will only get further explanations in terms of their dispositions to act and react when presented with certain stimuli. Furthermore, the argument goes, *any* further scientific study of this event will produce only further explanations in dispositional terms, all the way to fundamental particles (which are also understood through their causal interactions). For instance, electrons are understood in terms of attracting positively charges particles, among other interactions. On the other hand, the subject drinking the coffee can use the categorical terms that there is a sensation of certain flavors, aromas, more energy/alertness, etc. This is our only access to the intrinsic nature of the world and is really only an access to the intrinsic nature of our own brains since the corresponding causal structure studied by physics (or, indeed, any science) will be in the brain, not anywhere else. So, matter is the causal structure of the world, while our conscious experiences are the intrinsic natures of our brain states. We have no access to the intrinsic nature of anything else.

This insight about the aims of physics isn't merely something that Russell asserts without preamble. Historically, it pre-dates Russell by at least some 100 years.[32] The late British empiricists were operating under such a notion in their conceptions of the "natural philosophy" of Newton, although it isn't always so plainly stated. Even critics of empiricism came to have this understanding of physics. It is clearly stated by Dugald Stewart – a student of Thomas Reid and a critic of Hume – in his *Elements of the Philosophy of the Human Mind*. There, he states that physical science has been successful by putting aside questions about the real nature of matter and instead focusing on "the humbler province of observing the phenomena it exhibits" (1833, p. 4) while urging the study of the mind to follow suit. Later, Ernst Mach, as physiologist, and William James, as psychologist, did follow suit (whether or not they were aware of Stewart's suggestion) and developed their own versions of neutral monism – which seem to be nascent versions of Russell's formulation. Moreover, physicists seem to be operating under similar conceptions of their own field of study; ask a physicist what "matter" is and they'll answer that matter is just made up of the fundamental particles. But asked what *those* are, and the physicist will be hard-pressed to give any answer other than statements of causal relations that their fundamental particles enter into – to put it incredibly simply. Electrons are the things that are attracted by positively charged objects, repelled by other negatively charged objects, spin in certain ways, have a certain mass, etc. Neutrinos interact with other objects via gravity and the weak subatomic force, have some tiny mass, etc. In general, matter is whatever responds to such and such stimuli in such and such a way. And any further discoveries of how to break these particles into sub-particles would just lead to further descriptions of these sub-particles purely based on their dispositions to act when presented with various stimuli. These dispositions are exactly what are describes be the fundamental laws of nature.

The speculative placement of conscious experience into the intrinsic properties of the brain is a direct result of combining this insight with a recognition that any explanation of consciousness should be naturalistically respectable. Russellian monists agree with what I said in Chapter 1 about the naturalist presupposition in the question at the heart of the mind-body problem. That is, they agree that recourse to such entities as God's will or supernatural spirits simply isn't in the spirit of the question. In some sense, it

[32] Indeed, some variation of the insight may even be seen as early as Plato's dialogues. However, since those pre-date modern science, it seems a bit anachronistic to insist that Plato had some insight on the subject matter of *modern* physics.

is merely giving up the game and revealing that we are no longer seeking an answer. When these two ideas are accepted, Russellian monists find that the solution to the mind-body problem must lie in the intrinsic nature of the brain. There is no other place in the natural world to fit conscious experience. Thus, consciousness is not at odds with physics, the causal closure of the physical, or any other claims that physicists might come up with, now or in the future. Nonetheless, to discount consciousness from our explanation of everything in the world would be to miss something about the world; physics can't explain everything. Indeed, it's not even in the business of trying; it only explains the dispositional nature of things, leaving their intrinsic natures out of physical explanation entirely. This is not to say that Russellian monism attempts to explain everything in the world. It is intended to be an explanation of consciousness, specifically an explanation of how consciousness fits into the physical world. In other words, Russellian monism is intended to be a solution to the mind-body problem, but as I mentioned above, it need not give us a full explanation of what the intrinsic natures of physical objects (that are not the brain) are. Some formulations may attempt this, but others (like Russell's version) explicitly state that these are unknowable.

So, according to Russellian monism, consciousness cannot be reduced to the physical and yet it fits perfectly within physical theories about the causal structure of the world. There are no spooky interactions between the mental and the physical, the supervenience of the mental on the physical is preserved (though not asymmetrically), and consciousness provides some access to a part of the world that physics is unable to grasp. Furthermore, the reduction of the mental to the physical does not hold up. Rather, the world is composed of events which are physical in that they have a causal structure, but they are not only physical in that they have some intrinsic nature. At least part of the intrinsic nature of the world (the part where brains are) is conscious; what the intrinsic nature of the rest of the world is is unknown (and perhaps unknowable). Whatever we might discover about it, it won't be physical. On the other hand, were we to merely know the intrinsic nature of the world, we would be missing out on the causal structure. As such, the fundamental nature of the world is neither physical nor mental. It is neutral between these two, for without the intrinsic nature there would be nothing to have any causal effects in the world and without the causal structure it would make no sense to talk of the intrinsic nature of something. Or so the Russellians would have us believe.

In order to further clarify the view, we must understand two assumptions that the Russellian monist is making about causal structural properties, or dispositions and intrinsic, or categorical, properties. The

first assumption is that there is a genuine conceptual distinction between these two property types. Of course, this *seems* obvious, but the claim has been challenged.[33] Regardless of the specifics of this challenge or whether you think it is ultimately successful, the force of this challenge can be taken to be a demand that the conceptual distinction needs to be carefully laid out. Some ways of drawing the distinction simply won't do the trick. The second assumption is that these two sorts of properties never exist alone; they are always paired with the other.

The conceptual distinction between dispositions and categorical properties operative in Russellian monism is, roughly, the distinction between extrinsic and intrinsic properties. Russellian monism maintains that any physical property is dispositional, meaning that its identity is fixed extrinsically, particularly by its causal relationship to other things. Electrons have the property of being negatively charged precisely because, when in the appropriate conditions, they enter into certain causal relationships with similarly charged things and other causal relationships with oppositely charged things. Meanwhile, phenomenal properties are always categorical meaning that they have their identities fixed intrinsically. Someone has the property of being in pain because they experience pain, not due to anything that this pain may cause or be caused by. This is a natural way to understand the conceptual distinction; however, as we'll see below, what the metaphysical relationship is between dispositional properties and their categorical counterparts is not altogether clear. What is clear is that if this conceptual distinction cannot be fruitfully maintained, then the central insight of Russellian monism will have been shown to be no insight at all.

The second assumption is that neither dispositions nor categorical properties exist on their own. Dispositions are *always* connected to a categorical property and vice versa. For now, I'm happy to concede that this position is, at the very least, plausible enough. The idea of a bare disposition – i.e. a disposition with no connection to a categorical property – seems strange at best and impossible at worst.[34] The suggestion is that there has to be something intrinsic to the object that connects its causal relations to it. A bare disposition would be a property that makes an object causally interact with other things, but which doesn't do so because of anything intrinsic to the object. But then, on the face of it, it's more than a bit

[33] See (Mellor, 1974) for a classic argument against the existence of a genuine conceptual distinction between dispositions and categorical properties.

[34] Though, see (Ellis, 2001, 2002; Hildebrand, 2014; McKitrick, 2003; Molnar, 1999, 2003; Mumford, 2006) for arguments for the possibility of bare dispositions.

mysterious why the object should interact in the peculiar way that it does rather than some other way. This is a plausible reason to suppose that every disposition has some connection to a categorical property. On the other hand, the idea of a bare categorical property – i.e. a categorical property with no connection to a disposition – is equally unbelievable. We simply have no reason to think that there are properties (of any kind) that have no connection to any causal interactions with anything else in the world (indeed, this kind of reasoning often enters into arguments for materialism, as mentioned in the previous section). On this assumption, there may be properties that themselves don't cause anything, but they at least have to be connected in some way to a property that *does* enter into causal interactions. These aren't full-fledged arguments, though, so if it turns out that this assumption is incorrect, Russellian monism will again turn out false.[35]

With this, we can outline the essentials of a Russellian monist solution to the mind-body problem. First, Russellian monism must contain no supernatural entities or properties, where this is understood merely as the idea that no entity or property exists outside the laws of nature. Second, the view is a neutral monist view. That is, neither phenomenal nor physical properties are fundamental, but rather fundamental properties are neutral between these two categories. Third, physical properties are a subset of dispositions and phenomenal properties are a subset of categorical properties. What seems to have been recognized here is that the same principal consideration concerning causation that leads one toward materialism applies to certain monist views about dispositions while the same consideration that leads one toward dualism about the mind-body problem can similarly lead one toward a dualist view about dispositions and categorical properties. As such, looking to dispositions and categorical properties to do some work in solving the mind-body problem does make some sense.

In the next section, I'll consider ways in which the neutral, phenomenal, physical, dispositional, and categorical properties might all be related. This will cover all the logical space of potential formulations of Russellian monism. Previous taxonomies of Russellian monism have tended to focus on different metaphysical relations between the neutral and physical/phenomenal properties one might use to develop the view. In the following section, the focus will instead be on the relationship that might hold between the various kinds of properties that figure in Russellian monism, paying special attention to the relationship

[35] See (Hiddleston, 2019) for a challenge to Russellian monism on the basis of undermining these assumptions about dispositional and categorical properties.

between dispositions and categorical properties. So, the following varieties of Russellian monism are orthogonally related to the division introduced in Chalmers (2013, 2016) between "emergentist" and "constructivist" varieties. The arrows in the diagrams signify a determinable metaphysical determination relation, which can be read as whichever determinate relation you prefer – emergence, constructive, or other. The basic issues that I discuss will arise regardless – the view will fail to capture both considerations that motivate materialism and dualism. This is useful as it allows us to see that there is a similar range of problems that Russellian monism encounters regardless of the determinate metaphysical relation involved. As such, it allows us to see that if there is any hope for neutral monism, it lies in developing a view that does not rely on the so-called Russellian insight.

3. Varieties of Russellian Monism

The most direct objection to Russellian monism – or, indeed, any neutral monism – can be thought of as a dilemma. The dilemma argues that either the view is incoherent or, if it can be properly understood, it fails to meet the two demands adequately. The first horn of the dilemma is often not explicitly stated as the objectors are trying to be as charitable as possible and really understand the position. Invariably, though, according to the objectors, there simply isn't a coherent solution to the mind-body problem that meets both the Explanatory and Ontological Demands.[36] This isn't merely a concern that the solution isn't novel enough. The search for a neutral monist solution begins with a dissatisfaction with the standard solutions already on offer. So, if Russellian monism fails to capture the considerations that motivate dualism and materialism, they will have failed to differentiate themselves from one of these solutions. And if it fails to differentiate itself from one of these solutions, it will have failed in its central task.[37]

[36] The clearest presentations of a version of this argument can be found in (Cutter, 2019; Kind, 2015). However, these previous versions draw the conclusion that there can be no neutral monist solution to the mind-body problem. I think this is a mistake. The problem is with Russellian monism in particular, not with neutral monism in general. I'll have more to say on this in Section 4.

[37] Thus, I disagree with (Chalmers, 2013) where he argues that dilemmas such as these are merely verbal disputes about what to label the view. The dilemma is often framed in terms of whether neutral monism is really a different view from the three standard options, but the force behind the dilemma is that if Russellian monism is a coherent view, it will nonetheless face all of the problems of one of the standard solutions (though possibly dressed differently, and possibly more that aren't faced by the standard views). So, to say that the view is materialist (or dualist) means it has failed to meet the Explanatory (or Ontological) Demand. And to say that the view is idealist is to say that it has failed to capture either the Explanatory Demand or else will fail to respect the naturalistic presupposition. The labels themselves don't actually do any of the work.

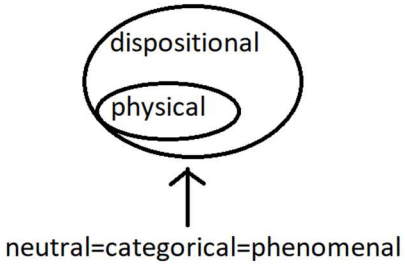
Formulating neutral monism in a coherent and stable fashion is no easy task, so the confusion as to what the view actually entails is understandable. Russell himself was aware of this when he stated that "Any attempt to classify modern views, such as I propose to advocate, from the old standpoint of materialism and idealism, is only misleading.…I shall try to persuade you…that matter is not so material and mind not so mental as is generally supposed. When we are speaking of matter, it will seem as if we were inclining toward idealism; when we are speaking of mind, it will appear as if we were inclining toward materialism" (Russell, 1921, pp. 34–36). Here, instead of trying to give the correct historical interpretation of Russell's view or his objectors' or the view of any other formulators of Russellian monism, I'll simply explore possible interpretations and judge how well each possibility can avoid both horns of the above dilemma.

The initial objections to Russellian monism tended to identify Russell's view as just a disguised idealism or phenomenalism.[38] The basic force of the objection revolves around how we are to think of the "neutral stuff" that lies "between" the mental and physical. This is one of the most difficult things for a neutral monist to spell out, and it seems that Russell struggled with it as much as anyone before or since. Under Ayer's interpretation, "this more primitive [i.e. neutral] stuff was thought by [Russell] to consist mainly of the sense-data, or sensibilia, out of which, as we have seen, he then believed the physical world to be constructible. These elements were also supposed to enter into the construction of minds" (1971, p. 110). From this, he infers that Russell's monism is merely a poorly veiled idealism. Indeed, on this reading, it is hard not to see it as a version of idealism. And when the intrinsic or categorical nature of the basic elements of the world are said to be phenomenal, which according to Russellian monism, it is in the case of the intrinsic nature of the brain, it is easy to interpret that as saying that the sensibilia are the basic building blocks out of which the more complex things of the world (like minds and physical objects) are made. Thus, one might combine the Russellian view of phenomenal and physical properties while taking a standard view of dispositional and categorical properties (categoricalism) which says that fundamental properties are intrinsic and therefore categorical. On this understanding, the neutral properties are categorical, thus I call this formulation Categoricalist Russellian Monism. This is even suggested by the importance placed on

[38] These objections can be seen in (Ayer, 1971; Stace, 1944). In some form or another, this criticism has stuck with Russell's view into the 21st century. See (Soames, 2003) for a more recent version of the criticism. Also, see (Pincock, 2005) where he argues that this criticism is a misunderstanding of Russell. As I stated in the main text, I'll try to steer clear of this interpretational morass as much as possible.

intrinsic properties in showing that physics cannot explain everything, but only the causal interactions. But there are three ways one might think that the neutral, phenomenal, and categorical are related.

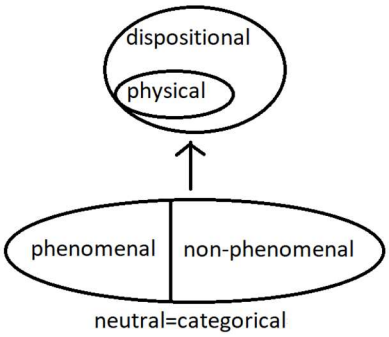
On the most natural understanding of Categoricalist Russellian Monism, phenomenal properties are identified with categorical properties. Many proponents of, and objectors to, Russellian monism have interpreted the view in this way and correctly see it as a form of panpsychism. The problem with this view is that it clearly doesn't meet the central task of developing a neutral monist solution. The neutral properties, being categorical, will be phenomenal properties. And since the neutral properties are fundamental, the physical properties will be properly seen as resulting from mental properties. But under this interpretation, Russellian monism will fail to capture the motivations behind either materialism or dualism. It will fail to capture the materialist causal consideration because either the causal closure of the physical will have to be violated or the fundamental properties will still not be able to causally interact with the physical properties. And it will fail to capture the dualist epistemic consideration because the same epistemic gap that dualists argue exists for materialism will exist on this view as well, though the fundamental property on this view is phenomenal, not physical. Thus, this panpsychist formulation of Russellian monism is clearly susceptible to the second horn of the dilemma.



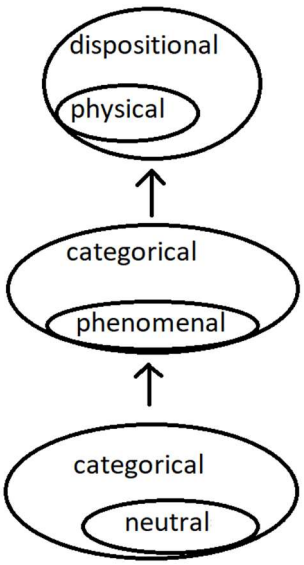
In the *Analysis of Matter*, Russell claims that we don't (perhaps can't) know much about the intrinsic qualities of things like electrons "for we know nothing of the intrinsic quality of the physical world and therefore do not know whether it is, or is not, very different from that of percepts" and even suggests that it is more natural to infer that the intrinsic nature of the physical world lacks any resemblance to our own phenomenal experience of it when he says that "the physical world, it seems natural to infer, is destitute of colour" (1927, p. 264). Thus, although we may have good reason to think that physical objects have *some* intrinsic quality[39], we have no reason to infer that they are anything like the intrinsic qualities of our own experiences of the world. This leads us to a second formulation of Categoricalist Russellian Monism which stems from the idea that if the neutral stuff is sensibilia, then we need to divorce our notion of sensibilia from our own phenomenal experience of the world. The intrinsic nature of the physical world (except for

[39] This is one of the assumptions of the view mentioned in the previous section.

that part occupied by our brain) is a mystery to us, and thus, we should not suppose that it is anything like our phenomenal experiences. Many developers of Russellian monism recognize the seeming absurdity of claims that fundamental particles might have something like phenomenal experience and make moves in this direction to distance themselves from such views. This interpretation, however, similarly fails to avoid the second horn of the dilemma. Although it captures the dualist epistemic consideration, it fails to capture the causal consideration in the same way that the first formulation failed to do so. By supposing that the neutral properties are intrinsic and therefore categorical, the fundamental properties are either phenomenal or non-phenomenal. Thus, we have two types of fundamental properties, not one. But if these fundamental properties causally interact with the physical properties, the causal closure of the physical is violated. If not, then it is not true that everything can (in principle) causally interact with everything else.



The third formulation of Categoricalist Russellian Monism is suggested by Chalmers' (2013) development of panprotopsychism which is meant to capture the best of the panpsychist view above while avoiding various problems associated with it. On this view, the protophenomenal properties are fundamental categorical properties of physical objects. Out of these protophenomenal properties, the phenomenal properties arise. These protophenomenal, categorical properties could be what a Russellian monist means by a fundamental property that is neutral between the physical and the phenomenal, but out of which both physical and phenomenal arise. By the Russellian insight, physical properties are all dispositional while the phenomenal properties are categorical. The neutral properties, being identified with the protophenomenal, in this case would be categorical, but they would be the *fundamental* categorical properties whereas the phenomenal properties would be non-fundamental categorical properties. There are two ways that this could work. Either the neutral properties can give rise to the phenomenal properties which then give rise to the physical properties, or the neutral properties could give rise to both physical and phenomenal properties directly. This second option (which I take to be Chalmers' view) will be mentioned below under the label



Reductionist Russellian Monism, so I will not address it here. The first option, though, is a very peculiar view. I'm not sure anyone holds it, and for good reason. On this view, there is a hierarchy of properties where the physical is least fundamental while the neutral is most fundamental and the phenomenal is between them. The problem is that this view fails to meet either demand in the same way that the panpsychist formulation failed to do so above.

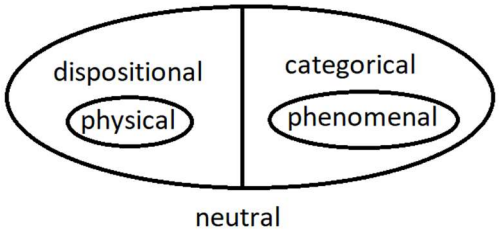
The problem with Categoricalist Russellian Monism seems to be its commitment to the idea that the fundamental properties are categorical.[40] Perhaps if we remove this commitment from Russellian monism, it may be able to avoid the above objections. As a first attempt, let's suppose the opposite is true. That is, the fundamental properties may all be dispositional, and thus the neutral properties would also be dispositional. This view is called dispositionalism and, combined with Russellian monism, would yield a Dispositionalist Russellian Monism. Dispositionalism is, traditionally, a less popular view than categoricalism, but it has been gaining popularity in the past couple of decades[41]. However, making this a part of Russellian monism will again fail to avoid the general dilemma in much the same way that the categoricalist formulations did above. If the dispositional is fundamental and is identified with the physical, then we have a straightforward failure to meet the Explanatory Demand. Thus, this view will be faced with all the attendant difficulties that plague materialism. If, on the other hand, the dispositional is instead divided into physical and non-physical properties, and the fundamental properties are all dispositions, then we fail to meet the Ontological Demand for the same reasons that the second formulation of Categoricalist Russellian Monism did. Finally, if the fundamental properties are these non-physical dispositions, then we have a position that, like the third variation of Categorical Russellian Monism canvassed above, simply seems unmotivated. The view would do nothing to resolve the problems faced by materialism since those problems arise from being unable to explain the relationship between the physical and phenomenal while simultaneously showing how the physical is more fundamental. This version of Dispositional Russellian

[40] As an historical aside, this view of the relationship between categorical and dispositional properties was highly popular among analytic philosophers around the time Ayer and Stace were writing and is still referred to as the "standard" position in the dispositions debate. So, it's possible that they were taking the standard position on this debate and applying it to Russell's monism. If this is true, then Russell is right when he responds to their objections by saying that they are misinterpreting the view. See (Lockwood, 1981) for a strong argument that something like this is the case. On the other hand, categoricalism was also popular when Russell was writing, and if he took this position, then his view is in fact susceptible to Ayer's and Stace's objection.

[41] This view is alternatively called "dispositional essentialism" or just "essentialism". See (Bird, 2007; Ellis, 2001, 2002, 2005; McKitrick, 2003; Mumford, 2006) for recent detailed defenses of this view.

Monism would have the fundamental properties be non-physical dispositions, followed by the physical dispositions, and the categorical properties (including the phenomenal) would be the least fundamental of the three. So, this view would fail to meet the Explanatory Demand.

Since the Russellian position maintains that the phenomenal are (a subset of the) categorical and the physical are (a subset of the) dispositional, the only option to avoid the line of reasoning of the preceding several paragraphs seems to be to take the properties that are neutral between physical and phenomenal properties as *also* neutral between the categorical and dispositional properties. In the dispositions literature, the view that the fundamental properties are neutral between dispositions and categorical properties has come to be known as the *powerful qualities view*.[42] This seems like a step in the right direction, but then we are faced with the question of how to understand what *this* neutral property is and how it is related to dispositional and categorical properties. One interpretation given by D.M. Armstrong is that the neutral has two "sides" or "parts" (1996, p. 250). This interpretation makes sense of C.B. Martin's (1993) suggestion that his neutral properties be considered a "two-sided coin" where one is dispositional and the other categorical.[43] However, this Dual Aspects Formulation falls to the second horn of the main dilemma by failing to meet the Ontological Demand. Perhaps there is only one neutral property, but nonetheless by admitting that the physical cannot be reduced to the phenomenal and vice versa, the physical and phenomenal are distinct types of properties of the neutral entities. And either the causal closure of the physical is violated, or it is not true that causal interaction is universal. Thus, this Dual Aspects Formulation fails to be sufficiently in between dualism and materialism.

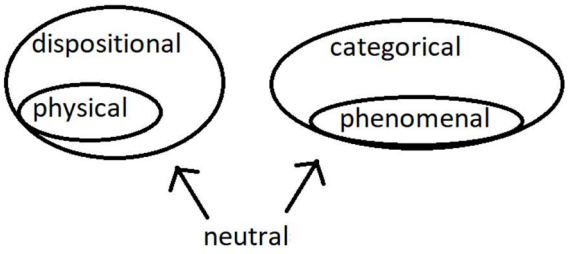


Of course, there is an obvious way to avoid this problem, by saying that the categorical and dispositional both reduce to the neutral. This Reductionist Russellian Monism seems better suited to the

[42] See (Coates, 2020; Heil, 2010; Ingthorsson, 2013; Jacobs, 2011; Martin & Heil, 1999; H. Taylor, 2018) for discussions of this view.

[43] Well, Martin and other powerful qualities proponents speak in terms of dispositions and qualities, not dispositions and categorical properties, but clearly, they are developing some form of neutral monism in the dispositions debate. So, I don't think that difference in terminology matters much for my purposes here.

task of meeting both the causal and the epistemic considerations.[44] We have many types of non-fundamental properties without serious ontological concerns about how they fit together. If all properties reduce to one neutral property type, then there are no concerns about how they interact (and any violation of the causal closure of the physical would be nominal) because they are merely different variations of neutral properties. Thus, the division between phenomenal and physical properties is not problematic.[45] The difficulties here arise because of the commitment to reduction. As I'll now argue, if one were in a position to defend the needed reductions, then one would have already accepted one of the various versions of reductionist physicalism already developed. To develop a Reductionist Russellian Monism would therefore be a little strange, unless it could help avoid some problem that reductionist physicalism runs into. This does not seem (to me) to be the case, and I'll briefly state why.



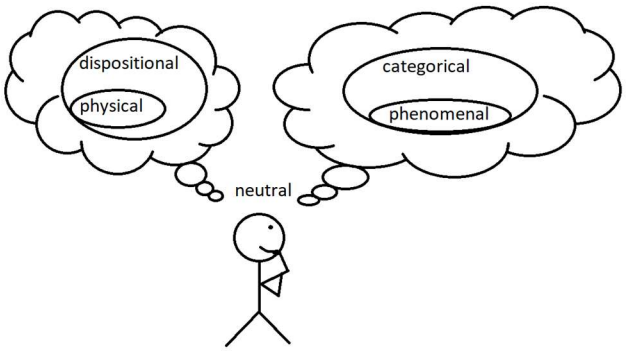
This is a difficult issue to canvas briefly because there are several notions of 'reduction' that might be in use here. A classic notion is that of intertheoretic reduction used by Ernest Nagel (1961) which makes use of the concept of bridge laws. Such laws bridge the phenomenal predicates to physical predicates in the following way. For any phenomenal predicate M there is a physical (i.e. neurological) predicate P such that the sentence '*x* is M if and only if *x* is P' is true. The difficulties with bridge laws are too numerous to get into here, but many of them center around the fact that multiple different physical predicates seem to correspond to the same phenomenal predicates – i.e. there are physical predicates P_1 (in humans), P_2 (in dogs), P_3 (in octopi), etc. which all correspond to the same phenomenal predicate 'pain'. This would be problematic enough, but the difficulty with a neutral monist version of a bridge law is that it would require neutral predicates. It's not clear what such predicates could be, but in the vein of Smart (1959), one might suggest that topic-neutral predicates could be deployed. However, such a move will not resolve the basic problem of multiple realizability. If these topic-neutral predicates hold a one-to-one relation with the physical predicates, then there would still be a many-to-one relation with the phenomenal predicates. On

[44] Chalmers' panprotopsychist variation of this view would add that the neutral properties are also categorical. This has the benefit of giving us a solid grasp of the neutral, but it threatens to collapse into one of the variations of Categoricalist Russellian Monism discussed above.

[45] See Leopold Stubenberg's Stanford Encyclopedia of Philosophy article on Neutral Monism, and especially the supplement to that article "Reducing Mind and Matter to Neutral Entities" (2016) for some ways in which this reduction can be cashed out.

the other hand, if the topic-neutral predicates held a one-to-one relation with the phenomenal predicates, then there would be a one-to-many relation with the physical predicates. So, it is hard to see how there could be the needed reduction of *both* to some neutral third type of property.

A final formulation is suggested by Michael Lockwood's (1981) understanding of Russell's view as well as by Martin and Heil's (1999) rejection of the two-sided coin analogy for their powerful qualities view of dispositions – they agreed that it gave license to too much of a dualistic interpretation. On this formulation, as Lockwood puts it, the "phenomenal-physical dichotomy is on Russell's view ontologically illusory – a reflection, merely, of the epistemological distinction between immediate acquaintance or introspection, on the one hand, and ordinary perception or scientific inference, on the other" (1981, p. 157). This amounts to an identity between the physical dispositions of the brain and phenomenal categorical properties of the brain.[46] The physical and phenomenal qualities just *are* the neutral stuff in different guises. I call this view Epistemic Neutral Monism. A different analogy might make the view a bit clearer. Imagine a holographic baseball card (the kind that were the best to get as a kid). Viewed from one angle, you see Ken Griffey Jr. with his hat backwards, smiling into the camera, but view it from another angle and you'll see Ken Griffey Jr. swinging the bat in a beautiful arc, hitting a homerun as he did so many times. The baseball card is neither one of Ken Griffey Jr. smiling nor of him swinging, but rather it is neutral between the two. The two "sides" have nothing to do with the baseball card changing. It is merely the viewer's perspective which changes and gives the appearance of two different cards, but they are the same card. This is the way that I viewed them as a kid anyhow. Of course, if we reflect more deeply on *why* the card has two appearances from different angles, we'll discover that it's because it has different parts. It's not merely a difference in the viewer of the card, but also a difference in the card itself. So, the analogy is *supposed* to illustrate the idea that the dispositional/physical and categorical/phenomenal are merely different appearances based on our epistemic perspective, but it may turn out that whenever there



[46] This view of the relationship between dispositions and categorical properties is also found in (Mumford, 1998). He calls his view 'neutral monism' rather than the powerful qualities view.

is such a potential difference, there has to be some different part in the property itself.[47] This nicely illustrates a central worry with this view, it's not clear if it can be sustained as different from the Dual Aspects Formulation above, which could not meet the Ontological Demand. It is easy to see why the Russellian monist might try for this epistemic approach. One way of knowing the world is via the causal interaction which things enter into – the scientific, or physical way of knowing – while the other way of knowing the world is via the intrinsic natures of things, to which we only have very limited access – the mental, or phenomenal way of knowing. However, by pinning the physical to dispositional properties and the phenomenal to categorical properties, it is hard to see how this merely epistemic trait of the view can be sustained. Nonetheless, this epistemic formulation shares the benefit of the reductionist formulation in that it provides a solution to the mind-body problem that might be properly considered neutral monist.[48]

So, in contrast to the other forms of Russellian monism, both Reductionist Russellian Monism and Epistemic Russellian Monism have the potential to be coherent and stable positions that could avoid both horns of the dilemma. As such, if it is to succeed in its central task of solving the mind-body problem while capturing the considerations behind materialism and dualism, any Russellian monist solution must be one of these two formulations. This would be a great starting point if that were the end of the story. Unfortunately for Russellian monism, there are problems that even these formulations cannot escape.

4. Problems with Any Formulation of Russellian Monism

In "Pessimism about Russellian Monism" (2015), Amy Kind argues that Russellian monism is incapable of reconciling the divide between materialists and dualists. Her analysis is that Russellian

[47] This same point holds for Martin and Heil's preferred analogies of the Necker cube and the duck-rabbit – the *whole* image's appearance changes because we look at different *parts* of it just as the *whole* image of the baseball card changes because we look at different *parts* of it. One plausible reason for this difficulty is that it is a defect of using common physical objects and drawings to make the analogy. I don't have the room to elaborate much on this thought, but common physical objects and drawings are always easily broken down into constituent parts. This makes it too easy to build that into the analogy, but it may not be the case for the neutral stuff that can be thought of categorically and/or dispositionally. These analogies have the benefit of being clearly understood though. Other analogies, like Michael Lockwood's (1989) analogy of particle/wave duality, are possibly better suited to the task of explaining the difference in epistemic perspective, but they have the drawback of being at least as obscure as the subject matter (at least to readers who aren't quantum physicists). So, such analogies aren't helpful in clarifying what the position is trying to say.

[48] This sort of merely epistemic difference is similar to what is posited by a materialist who says that phenomenal concepts don't reduce to physical concepts. Thus, there is an epistemic difference without a metaphysical difference. Indeed, the same goes for categoricalists and dispositionalists that say that the opposing class doesn't reduce to whichever they say is fundamental. The difference in this epistemic formulation of neutral monism is that there is no preferred class of concepts. Thus, the property that is in the world is neither physical nor mental, though physical and mental concepts may correctly refer to it. The materialist must show why even though the concepts refer to the same property, the property *really is* physical rather than mental. Epistemic Neutral Monism needn't do this.

monism would shift the mind-body debate from the impasse between materialism and dualism to one between panpsychism and panprotopsychism. This obviously doesn't reconcile much. More importantly, she argues that the same considerations that push people toward dualism will push people toward panpsychism and the same considerations that push people toward materialism will push people toward panprotopsychism.[49] I'm not sure that this claim is true given that the most vocal proponent of panprotopsychism is David Chalmers who has made it clear that given the option between only dualism and materialism, he favors dualism. But suppose it were true that, given the option between panpsychism or panprotopsychism, the same philosophers who find materialism attractive would choose panprotopsychism while those who prefer dualism would choose panpsychism. If she gets this claim right, she has proven that Russellian monism cannot reconcile the debate between dualists and materialists. Nonetheless, the shift to Russellian monism has the ability to change the structure of views and arguments on offer, so it would be a significant philosophical move forward if it is right.

I understand Kind as arguing that any form of neutral monism will end up falling to the general dilemma presented in Section 3. But although I think Kind is right to be pessimistic about the prospects for *Russellian* monism to resolve the impasse between materialists and dualists, I think she's wrong to be pessimistic about the prospects for formulating a neutral monism that captures the best of materialism and dualism. My dispute with Russellian monism is not that I think any neutral monist view is doomed to fail. Rather, my problem is that no matter what formulation of Russellian monism we choose, the insight that makes it *Russellian* simply doesn't do any work in finding a solution to the mind-body problem that meets the Explanatory and Ontological Demands. This has to do with the way Russellian monists try to formulate their solution by making the mind-body problem a special case of the dispute about the relationship between dispositions and categorical properties.

To begin, the dialectical problems with the Russellian monist solution make it an unattractive way to argue for neutral monism. The insight that all physical properties are dispositional takes for granted substantial claims about the project of physics as a science and proclaims that this applies to all physical

[49] It should be noted that Kind seems to only be considering Categoricalist Russellian Monism as her issue is with whether the *intrinsic* properties are phenomenal or not. And since it is these intrinsic properties which reveal what the fundamental properties are for her, she seems to be taking categoricalism for granted. This may influence her thinking about the ways materialists and dualists will divide themselves given the option between panpsychism and panprotopsychism. If this is right, then Kind's argument is not only problematic for overgeneralizing, as I say in the main text, but also misses some interpretations of Russellian monism.

properties. This move from what physics can tell us about the world to a claim about what kinds of physical properties exist is contentious.[50] Moreover, we should remember all of the assumptions discussed in Section 2 about dispositions and categorical properties that we must accept. When discussing those, I said that they are plausible enough, but that is far from saying that they are without problems of their own. In trying to solve a philosophical problem, it would be best to avoid merely exchanging one problem for another. However, this exchange would be acceptable if it were to meet two conditions. First, this exchange should be necessary for making progress on the target, in this case the mind-body problem. If we can make the same (or similar enough) argumentative moves without taking on board extra philosophical problems, then there is no motivation for taking these extra problems on. Second, the exchanged problem should be no more intractable than the target problem. If the problems are equally difficult to solve, then we've gained no ground. We've simply changed the subject. In the case of Russellian monism, the idea is to argue that the phenomenal/physical divide is a subset of the categorical/dispositional divide, and then solve the problem of how dispositions and categorical properties are related, thereby also solving the mind-body problem. For someone inclined toward neutral monism, the mind-body problem has proven itself to be particularly intractable, so one may think that this is progress. However, as I'll show here, this exchange meets neither condition. So, we'd be better off trying to solve the mind-body problem on its own terms rather than exchanging it for the dispositions problem.

For the first condition, it is clear that the two types of Russellian monism – Reductionist and Epistemic – that are best suited to avoiding the general dilemma faced by any neutral monism can be just as easily formulated without any reference to categorical or dispositional properties. This can be seen by looking at the diagrams above and noticing that the larger circles that denote the full set of categorical and dispositional properties are entirely redundant in those two variations of Russellian monism. If a reductionist or epistemic relation from some neutral fundamental type of property to two non-fundamental property types is able to avoid the general dilemma, then the further detail about whether the phenomenal are categorical and the physical are dispositional is doing no work. We can just as easily argue for a Reductionist or Epistemic Neutral Monism that makes no reference to categorical or dispositional

[50] See (Stoljar, 2001) for a clear argument that the theories of physics do not tell us about all physical properties. This claim is also argued against by categoricalist physicalists who nonetheless agree that the sciences can only tell us about dispositions.

properties as we can for the Russellian version of the view that makes essential reference to these properties. The general point for both of these attempts to craft a neutral monism is that there is some property that is neutral between the phenomenal and the physical which can help us incorporate the dualist's Explanatory Demand that there should be no epistemic gap between two things where one necessitates the other while also incorporating the materialist's Ontological Demand that everything can, in principle, enter into causal relations with everything else. Nothing about this property requires any reference to the Russellian insight about physical properties being dispositional or the speculation that phenomenal properties are their categorical correlates. Thus, the Russellian move is not necessary for the best options available to a neutral monist.

Showing that the Russellian move does not meet the second condition requires recognizing that the principal considerations at play in the mind-body problem also apply in the debates about dispositions, and in very similar, if not in exactly the same ways. The idea that all properties can, in principle, enter into causal interactions with one another drives the debates between categoricalists and dispositionalists. For the most part, each side accepts this general view, but argues that their preferred property type is the one that is the basis for the other. On the other hand, both sides attempt to show that there isn't the type of an epistemic gap between the two property types as there would need to be to justify saying that they are both fundamental types of property in the world. Meanwhile, there are those who argue for a dispositional/categorical dualism on the basis of the existence of an epistemic gap between dispositions and categorical properties.[51] And, of course, we've already mentioned the powerful qualities view that tries to accommodate both considerations.

Since the same considerations apply to the dispositions problem, we may think it would be a good thing to try to combine the problems as Russellian monism does. However, because the same considerations apply, either the problem will be just as intractable as the mind-body problem, or it will be more easily solvable only due to some serious differences in the types of properties involved. So, purely on dialectical grounds, the Russellian move toward discussing dispositions and categorical properties in order to solve the mind-body problem is not as helpful as it may have first appeared.

[51] See (Molnar, 2003; Prior, 1985) for two prominent examples.

There is a further, more philosophically important, problem with the Russellian insight's ability to help solve the mind-body problem. Even if Russellians are correct in making the physical a subset of the dispositional and the categorical a subset of the phenomenal, the phenomenal is a *special* subtype of categorical property. In a way, this is precisely the issue that nags the debate between panpsychist Russellian monists and panprotopsychist Russellian monists. The panprotopsychists point out that the phenomenal properties you and I experience are definitely different from the categorical correlates of the physical properties of things like rocks and electrons. As I see it, panpsychists have no recourse but to accept this. The physical structure of our brains is quite different from the physical structure of a rock or electron, and if phenomenal properties are just the categorical correlates of these dispositional/physical structures, then it would be absurd to think that the exact same categorical properties could be directly correlated with such vastly different physical structures. So, even if the panpsychists are right to think that the categorical properties of rocks and electrons are phenomenal, there is a major problem of how those quite different phenomenal properties are related to the phenomenal properties that are more familiar to us (i.e. our phenomenal properties).[52]

Moreover, explaining (away) the fact that our phenomenal experiences are *different* is precisely what the mind-body problem is about. This special nature of our phenomenal properties makes Russellian monism not only unnecessarily more complex, but also shows that shifting the discussion to the dispositions debate simply won't help. Even if phenomenal properties are categorical and physical properties are dispositional, the fact that *our* phenomenal properties are a special subtype of categorical properties means that the mind-body problem will simply be regenerated. The big hope of Russellian monism is that one can show that the mind-body problem is a special case of the more general dispositions problem, solve this latter problem, and thereby say that the mind-body problem has been solved. However, the special nature of our phenomenal properties guarantees that the mind-body problem will still pose difficulties.

The point of all this is to say that if we want to develop a solution to the mind-body problem that is neutral with respect to what it posits is the fundamental property type of the world, shifting the discussion to debates about dispositions and categorical properties is a distraction. We'd be better off focusing on the

[52] A problem of this kind often goes under the name "The Combination Problem". For discussion of this problem, see (Chalmers, 2016; Goff, 2017; James, 1890; Lockwood, 1993; Seager, 1995).

peculiarities of the phenomenal and physical properties rather than trying to rope in dispositions and categorical properties. On the other hand, I'm not prepared to argue that there is *no* relationship between these two problems. But the relationship seems to be one of sibling problems rather a genus/species relationship. That is, both problems concern how two deeply different (at least on first appearance) types of property are related to one another and trying to figure out which is the fundamental property type, if either. The two problems are linked by the principal considerations in favor of the two standard views – that is, the idea that all things in the world can, in principle, enter into causal interactions with one another without excess overdetermination, but also the idea that if one type of thing necessitates the other, then no epistemic gap should exist between them. In both problems, these considerations, in combination with further facts, pull us in two directions – toward monism on the one hand and toward dualism on the other. This leads us to the impasse in each debate. Both sides have strong arguments because both sides make use of general considerations that are highly plausible. But that doesn't mean that the solution to one problem has to be the same as the solution to the other problem. Even though the underlying principal considerations are generally applicable, in each case, it is possible that some peculiarity about the two property pairs in question should lead us to believe that one or both considerations don't apply. Or, at the very least, that in one problem the further facts show that one argument is stronger than the other.

As such, I think that the general strategy of the Russellian monists of hitching their wagon so strongly to the dispositions debate is misguided. Of course, some solution to the general problem of which principal consideration is stronger (if indeed, either is, in the general case) *will* have some bearing on both problems. However, due to the variation in the property types involved, it won't completely answer either problem straightaway. That doesn't mean that nothing can be learned about the mind-body problem from debates about dispositions, but just that the solutions to the problem of dispositions aren't so easily mapped to the mind-body problem. If we want a genuine solution to the mind-body problem, we'd be better off looking directly at the phenomenal properties and the physical properties and whether it is best to make use of the metaphysical considerations of dualists, materialists, or both.

Chapter 4: Toward a Non-Russellian Neutral Monism

> "Usually, when we are told that X is Y we know how it is supposed to be true, but that depends on a conceptual or theoretical background and is not conveyed by the "is" alone. We know how both "X" and "Y" refer, and the kinds of things to which they refer, and we have a rough idea how the two referential paths might converge on a single thing, be it an object, a person, a process, an event, or whatever. But when the two terms of the identification are very disparate it may not be so clear how it could be true. We may not have even a rough idea of how the two referential paths could converge, or what kind of things they might converge on, and a theoretical framework may have to be supplied to enable us to understand this. Without the framework, an air of mysticism surrounds the identification."
>
> - Nagel, "What is it Like to Be a Bat?" (1974, p. 447)

1. A Path Forward

A proper neutral monist solution will have to avoid the dilemma where on the one hand opponents seek to show that the view is incoherent and on the other hand, they seek to show that the view fails to meet one or more of the demands that motivate materialism and dualism. As I argued at the end of the previous chapter, Reductionist Russellian Monism and Epistemic Russellian Monism can both be developed without accepting the Russellian claim that the physical is dispositional and the phenomenal is categorical. Since that claim does nothing to move us toward a solution to the mind-body problem, anyone seeking to develop a neutral monist solution would be best served by keeping it out. Nonetheless, both formulations show us a model for developing a neutral monist solution to the mind-body problem that has the chance of avoiding the dilemma. The idea of offering a fourth type of solution to the mind-body problem is a matter of fulfilling the goal of capturing the best parts of materialism and dualism while avoiding their problems. The best chance of doing so lies not with Russellian monism, but rather with some form of either a reductionist neutral monism or an epistemic neutral monism. In attempting to develop a non-Russellian neutral monism that can meet the demands on a satisfactory solution to the mind-body problem, I will first show why I think that the reductionist option is a dead end. Then, I'll discuss some of the pitfalls that an epistemic neutral monism will have avoid if it is to have a chance at succeeding.

The problem with Reductionist Neutral Monism is that even if we can develop a version that successfully avoids the general dilemma against neutral monism, it will face a further dilemma I think is inescapable. The dilemma is as follows. Either the phenomenal can be reduced or it cannot. If it cannot, then obviously Reductionist Neutral Monism is false. If it can, then all of the tools developed to show that the phenomenal can be reduced to some other property type will be available to a reductionist materialist. The problems for reduction don't depend on the reduction base being physical, but on the epistemic gap

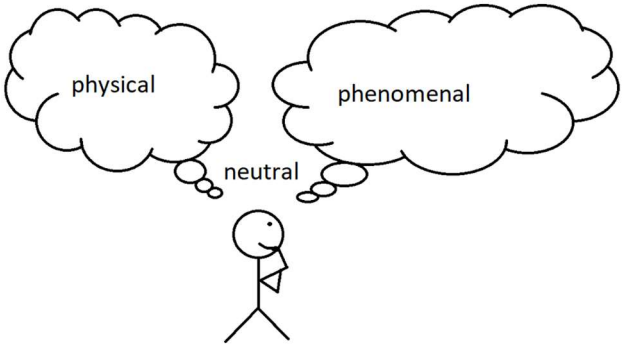
between it and the phenomenal. And this will be present for any neutral base as well. Of course, this is necessarily brief and speculative. I do not intend this dilemma to be a knock down argument against any future formulation of Reductionist Neutral Monism. Instead, I mean it to do two things. First, it provides an important constraint on the pursuit of reductionism by neutral monists. They will have to not only show how to reduce phenomenal properties (and physical properties) to the neutral but also show why it only works for reducing the phenomenal to the neutral and not to the physical. Second, since I see no way to do this, it provides me a reason to forego exploring Reductionist Neutral Monism and instead pursue the option that I believe has better hopes of ultimately working. If the reader disagrees, of course, I'd be very interested to see how a Reductionist Neutral Monism might be developed that avoids both this dilemma and the general dilemma that faces any neutral monism.

In order to develop a neutral monist solution to the mind-body problem, we'll need to continue to keep two criteria clearly in sight. First, any neutral monist will have to appropriately meet both the Explanatory Demand and the Ontological Demand. Meeting both of these demands seems to require one type of property and some way of understanding the phenomenal and physical such that they can both be that property type. Materialists fail to adequately explain the epistemic difference between the physical and phenomenal while dualists fail to explain the causal significance of the phenomenal. Neutral monism will have to succeed in both.

Second, a properly neutral monist solution will have to avoid the dilemma where on the one hand opponents seek to show that the view is incoherent and on the other hand they seek to show that the view is merely a repackaging of one of the three more well-known solutions to the mind-body problem – dualism, materialism, or idealism. This second criterion is really a direct result of trying to meet the first. Meeting both demands is ordinarily presented as something that is impossible to do. We are meant to choose between one or the other, and so trying to meet both at the same time is met with the suspicion that the view on offer is incoherent. Nevertheless, it is obvious that a solution that meets the two demands is not logically inconsistent. Thus, the logical space is available for a neutral monist view, at least in theory. In practice, though, each neutral monist solution, once it is accepted as coherent, seems to fall on the other horn of the dilemma; it is not really a fourth type of solution, but only a rehashing of one of the more familiar three. This is also closely tied to the first criterion in that a view that meets the Ontological Demand and not the Explanatory Demand is materialist, while a view that meets the Explanatory Demand and not the

Ontological Demand is dualist. And a view that meets neither is simply unmotivated. So, the idea of offering a fourth type of solution to the mind-body problem isn't merely a matter of offering something novel for novelty's sake, but rather it is a matter of fulfilling the goal of meeting both demands of materialists and dualists, thus creating a solution to the mind-body problem that inherits the best parts of both of these views while avoiding their problems.

Due to the arguments up to this point, Epistemic Neutral Monism is left as the last hope for neutral monism in the effort to accommodate both demands. In this formulation, the physical/phenomenal distinction is merely epistemic. We have two different ways of epistemically "carving" the world, but nevertheless, ontologically, there is only one neutral type of stuff. Thus, the physical and phenomenal are just neutral stuff understood differently. This promises to meet the Ontological Demand that everything in the world be able to interact with everything else in the world because everything is really just the same type of thing, neutral. And it also promises to meet the Explanatory Demand that the supposed distinctness of the phenomenal be explained. It can't be explained in physical terms, so materialism fails, but the phenomenal is nevertheless not a distinct type of property because it is merely a different way of understanding the neutral property type in the world. Of course, merely saying that the difference between the physical and the phenomenal is just epistemic does not make it so. There are many questions about how this might work. One obvious worry is that we have two different understandings because there are two different types of thing. This is the dualist's approach, but then that runs directly up against the Ontological Demand. So, to meet both, we must say something different. The remainder of this chapter will be an exploration of what must be said and what must be avoided in order for such an approach to work. And then in Chapters 5 and 6, I'll attempt to do exactly that.



2. Epistemic Neutral Monism

The basic statement of Epistemic Neutral Monism is that phenomenal concepts and physical concepts both refer to the same thing, neutral properties and/or objects. They are two conceptual schemes that refer to the same stuff out in the world. To those familiar with it, this will sound strikingly close to *a*

posteriori physicalism (or *a posteriori* identity theory, or type-B materialism, depending on the preferred name of the theory). Indeed, the basic stance is different in only one respect. Both theories agree that there are two conceptual schemes that are not *a priori* identical, but nonetheless they both refer to the same thing in the world. However, the materialist maintains that what is out in the world is physical and therefore the physical conceptual scheme tracks the truth of the matter more directly than the phenomenal conceptual scheme. On the other hand, the neutral monist says that neither conceptual scheme is to be so preferred because what both conceptual schemes refer to is stuff that is neutral between the two.

Two analogies will help clarify the difference between these two very similar views. Take the standard view that water is identical to H_2O. This is an *a posteriori* identity because conceptually water and H_2O come apart, at least in different possible worlds. One is a term within a conceptual scheme based on everyday human interactions while the other is a term within a conceptual scheme based on the underlying chemical microstructure, which presents a more fundamental way of understanding what is in the world. This is most similar to the *a posteriori* physicalist view because the chemical concept more directly tracks what is fundamental out in the world. On the other hand, take the standard view of Hesperus and Phosphorus. Each of these is introduced via a different interaction with, or perspective on, Venus; one using the conceptual scheme of the evening while the other uses the conceptual scheme of the morning (in a loose sense of "conceptual scheme"). Neither more directly refers to the object that revolves around the Sun at a closer orbit than the Earth. They both refer to an object that is neutral between the two conceptual schemes. This is directly analogous to the neutral monist's position.

This raises an immediate issue concerning the relationship between a concept and its referent.[53] For my purposes here, the relevant referents are properties, kinds, types, etc. rather than simply things that fall in the extension of the term. As I see it, the issue that I'm raising applies equally to both *a posteriori* physicalism and epistemic neutral monism. Understanding this issue requires distinguishing between concepts that refer via essential properties and concepts that refer via contingent properties of their referent. Referring via an essential property means that the concept cannot possibly refer to any other referent than the one that it refers to. The essential property of the referent is necessarily attached to the

[53] Much of the following discussion is indebted to (Woodward, 2018). His description is in terms of connotation and denotation, but I've put in Kripke's terms as I find them better suited to the task. Of course, one might put it in Chalmers' terms of his two dimensional sematics or, following (Goff, 2017) and others, put it in terms of transparent and opaque concepts. Nothing hinges on the terms.

referent, so any concept that refers via essential properties will necessarily refer to that same referent every time. On the other hand, referring via a contingent property means that the referent of the concept could very well be something other than the actual referent. In the case in of water and H_2O, the concept "water" refers to the stuff that falls from the sky, sustains life, turns to ice when below zero Celsius, etc. These are all properties that are contingent to the stuff that "water" actually refers to. It is possible that the stuff that is associated with all of these properties in the actual world could be associated with some other kind of thing. So, the concept "water" refers via contingent properties of its referent. On the other hand, the concept "H_2O" refers via an essential property of its referent. There is no way that "H_2O" could refer to anything that isn't what it actually refers to. This is why the analogy works for materialism. One concept directly refers to the essential nature of the object out in the world while the other refers via a contingent property of that object.

In the case of "Hesperus" and "Phosphorus", both concepts denote Venus, but one refers to the object with the property "is the first star visible in the evening sky" while the other refers to the object with the property "is the last star visible in the morning sky". Both of these properties are contingent to Venus, so both "Hesperus" and "Phosphorus" refer via contingent properties. This explains why it is not *a priori* that these identities are, in fact, identities. At least one concept in the equation of each identity refers via contingent properties of its referent. Nonetheless, in each case, there is an identity because they refer to the same thing – water=H_2O and Hesperus=Phosphorus. This is also what makes Hesperus and Phosphorus an apt analogy for neutral monism; both concepts refer to something via contingent properties.[54]

The issue for our purposes is that it seems as if all *a posteriori* identities involve at least one concept that refers via contingent properties, and we seem to have good reason to think that *both* phenomenal *and* physical concepts refer via essential properties. The physical concepts, at least, seem to refer via essential properties, if we're to follow the example of H_2O and other microstructural properties. So, it would seem a real problem for both materialism and neutral monism if phenomenal concepts also referred via essential properties. Of course, materialists have thought otherwise in the past. The original *a posteriori* materialists of the 1950's – e.g. J.J.C. Smart – thought of phenomenal concepts as referring via contingent properties.

---

[54] You might have recognized that the analogy isn't perfect. Water and H_2O are kinds and Venus is an individual. So, these are all objects whereas phenomenal and physical properties are, well, properties. We'll explore this issue more in later chapters.

This would mean that "pain" and "c-fibers firing" (or whatever is going on physically when one experiences pain) have the same referent but while "c-fibers firing" refers essentially to the physical stuff in the brain, "pain" refers to the physical stuff happening in the brain via the contingent property of experiencing pain. However, it is difficult to argue with the idea, presented vividly in Kripke's (1980) *Naming and Necessity*, that pain is essentially painful. That is, that the referent of "pain" just is the experience of pain. It does not refer to whatever neuroscientists tell me is going on in my brain. I know whether my use of "pain" refers by my experience of pain. At least, normal users of the concept "pain" take themselves to be referring to the experience of pain when using the concept, not something happening in the brain, they know not what. This is different from the case of water. Users of the concept "water" and other concepts that refer via contingent properties are fully willing to accept the idea that they are referring to some underlying structure they know not what (or, if they know the empirical work, they *do* know that it is H_2O). If this is right, then at least one phenomenal concept refers essentially with no reason to suppose that it is unique amongst phenomenal concepts. And thus, the supposed *a posteriori* identity between phenomenal and physical properties doesn't hold.

One solution to this problem offered by materialists[55], is to argue that phenomenal concepts refer via neither essential nor contingent properties but are rather more similar to demonstratives like "I" or "now". These concepts (the theory goes) refer without using any property or set of properties. That is, when I refer to myself as "I", there is no additional way of thinking of my referent; it is more like me linguistically pointing to myself rather than referring by use of properties. If phenomenal concepts refer without the use of any property at all, then the possibility of an *a posteriori* identity is back on the table. The more apt analogy would be the *a posteriori* identity between that stuff (where it's clear what the referent is by context) and H_2O.[56] Both "that stuff" and "H_2O" refer to H_2O, but "that stuff" does not do so through any property at all. It refers via linguistic pointing (perhaps along with physical pointing). Meanwhile, again "H_2O" refers to H_2O, and so refers via the referent's essential property. This allows for the possibility of the subject not knowing *a priori* that the referents of the two concepts are the same because although users of the concept

[55] Amongst them are (Levin, 2007; Loar, 1997; McLaughlin, 2001; Perry, 2001).

[56] This example is adapted from John Perry's (1979) example of a man wondering who is spilling sugar on the floor only to realize that it is he who is the one spilling sugar. Perry's example presents an interesting case where an indexical has the same referent as a term that refers via contingent properties. But this is a disanalogy for the materialist case at hand. Thus, the adaptation.

"that stuff" know that they are referring to something specific, they may not know exactly what it is. They may not even know that it is water. Only after some empirical investigation could they discover that it is H_2O.

As with concepts that refer via contingent properties, the idea that phenomenal concepts refer via linguistic pointing seems problematic. Normal users of "pain" don't take themselves to be pointing at some experience of theirs without any understanding of what it is. They understand "pain" to be the experience of pain. Moreover, if phenomenal concepts did refer via linguistic pointing, any comparisons between different phenomenal experiences would no longer make much sense. Take the judgement that pleasure is better than pain. This would amount to saying "this is better than that" where "this" refers to some brain activity and "that" refers to some other brain activity without any reference to the experience. But it is the *experience* of pleasure that one means is preferable to the experience of pain, not the brain activity. This reveals that phenomenal concepts do refer via at least some property, an experiential property.[57] But of course they refer via at least some property if Kripke is right that phenomenal concepts refer via essential properties.

Another possibility is to accept that phenomenal concepts refer essentially, but to argue that they are somehow special such that they do not reveal their actual referent to the subject using them. Given that the *a posteriori* materialists want to say that phenomenal concepts are unique amongst those concepts that refer via essential properties, they will have to give an account of how they are different such that normal users remain unaware of phenomenal concepts' referents despite the fact that they refer to those referents via the properties that are essential to them. Without an adequate explanation of why phenomenal concepts work this way, and no others, this strategy will be an ad hoc explanation devised merely to save *a posteriori* materialism.

My goal here isn't to discuss the merits or demerits of each of the proposals that follow this strategy.[58] Rather, there is a general problem made clear by Woodward (2018). The problem is that if a

[57] Goff (2011) makes a similar point that phenomenal properties are not, as he puts it, "radically opaque". Also, see (Demircioglu, 2013; Horgan & Tienson, 2001) for an argument that the indexical view of phenomenal concepts would have the implausible result that normal perceivers and blindsighted perceivers would think of their experiences in the same way. See (Holman, 2013; Levin, 2007; Loar, 1997) for responses.

[58] See (Woodward, 2018, pp. 136–141) for such a discussion. His reason for rejecting the phenomenal concepts strategy is intimately tied to the problem of strong necessities discussed in the main text. His "dialectical bad news" is one of the

concept refers via something essential to its referent, in the normal case, we know what that referent is. So, if phenomenal concepts worked like other concepts, when we say that 'pain' refers to something that is essentially painful, we would know that it refers to a phenomenal property of experience, not to anything necessarily physical. But the phenomenal concepts strategy claims that phenomenal concepts work differently. Something obscures the reference of the concept from us even though we are correct about the fact that the concept refers via the very thing that makes the referent what it is. Thus, the concept 'pain' could really be referring to something physical rather than something phenomenal. But this purported fact about phenomenal concepts would also make it impossible for us to explain away the apparent distinctness of phenomenal and physical properties because we would have no access to the actual reference of phenomenal concepts. Without knowing what phenomenal concepts refer to, we can't show that their reference is necessarily connected to the physical. So, the *a posteriori* materialist is put in a dilemma to either accept the apparent distinctness of phenomenal and physical properties and so, be a dualist, or to insist that they are necessarily related despite no evidence that explains away their apparent distinctness. This is a problem for which the *a posteriori* physicalist has one final strategy, what Philip Goff (2017) calls the *dual carving strategy*. I'll discuss the details of this strategy below, but first I'll connect the previous discussion of *a posteriori* materialism with Epistemic Neutral Monism.

Thus far, we've discussed many of the strategies, and problems with them, that have been proposed to save *a posteriori* materialism from the problem that results when the apparent reference of physical and phenomenal concepts is different. But it may be a bit mysterious what all this talk has to do with Epistemic Neutral Monism, the purported topic at hand. The first point is to recognize the general strategy that the *a posteriori* materialist has taken. At each point, they accept that physical (scientific) concepts refer via essential properties and propose different ways in which phenomenal concepts don't refer essentially.[59]

---

same dialectical problems that strong necessities face; we must accept them without the existence of any conceivable evidence that would prove them true.

[59] An alternative way of understanding Russellian monism to my presentation in Chapter 3 is that Russellian monism accepts that phenomenal and physical terms refer via different kinds of properties while denying the materialist claim that physical terms refer essentially. Rather, they say that physical terms refer via contingent properties. One might think that this conceptualization of Russellian monism can avoid the problems discussed in Chapter 3. However, the same problems presented in Chapter 3 will arise when we try to think of whether phenomenal terms refer via essential, contingent, or no properties. They still seem to refer essentially, and if the Russellian monist is right that physical terms refer via contingent properties, then the view seems to fall to A.J. Ayer's and others' contention that the view is a veiled idealism. On the other hand, if phenomenal terms refer via contingent properties or linguistic pointing, then the view runs into difficulties discussed in the main text.

This has been dubbed the *phenomenal concepts strategy* by Daniel Stoljar (2006). As we have seen, the basic problem with the phenomenal concepts strategy writ large is that we have no reason to suppose that phenomenal concepts refer via anything but properties that are essential to their referents. Insofar as the materialists are pursuing a reasonable goal, they are doing so within the materialist paradigm which presupposes the truth of materialism. I don't mean that they are presupposing it without reason; they have very good reasons for believing the truth of materialism – the most important of which I discussed in Chapter 2. However, since we are not in that paradigm, we cannot presuppose the truth of materialism and so we need more in-depth reasons for supposing that language users are radically and systematically wrong about what their concepts are doing. Without presupposing the truth of materialism, we have no reason to suspect that anything is amiss with our phenomenal concepts. The epistemic formulation of neutral monism, however, claims that both physical and phenomenal concepts refer via the same kind of properties – both essential, both contingent, or both without using any property of their referent. The puzzle then arises which kind of concepts they are. The least likely option would be to argue that both phenomenal and physical concepts refer via no properties at all. This seems absurd on its face. It certainly *seems* like we can understand these concepts in some way or another, and so we'd be better off trying something else.

The most natural suggestion is that both concepts refer via contingent properties. This is indeed the suggestion given by the analogy with Hesperus and Phosphorus and seems to be the type of thing that the early neutral monists were thinking.[60] However, this runs into the same problem as the early *a posteriori* materialism of Smart; phenomenal concepts just seem to refer via essential properties. So, an Epistemic Neutral Monism that takes this approach is going to face the same style of argument that Reductionist Neutral Monism faced in the previous section. Any way of arguing that phenomenal concepts *do* refer via contingent properties will be available to the materialist at which point the motivation for pursuing the neutral monist strategy is gone. Furthermore, scientific physical concepts *also* seem to refer

[60] There is some sense to be made of this being equivalent to the reductionist neutral monism discussed above. In fact, it ends up falling to a very similar criticism that leveled at reductionist neutral monism. Between this footnote, the discussion in Chapter 3, and the previous footnote, it should become clear just how difficult it is to come up with an adequate classification scheme of the available options for a neutral monist view. There are many different possible ways of dividing things up, and while many of them are equivalent (or near enough), many of them are not. The reference to dispositions and categorical properties is sometimes made explicit, sometimes hides in the background, and sometimes is not there at all.

essentially, and so this approach will face the additional burden of proving that they refer via contingent properties.

Of course, if neutral monism is successful in both of these tasks, then it will show a flaw in *a posteriori* materialism. However, it will then face the difficulty of saying just what is this neutral property that is the referent of both types of concepts. Why don't we have a neutral conceptual scheme such that we could very easily understand what it is that our phenomenal and physical concepts are elliptically referring to? One might be a radical skeptic about this and say that it is impossible to know what the neutral properties are and thus we have no possible way of building such a conceptual scheme; this seems to be what Russell thought. However, this runs into a variation of the same problem that the *a posteriori* materialist eventually had to face. If there is such a neutral property type, we have no way of knowing it. And therefore, we have to accept that physical and phenomenal concepts refer to the same thing without any evidence for this fact. Given the fact that a neutral conceptual scheme seems necessary for us to obtain such evidence, one would think that we'd be trying much harder to develop this neutral conceptual scheme. Humans are quite good at developing new conceptual schemes, and since the interest is there, we should at least have some explanation for why the conceptual scheme has eluded us for so long. After so much thought and effort, we are led to think that the world is neutral, and we can't discover what that neutral type of thing is.

The final option is that both phenomenal and physical concepts refer via essential properties. This has the benefit of taking for granted the evidence of how we understand what is going on with both phenomenal and physical conceptual schemes. It also leads into a bit of a mess. Above, I mentioned the final strategy to save *a posteriori* materialism – the dual carving strategy. It also accepts that phenomenal and physical concepts refer via essential properties. So, by two different routes, we are led to the same position about phenomenal and physical concepts; both refer essentially. But this is *also* exactly what the dualist is arguing. So, everyone is agreed about what kinds of concepts we are using, but they disagree about how the referents of those two conceptual schemes are related. So, what's going on? Where does the disagreement arise? Dualists argue that because we are well aware of what phenomenal concepts essentially refer to (an experiential property) and what physical concepts essentially refer to (a physical property), we should accept that the apparent distinctness between the two types of property is a *real* metaphysical distinction. So, unless there is another way to understand what is going on with concepts that refer

essentially, this way of formulating neutral monism (or *a posteriori* materialism) will simply collapse into dualism. But as I mentioned before, there is one last strategy available.

The dual carving strategy will help illuminate what is really at issue. This is a strategy where the materialist admits that there is no difference between the two different types of concepts. Instead, this strategy states that "there are multiple ways of grasping the essence of some properties or kinds" (Goff, 2017, p. 125), which implies that the conceptual schemes in use are both referring to the essence of the property, but nonetheless there is only one (kind of) property. There's a sense in which this is what the *a posteriori* materialist and epistemic neutral monist have been trying to do all along, of course. So, it will be worth exploring this dual carving strategy in more detail, both to see what it has to offer and to see how the materialist and neutral monist might deploy it differently.

The clearest instance of the strategy comes from the powerful qualities view (which, as we saw in chapter 3, is closely related to Epistemic Neutral Monism, often acting as its inspiration).[61] As such, I'll adapt much of what they say to the current context of the mind-body problem.[62] John Heil's presentation of the powerful qualities view summarizes the dual concepts strategy well:

> "If *P* is an intrinsic property of a concrete object, *P* is simultaneously dispositional and qualitative; *P*'s dispositionality and qualitativity are not aspects or properties of *P*; *P*'s dispositionality, P_d, is *P*'s qualitativity, P_q, and each of these is *P*: P_d=P_q=*P*" (2003, p. 111).

It is easy to see how such a strategy could be adapted to the mind-body debate (which is one reason why Russellian monists have seen the debates to be so intimately related); simply substitute "dispositionality" and "qualitativity" for "phenomenality" and "physicality". This shows a valiant attempt to avoid the dualism objection against a neutral monist view. The idea is that "even though the qualitative [or phenomenal] and the dispositional [or physical] may be separable in thought, they are not separable in reality" (J. H. Taylor, 2013, p. 1291). However, while it is easy to see how this might be true for standard *a posteriori* identities, it is difficult to understand how this is supposed to work for identities where the concepts expressed by the terms on both sides of the identity refer via essential properties of their referents.

---

[61] There's an interesting interpretational question as to whether they *are* implementing this view. Goff (2017) makes a strong case that they are not, in fact, appealing to this view. However, if they aren't, then it seems impossible for them to distinguish their view from a dispositional/categorical dualism. The type of view that I develop in the next chapter is open to them, but I see little indication that they see modality as conventional.

[62] Specifically, I'll follow the developments of (Heil, 2003, 2010, 2020; Martin, 1993; Martin & Heil, 1999; J. H. Taylor, 2013), as well as Goff's (2011, 2017) objections to the strategy.

In these instances, we are aware of the referents of the concepts, and in the case of phenomenal and physical concepts, they are referring essentially to two different property types. So, how is this dual carving strategy supposed to work?

One way of considering this position is to suppose that although our phenomenal concepts refer to the experiences we suppose them to refer to, the property of experiencing (pain, say) isn't all there is to that property. There is also the activity in the brain (or c-fibers firing, as philosophers are wont to say). On the flip side, our physical concepts refer to the properties of the physical brain activity that we suppose they refer to, but there is, in addition, the experience which is also part of the property.

This "two aspects" approach clearly won't do, as it fails to distinguish itself from dualism. Moreover, it is not the approach that the powerful qualities view is advocating. It's a simple way of understanding what they might be saying, but it is mistaken. As we saw above from the Heil and Taylor quotes, the phenomenality and the physicality are identical to each other, so they can't be different aspects of the one type of property. Rather, they are different ways of understanding or *modes of presentation* of the same property.

The dual carving strategy states that we can attend to the property in different ways such that it presents to us as different but is nonetheless the same property. The analogy used here by Heil (2003) is that of the Necker cube or duck-rabbit. By focusing on the *whole* image in different ways, we are presented with the cube "popping out" or "receding" or with the duck-rabbit, we are presented with a "duck" or a "rabbit". In both instances, it is the whole image which appears in different ways but is unchanged; the duck-rabbit isn't part duck and part rabbit. Similarly, the property in the world is not part phenomenal and part physical, but rather our concepts present the property in two different modes. As such, the properties are identical. However, it's one thing to offer up an analogy that's supposed to illustrate a point, and it's another thing to be able to explain exactly what is going on. In the duck-rabbit case, it's pretty clear that we are attending to different *parts* of the image which makes it such that it presents as a duck oriented vertically or a rabbit oriented horizontally (and similarly for the Necker cube, we attend to one corner and it pops out and a different corner and it recedes). It's hard to see why we shouldn't think something like this is going on in both the mind-body and the dispositions-qualities cases. That is, one conceptual scheme "attends to" the phenomenal "parts" or properties of an experience while the other "attends to" the physical "parts" or properties. If this is what is going on, then despite claims to the contrary, the view is dualist.

The main problem here is that these concepts point to what is essential about their referents, and so if two concepts refer to the same property via different essential properties of that property, there seems to be nothing that could support that these different essences are necessarily tied to each other in the world. Taylor asks a pertinent question, "Why would we assume that in thinking of an experience under one concept, this stops us from being able to think of it under another concept?" (2013, p. 1293). Of course, when one of the concepts refers via contingent properties of its referent, we *shouldn't* assume that there isn't another concept under which we could think of the property. There is probably (perhaps always) another concept in the vicinity that picks out the same property using something essential to it.[63] This is exactly what the instances of *a posteriori* identities like Hesperus and Phosphorus or water and H_2O have shown us.

However, when both of the purported concepts refer via essential properties, we are led to ask what might prove that the properties that the concepts pick out are identical (or, at least, necessarily related) to one another. As Phillip Goff points out, the dual carving strategy has given us no answer and seems, more importantly, incapable of giving an answer. As he states it, the dual carving strategy commits us to a world that is "radically unintelligible" because "no amount of reasoning can reveal to us that the numerous ways of carving up the world 'hang together' (i.e., are just different ways of understanding the same reality)" (Goff, 2017, p. 128). Although Goff doesn't mention it here, the issue that he is referring to is whether there are *strong necessities*. Strong necessities are brute metaphysical necessities that have no explanation. This isn't simply that we've yet to find the relevant detail that would explain the necessity. Instead, strong necessities have no explanation *at all*. The dual carving strategy commits us to the existence of such necessary truths because it says that both phenomenal and physical concepts refer via the essence of their referents, but nowhere in the physical concept is there any reference to what the phenomenal concept refers to and vice versa. As such, there exists no reason to suppose that their referents are the same. Moreover, there is good reason to suspect that they are different because they seem capable of independent instantiation, even given full empirical information. If both conceptual schemes work by reference via essential properties (as the dual carving strategy says that they do), then the concepts tell us the essential

[63] As far as I've discussed the situation, there's a possibility that this is precisely what is going on in the dispositions/categoricals debate, but the way in which Heil, Martin, and others speak about these concepts seems to suggest otherwise.

properties of their referents. If these referents are different, then there simply isn't a necessary relation between them. The main problem with this is that the existence of such strong necessities makes the world itself radically unintelligible. Even if we have all the right concepts (that refer via essential properties), we wouldn't be able to tell if the referent of one is necessarily related to the referent of the other. There may even be identities between properties that could never be discovered even in principle.

3. The Puzzle – Final Version

The dual carving strategy makes it so that three different ontological views converge – *a posteriori* materialism, epistemic neutral monism, and property dualism – and it appears that Epistemic Neutral Monism finally comes face to face with the dilemma that plagues any neutral monist view. On the one hand, neutral monism can accept that there are strong necessities and that the necessary relation between phenomenal and physical properties is one of them. This give rises to a host of general epistemological concerns, but more pertinent for neutral monism is the fact that the view then becomes indistinguishable from *a posteriori* materialism and is forced to accept its attendant inability to meet the Explanatory Demand. On the other hand, Epistemic Neutral Monism can say that there are no strong necessities. This is essentially what the property dualists have been saying all along, and so going this route makes Epistemic Neutral Monism indistinguishable from dualism and is forced to accept its attendant inability to meet the Ontological Demand. This issue of whether there are strong necessities is what I think has been the central problem with interpreting any instance of neutral monism such that it is both coherent and stable. Even after the more obviously problematic formulations of neutral monism have been set aside, the issue of strong necessities still vexes the neutral monist position and is what gives so much force to the coherence/stability dilemma martialed against it.

Unlike the case of Reductionist Neutral Monism, this is a case where Epistemic Neutral Monism has an available resource that is unavailable to materialism – in fact, it's an essential resource that makes this view neutral monist rather than materialist or dualist. It bears some similarities to the dual carving strategy, and in some ways should be construed as a variant of this strategy. However, the ways in which this strategy is implemented is extremely important. The proponents of the powerful qualities view fail in their implementation of the dual carving strategy. Their failure is due to the way in which they construe identity and modality more generally. They believe that dispositional and categorical properties share a

strongly necessary relation and fail to see how this makes their view untenable. As such, I will not be interested in pursuing their implementation. However, I'll seek to learn from their mistake. On the other hand, dualists also make a mistake. Their mistake is in not seeing that the reasons for why there are no strong necessities should lead them away from dualism. As I'll explain further in the next few chapters, these reasons show that modal facts are conventional. Dualism says that the epistemological gap between phenomenal and physical properties exists because there is an ontological gap between them. However, a detailed understanding of why strong necessities don't exist will reveal that dualists have this backward. There is an ontological gap because there is an epistemic gap and there is an epistemic gap because our conceptual schemes of phenomenal and physical properties work in different ways. I will formulate this variant of Epistemic Neutral Monism – what I will call Conventionalist Neutral Monism – in the final two chapters. For the neutral monist, the issue isn't that phenomenal concepts are a special kind of concept that refers via essential properties, as the phenomenal concept strategy would imply, but rather that we have misunderstood the relationship between our concepts and the world.

Chapter 5: Conventionalist Neutral Monism

> "The world provides some material, the *substratum* (or *stuff*), which is neutral with respect to the features that are taken to be conventional. Onto this substratum, features of the kind in question can be conventionally imposed in many different ways."
>
> - Iris Einheuser, "Counterconventional Conditionals" (2006, p. 461)

1. Revisiting the Neutral Monist's Dilemma

The main dilemma faced by neutral monism is that it is either incoherent or is merely a restatement of one of the commonplace solutions of the mind-body problem. In the previous chapter, I argued that an epistemic version of neutral monism – that is, a version whereby the phenomenal and physical properties are merely two different ways of understanding the fundamental neutral property type – has the best chance of navigating this dilemma. However, I also showed that the standard ways of building such a view runs headlong into it, making it appear as if this final attempt at creating a neutral monist solution also doesn't work. In this chapter, I'll sketch a variant of epistemic neutral monism, which I will call conventionalist neutral monism, that can handle the dilemma in ways that the standard epistemic neutral monism cannot. I begin by reviewing the central issues that the standard epistemic neutral monism runs into in order to show what is needed. Then, I'll show how conventionalist considerations can meet these needs such that it successfully deals with both horns of the dilemma faced by neutral monism. I will not argue that this view is correct. Rather, my main goal for this chapter is to show how this view is both coherent and differs from the standard solutions to the mind-body problem. In the next chapter, I will show how this solution meets the demands from Chapter 1. Since the other solutions fail to meet these demands, I will conclude that conventionalist neutral monism is the best answer that we have.

At the end of the previous chapter, we ended up in a position where three strands of thought – *a posteriori* materialism, dualism, and epistemic neutral monism – converged. Following similar patterns of thinking, each view ended up concluding that the concepts used to refer to both phenomenal and physical properties make reference via essential properties. This seems to lead all three views into a position where there are really two different types of properties, corresponding to the two different types of concepts. Thus, these views converge on a dualist position whereby since the concepts refer to the essence of their referents, and since the concepts in no way refer to the other concept's referent, there must exist two different types of property. Of course, where dualism readily accepts that this is an ontological difference, *a posteriori*

materialism and epistemic neutral monism seek to keep this as a purely epistemic difference. In order to attempt this, *a posteriori* materialism invokes what Philip Goff (2017) calls the dual carving strategy.

This strategy accepts that both phenomenal and physical concepts make reference via essential properties and accepts that these essences in no way mention the essence of the referent of the other type of concept. However, since neither physical nor phenomenal concepts explicitly state that their referents are incompatible with a necessary ontological connection to each other, the thinking goes that this leaves the door open for them to have the same referent even though what is essential to the two concepts' referents is distinct. Thus, what appears to be two properties by conceptual analysis turns out to be one property that is referred to by two different concepts.

Of course, any non-trivial identity statement will be one where two different concepts refer to the same entity. The difference is that in the dual carving strategy, both concepts have their reference fixed by essential attributes that are not reducible via conceptual analysis to one another. Thus, the metaphysical dependence is not *a priori*. Moreover, it is different from usual *a posteriori* identities because in usual *a posteriori* identities, at least one of the concepts has its reference fixed by contingent attributes. It is this peculiarity about how these concepts refer that explains why we can conceive of the reference of the two concepts as different even though both concepts in fact refer to the same thing. However, those who employ the dual carving strategy agree that the concepts both refer via essential attributes. As such, the dual carving strategy, as employed by *a posteriori* materialism, claims that the existence of physical properties is necessary for the existence of phenomenal properties but that this fact cannot be explained either *a priori* or in the usual *a posteriori* way. Without some different way to explain this fact, it must be a necessary truth that is completely inexplicable. Not merely have we not found the explanation, but rather the explanation simply does not exist. One type of explanation of this necessary truth that will not work is to say that the existence of the phenomenal requires the existence of the physical precisely because the phenomenal is identical to certain kinds of brain states. This is, to be fair, a certain kind of explanation in that I had been speaking of a determinable relation of metaphysical necessity where I have now specified the determinate relation of identity. In some sense, the determinate relation does explain the fact that a determinable relation exists. However, that two seemingly different things are identical has an even greater demand for explanation.

This, then, is the basic difference between dualism and the current version of *a posteriori* materialism. Where they agree on the features of phenomenal and physical concepts, they disagree on whether there can be a different kind of *a posteriori* explanation of the necessary dependence of phenomenal properties on physical properties. Dualists argue that such a necessity can only be explained in the usual way while *a posteriori* materialists argue that there is some other explanation available. Without being able to explain the necessary truth, dualists argue that there simply is no such truth. Thus, there really are two properties.

Given that epistemic neutral monism also agrees with dualism and *a posteriori* materialism on the reference of phenomenal and physical concepts, epistemic neutral monism finds itself in a tough spot. It seems that epistemic neutral monism must choose. It can deny that such a necessary connection between the phenomenal and physical exists, and thus deny that any necessary connection exists between them at all. This way lies dualism. Or it can accept the necessary connection between the phenomenal and physical. But this would make it indistinguishable from *a posteriori* materialism. Neither option should sound appealing to the neutral monist not merely because neutral monism is supposed to be different, but rather because neither is a satisfactory solution to the mind-body problem in that neither solution can meet the demands laid out in Chapter 1. As discussed in the previous chapter, standard ways of formulating epistemic neutral monism are unable to navigate this final version of the neutral monist's dilemma. In the next couple of sections, I'll explain how a conventionalist neutral monism can give us what we've been looking for. The first step will be to show why we shouldn't accept the type of necessities that are demanded by the dual carving strategy as employed by *a posteriori* materialism. The second step will be to show an alternative to the dualist's conclusions about what this means, ontologically, for the physical and the phenomenal.

2. The Problems with Strong Necessities (or, I got a lot of problems with you necessities, now you're gonna hear about it!)

All *a posteriori* necessities are statements which are true in every possible world but for which there seems to be a world where the statement is false. In the case of supposed identity statements, this is usually explained in the following way. Each term is a rigid designator but where one term has its referent fixed by contingent features of it. For instance, in the case of 'water = H_2O', this statement is true in every possible world and both terms are rigid designators, but the term 'water' determines its referent via contingent

properties. Thus, while the statement is necessarily true, there are worlds where there is a clear, colorless, tasteless liquid in lakes and oceans, that falls from the sky, and nourishes life on earth which is not H_2O. Such a world looks and feels and tastes exactly like ours, but nonetheless, the liquid in that world is some other chemical or something else entirely. The usual philosopher's stand-in chemical is XYZ. So, in order to prove that we are in this H_2O world and not the XYZ world, we have to perform some empirical investigations. Thus, the statement 'water = H_2O' is *a posteriori* and seems merely possible despite being necessarily true. This is how the standard *a posteriori* materialist argues we should understand the identity between phenomenal and physical properties; phenomenal terms refer via contingent attributes and thus the physical properties are the "real" ones despite the fact that they are identical to phenomenal properties.

A posteriori materialism that employs the dual carving strategy states that the relationship between the phenomenal and physical is a different kind of necessary relationship, one where both sides are referred to by essential attributes. Thus, there must be a different explanation for how it is that every possible world that is physically identical to ours is also phenomenally identical even when it *seems* to be quite clear that physically identical worlds with phenomenal differences are metaphysically possible. This kind of necessity, it is claimed, is just like every *a posteriori* necessity except that there would be *no worlds at all* where there is a clear, colorless liquid, etc. which is not H_2O. This is supposed to show that some *a posteriori* necessities might appear to be contingent even though the terms used to state them and the corresponding qualitative facts determine that they are necessary. Thus, the worlds that are physically identical to our world with phenomenal differences *seem* possible but do not exist. And so, all and only the worlds where the identity statement between the phenomenal and physical is true exist. In the case of supposed identity statements, there is only one way for such statements to be necessary *in this way*. Each term must be a rigid designator and they also must have their referents fixed by necessary features of it (this shouldn't be confused with me saying that *all* rigid designation works this way. Of course, in standard *a posteriori* necessities, one term refer via contingent properties, as I said in the previous paragraph). This would be the only way for statements such as 'water = H_2O' to be true in all possible worlds. Of course, as I just discussed, this statement is *not* true in all worlds where the watery stuff has a different composition and where H_2O is not watery stuff. There are just some worlds where the substance that looks like water is not water. For this to be a necessity in the sense used here, those worlds that seem like they have a substance that is water but isn't H_2O must be illusory. Either there isn't really a substance that looks exactly like water or it is in fact

water (and also H_2O), despite our stipulating that it is not. We somehow have to be mistaken about our own descriptions of this possible world or mistaken about whether it is possible.

In other words, according to this version of materialism, there is a necessary connection between the physical brain and our phenomenal experiences such that the apparent world(s) where these come apart is spurious. It does not represent a genuine possibility. Allowing this kind of necessity has some consequences that we should think are strange at best, or even worse, entirely incoherent. David Chalmers (1996) has dubbed this kind of necessity a *strong necessity*. As I'll show, while strong necessities are supposedly *a posteriori* necessities, they turn out to be necessities which have no explanation at all. So, they end up being metaphysical necessities to which we have no epistemic access.[64]

Being mistaken about a description of a possible world, of course, seems reasonable. We can misdescribe *this* world, and it is where we *are*. So, we might ask what happens when we misdescribe the actual world and try to see if one of these kinds of mistakes fits the model of being a mistake about a possible world. It seems to me that when we misdescribe our own world, we make one of two mistakes. Either we make a mistake in *a priori* reasoning – or we get the empirical facts wrong – an *a posteriori* mistake. Since we are stipulating the empirical facts of possible worlds, it is difficult to understand the notion that we might get these facts wrong, but I'll return to this.

One kind of *a priori* mistake is a misuse of our own language. In this sense, we can easily misdescribe possible worlds. I can ask you to consider a world in which '2+2=5' is true and then ask you to infer various things about that world. But something has to have gone wrong if we are actually making any sense. Either we are using the symbols in different ways from normal, and so not really speaking English, or we are simply speaking nonsense. When we are making actual sense, it is always the former. For instance,

---

[64] To reiterate a point made in a footnote in Chapter 2, this is the point where what is commonly called "*a posteriori* materialism" in the literature is no longer, properly speaking, *a posteriori* materialism. As I understand *a posteriori* materialism, it denies Premise 4 of the Epistemic Argument as presented in Chapter 2, and as such, states that the ontological dependence relation between phenomenal properties and physical properties is discoverable *a posteriori*. Clearly, once materialism has advanced to the point of deploying the dual carving strategy and claiming that the dependence relation is a strong necessity (which cannot and need not be explained), this relation is no longer discoverable *a posteriori*. The view which began with a rejection of Premise 4 has thus transformed into a view which rejects Premise 1 of the Epistemic Argument. Thus, this type of "*a posteriori* materialism" is better named "obscurantist" or "strong necessitarian materialism", which follows the taxonomy of positions outlined in Chapter 2. However, as I mentioned in Chapter 2, I refer to the view as "*a posteriori* materialism" for the sake of consistency with the rest of the literature. One might think that I'm not giving this type of *a posteriori* materialism fair treatment by pinning them with the view that these necessities are inexplicable but see (Goff & Papineau, 2014) for at least one explicit endorsement of the view that the identity between the phenomenal and physical is a strong necessity and that strong necessities needn't be explicable.

if '2+2=5' were true, then would '2+2+2+2=8' still be true or would '2+2+2+2=10' be true? If the former, then we can glean that what has happened is that we are using the symbol '5' to refer to the number 4. If the latter, then we are using the symbol '2' to refer to 2.5.[65] But barring a mistaken (or at least non-standard) usage of our own language, what we seem to do when considering a world in which '2+2=5' is true is describe a world which is incoherent and therefore not possible.

The problem is that it *doesn't* seem to be the case that we are making this kind of mistake where we are misdescribing these worlds in which we say that the physical exists without the phenomenal or vice versa. We aren't using the term 'pain' in a non-standard way when we consider a possible world in which the physical facts are exactly the same, but wherein I lack the relevant experience when I stubbed my toe this morning and say that 'pain' does not apply. This is neither nonsense nor is it using the term 'pain' any differently from the way in which we actually use it. In fact, it is this fact to which the *a posteriori* materialists have agreed when they agreed that phenomenal terms and physical terms have their reference fixed by necessary attributes. The disagreements about whether the terms refer via contingent or necessary attributes just is a disagreement about whether we are making a linguistic mistake when we say that 'pain' might be used to describe something that isn't painful. By agreeing to the claim that phenomenal terms refer via necessary properties, the type of *a posteriori* materialism that uses the dual carving strategy cannot mean to say that we are misdescribing possible worlds by misusing our phenomenal vocabulary in describing these genuinely possible situations.

Instead, the mistake that must be happening in the case of strong necessities must be an empirical one about the world in question. The idea is that just like in our world where we might *think* that Hesperus is not Phosphorus and so describe a possibility that isn't genuine because of a mistaken observation in our world, we might be mistaken about which empirical facts might obtain in other worlds. So, although it *seems* like there are worlds where the phenomenal and physical are not the same, we are just wrong about that. Above, I disregarded this idea because we *stipulate* the empirical facts of possible worlds ourselves (or, perhaps more accurately, we stipulate the *relevant* empirical facts), so it seems like we cannot be mistaken

[65] Of course, the example is ambiguous. If the former, we could actually be changing the usage of both '2' and '8' to correspond to 2.5 and 10, respectively, holding fixed the normal usage of '5'. Similarly, in the latter case, we could be changing both '5' and '10' to refer to 4 and 8, respectively. Or we could be changing normal usages of the symbols '+' and/or '='. But the point in the main text still stands and is in fact strengthened by these observations.

about them. They do not need to be discovered in the same way that empirical facts are in the actual world, and so the epistemic difficulties associated with them simply do not apply.

However, there is one way in which this mistake might come about. We might stipulate an empirical fact about a world but not realize that this stipulation brings along other facts necessarily. For instance, when we stipulate that some world contains gold, we are also stipulating a world with protons, electrons, and the like. And this is true even if we have no idea that gold is made up of atoms or that atoms contain protons and electrons. In this way, when we stipulate the fact that the world under consideration is physically identical to ours, this set of facts may bring with it all of the facts about phenomenal properties. So, all worlds wherein the physical facts are the same, the phenomenal facts will also be the same even though we may not recognize this fact when making the stipulation that all the physical facts are the same. As a result, when we add on the further stipulation that the physically identical world under consideration is one where there was no pain in my stubbed toe, we have, unbeknownst to us, stipulated a world that is not possible – a world wherein the phenomenal facts are both the same as ours and different. Just as we can seemingly coherently state that a triangle might not have angles that add to 180 degrees but end up stating something incoherent and therefore state something impossible, we can seemingly consider a world where the phenomenal and physical come apart. This makes sense in the usual Kripkean *a posteriori* necessities because one of the terms refers via contingent properties of the referent. As such, it is easy to see how, for instance, we might be mistaken about whether gold or water or Hesperus come along with some other properties. All of those terms refer via contingent properties and so they could very well *not* have been associated with a referent with some other property – e.g. having atomic number 79, being H_2O, or being Venus. However, in these cases, we can explain away the apparent contingencies (following Kripke's examples) in ways that we can't in the case of strong necessities.

In other cases like that of the triangle, both terms do refer via essential properties of their referent, so they are closer to the situation that the dual carving *a posteriori* materialist agrees we are in with respect to phenomenal and physical properties. However, in these cases, we (rightly) require *proof* of the necessary connection before agreeing that the world being described is not possible. For triangles, we can geometrically prove that all closed three-sided figures necessarily have angles that add up to 180 degrees. So, we accept that a world where this is not true is not possible, even though many people may not recognize it at first. For phenomenal and physical concepts, these kinds of purported *a priori* necessary connections

could only be proved by conceptual analysis. And the relevant conceptual analysis has always turned up empty of any necessary connection. There's nothing in the concepts of pain and c-fibers firing that makes facts about pain necessarily follow from facts about c-fibers firing or vice versa. So, if this kind of strong necessity between physical and phenomenal properties exists, we have no proof of it.

This is the case with the version of *a posteriori* materialism under consideration. We take 'pain' to refer essentially to the experience of pain and 'c-fibers firing' to refer essentially to an activity in the brain. Thus, where we have c-fibers firing we know we have some brain activity that, in *our* world, is always empirically associated with pain. But what could possibly justify this jump from knowing what happens in our world to what is the case in some *other* world? Even with the stipulation that every physical fact in the other world is the same as the physical facts in our world, there is simply nothing about the empirical situation that could force us to conclude that the term 'pain' applies to a situation in this other world where all we've stipulated is that c-fibers are firing. The referent of 'c-fibers firing' only necessarily gives us the fact that some brain activity is occurring, and the referent of 'pain' only necessarily gives us a phenomenal experience. What is missing is that when one term applies, the other necessarily applies as well. There is nothing in the terms themselves that tells us this, of course. If there were, then such a necessity would be true *a priori*.

So, what could it be about the world that would make this necessary link true? Not the fact that they perfectly covary in our world. Many properties perfectly covary in *a* world and yet are not *necessarily* connected in any way – for example, having a kidney and having a heart. But if perfect covariation is not enough to prove that certain brain activity necessarily comes along with experiences, then it doesn't seem like *anything* could explain the supposed necessity between physical and phenomenal properties. With standard Kripkean *a posteriori* necessities, the appearance of contingency can be explained away by showing that we have either misunderstood the use of the terms or that we have misunderstood our stipulations about the world in question. For instance, in the water and H_2O case, although we had to do some empirical analysis to determine that it was H_2O in the lakes, rivers, and oceans, and falling from the sky when raining, etc. we can then use these facts to explain why a world wherein water exists without H_2O is impossible. Unlike with terms like 'pain' and 'c-fibers firing', there *is* something in the term 'water' that makes it the case that when it applies, the term 'H_2O' necessarily applies as well. Following (Sidelle, 1989), we can analyze 'water' to mean something along the lines of 'necessarily whatever deep structure in our

world that explains the stuff we find in rivers, lakes, oceans, falls from the sky when it rains, etc.' Since we've found that in our world, this stuff is explained by H_2O, we have a proof by conceptual analysis that a world with water but without H_2O or vice versa is not possible.[66] Unlike these standard cases, the strong necessities under consideration here are necessary truths that can in no way be explained. The term 'pain' simply refers to the painful experience, not to 'whatever deep structure in our world that explains the painful experience'. So, the appearance of contingency does not go away even for an ideally rational agent. And thus, we have worlds that *seem* possible in that they are conceivable by ideally rational agents, but which are not possible, and if we ask *how we know* that they aren't possible, no explanation is available. It is a brute, inexplicable fact that a world physically identical to ours but with no phenomenal properties is not possible.

The problem with strong necessities is this inexplicability. Because we cannot explain away *this* appearance of contingency, we cannot know anything about which worlds are possible and which are not. Again, the case is not one where we are simply lacking some empirical facts, as we were before the discovery that Venus had two names, Hesperus and Phosphorus. There is, instead, a necessary fact which is untethered to the other facts of the world. If it were tethered to the other facts, then it would be possible to explain *this* fact using the other facts connected to it. But without the ability to explain away the illusion of contingency, one could know every other fact in the world and never be able to learn this one. But if this necessity is unknowable, then any ideas we have about the space of possible worlds gives no indication about what is possible. Even worse, there is nothing that could *dis*prove the existence of a strong necessity. As we've been discussing, all evidence we have points to the contingency of the relation between phenomenal and physical properties and as we've shown this evidence cannot be explained away. A strong necessity is a necessity which always retains its associated illusion of contingency and thus exists despite all evidence to the contrary.

Of course, one might try to justify the existence of such necessities with a story about the limitations of human knowledge, and while there certainly are limitations on such knowledge, that fact does not help establish the plausibility of strong necessities. A limitation on human knowledge implies that there is

[66] This also follows the notion that (Chalmers, 2014) dubs [MNM] for "modal/nonmodal" that "For every a posteriori metaphysical necessity 'Necessarily *S*', there is a nonmodal truth *D* such that 'If *D*, then necessarily *S*' is a priori" (pg. 791). In both Chalmers' analysis and Sidelle's, these *a posteriori* necessary truths end up being necessary truths because of some *a priori* fact about the terms being used.

something that *human* beings cannot discover, generally some empirical fact that is inaccessible to the types of beings we are. In this case, though, there isn't anything that could be discovered by *any* being that could show that a strong necessity exists. So, it cannot be some limitation to human knowledge, but rather a strong necessity would be some *unknowable* fact if it were to exist. The only possibility to learn that such a fact exists would be some sort of divine revelation. But if that's the *only* way to discover some fact, then it either (a) violates the Ontological Demand because it would not be connected to the other facts about the world, and we would be right to reject an account that requires this, or (b) ignores the naturalistic presupposition that is part of the current debate over the mind-body problem, and we would again be right to reject an account that requires this.

So, maybe we should try to find some third way of misdescribing possible worlds. We have been thinking of these mistakes as analogous to the types of mistakes that we make when we misdescribe the actual world, but of course, we are now talking about *possible* worlds. So, the scenario may be different. However, it's doubtful that there could even be some third mistake.[67] As I've already mentioned, the mistake about our description of the world in question must be either an *a priori* mistake or an *a posteriori* mistake. There are no others. There are, perhaps, different types of *a priori* and *a posteriori* mistakes from those that I discussed, but I think they will be handled in the same ways that I handled these types of mistakes above. The question remains, what kind of epistemic mistake might there be that is neither *a priori* nor *a posteriori*? In order for there to be a strong necessity between the phenomenal and the physical, we would have to be mistaken about either the empirical content of the world or about the referents of the terms we've used to describe the world where the phenomenal and physical come apart. These are the only two types of mistakes we *could* be making. Since we aren't making either of those mistakes, the *a posteriori* materialist is left boldly claiming that phenomenal and physical properties are necessarily connected despite the fact

[67] In *Natural Minds* (2004), Thomas W. Polger suggests that the problem is that we don't yet know the full identity conditions of the relevant phenomenal and physical properties. There's more to be said on this suggestion, but I can make some preliminary remarks here. The fact is that we don't need to know all of the details of the identity conditions of two properties in order to know that they are not identical. We just need to know that one set contains a condition that is not contained in the other set of identity conditions. By his own lights, pain is necessarily painful. However, being painful is not any part of the identity conditions of the brain process associated with pain. So, there is one thing in the set of identity conditions for pain that is not in that of the brain processes associated with pain. A natural response is to state that I'm begging the question on this, but then I'd point the reader to the main text where I think it becomes clear that we know enough of the identity conditions to know that they are not the same.

that both *a priori* and *a posteriori* reflection leads us to conclude that the phenomenal and the physical have no necessary connection.[68]

3. Conventional Properties

By rejecting the existence of strong necessities, we seem to be led to accept dualism. However, the arguments of the preceding section might also lead us to a different conclusion based on how we view the relationship between a concept and its referent. In the remainder of the chapter, I'll sketch a way of understanding the above discussion that allows us to keep the epistemic understanding of the situation such that the physical and the phenomenal are different while avoiding the dualist's metaphysical view.

Seeing this difference begins by noticing the tight connection between the epistemic and the metaphysical that is required once we have abandoned strong necessities. Essentially, what we have accepted is the dualist's premise that what is (ideally) conceivable is metaphysically possible.[69] The dualist would like us to just leave it at that. Since we can conceive of a physical duplicate of ourselves without any phenomenal experience, it is possible, and since it is possible, there is no metaphysical (i.e. necessary) connection between the physical and the phenomenal. The phenomenal and physical are thus ontologically distinct and dualism is thus proved true. However, what we have yet to explore is *why* it is true that conceivability entails possibility. An exploration of this, I'll be at pains to show, reveals the position that we've been searching for, a neutral monism that is both coherent and different from standard solutions to the mind-body problem.

---

[68] In *The Character of Consciousness*, Chalmers puts the point as "the space of (centered) metaphysically possible worlds is *smaller* than the space of epistemically possible scenarios" (Chalmers, 2010, p. 169). I'm not a fan of putting the point this way for a few reasons. First, there is a significant sense in which there are more epistemically possible scenarios than metaphysically possible worlds. We can be mistaken after all. Of course, when Chalmers is being more careful, he makes sure to specify that the epistemic scenarios are those of an ideally rational agent and are maximally specified and so the scenarios that we may mistakenly *try* to refer to are not epistemic scenarios in the relevant sense. Nonetheless, putting the point in this way tends to lead to some confusion. Secondly, saying that there are more epistemically possible scenarios than there are metaphysically possible worlds entails that strong necessities are metaphysical necessities for which we have no epistemic access, which is more in line with the way I discuss it in the main text. This puts better emphasis on the strangeness of the view. Despite a difference in emphasis, the basic point remains the same. These purported necessities are relations for which there is no evidence, not because we haven't found it yet, but because it doesn't exist. The big question for *a posteriori* materialists who believe that strong necessities exist is how to square these necessities with such a strong commitment to the Ontological Demand, which is the usual starting point for materialists.

[69] From here forward, when I mention that conceivablity entails possibility, it should be understood to include all of the qualifications about empirical information, ideal rationality, etc.

It should be pointed out that at this point, we've not only accepted that conceivability entails possibility, we've accepted *all* of the premises of the Epistemic Argument that was presented in Chapter 2.[70] So, what is going on? First, I'd like to remind you that the conclusion of that argument is the falsity of materialism (although it also works as an argument against ontological idealism as well, though this is generally not emphasized due to the position's relative lack of support). The fact is that any argument that employs a premise like 'conceivability entails possibility' cannot, by itself, establish that dualism is true. Even if we grant that the antecedent is also true – that it is true that we can conceive of a world that is exactly the same as ours physically but which is devoid of phenomenal properties of experience – we have only proved that materialism is false. The idea that such arguments thereby prove that dualism is true is based on the false trilemma that either materialism or dualism or idealism is true. This trilemma (and, ultimately, the central dilemma posed against neutral monism) gets its force from a hidden assumption that is shared between all of the standard solutions to the mind-body problem. It is also shared by the variations of neutral monism that have been explored in previous chapters. This is an assumption about the basic ontology of the world, that there is one privileged ontology that our theories can either describe correctly or incorrectly. It is this shared assumption that makes the impasse presented in Chapter 2 seem so intractable. Assuming that there can be only one ontology that correctly divides the world into its true fundamental components, the arguments over the mind-body problem are about which of the ontological views is correct. However, by turning this assumption on its head in the appropriate way, a neutral monism will emerge that can preserve the idea that conceivability implies possibility while denying its apparent dualist conclusion. Here, I'll present this standard understanding of the basic ontology of the world in more detail and then describe two alternative understandings of what there is in the world. Finally, I'll illustrate how these alternatives can help specify a neutral monism that escapes the dilemma that opponents have leveled at it.

---

[70] For your convenience, I've copied my presentation of those premises from Chapter 2 here:
1) All ontological dependence relations have a corresponding entailment relation.
2) Anything with an entailment relation can be proven on the basis of either *a priori* evidence or *a posteriori* evidence.
3) There is no *a priori* evidence that could prove an ontological dependence relation between the phenomenal and the physical.
4) There is no *a posteriori* evidence that could prove an ontological dependence relation between the phenomenal and the physical.

The standard view of the world is that the world comes pre-made into ontological kinds, and particulars within those kinds, which our language can either correctly track or incorrectly cut across.[71] In other words, the world contains ontological "perforations" along which our terms can "pop out" the correct ontological objects[72] if they are operating correctly, otherwise they act as scissors cutting out "things" which are not real, though they might have been real if the perforations had been made differently. So, some privileged few of our terms, but not all of them, pick out the *real* objects in the world.[73] To speak in a fundamental language, then, is to speak in a language that pops out the real objects at the fundamental level of description. To speak in a complete language is to speak in a language that pops out all and only those objects determined by the world's perforations. Much of our natural languages is a mixed bag, sometimes finding the perforations in the world and other times cutting at the world in ways that do not yield real objects or kinds or properties.

Importantly, on this standard view, very often our language does no cutting or popping out in any direct way at all. Instead, language merely *indicates* an object for study, and it is up to our investigations to determine where the boundaries of the indicated object lie. This is the project of the *a posteriori* discovery of the identity conditions for the referents of our terms. Those terms which indicate an object for which there are perforations are then natural kind terms while others are decidedly non-natural since they indicate objects for which there are no perforations provided by the world. So, no object can be popped out by this indication. If we want to refer to an object at all with such indications, we'll have to go into the world and cut across the boundaries provided by it. But cutting across the perforations provided by the world will yield, at best, an object that is not *real* (or, as people like David Lewis would put it, is not *natural)* because

---

[71] The standard metaphor is that of "cutting nature at its joints" from Plato's *Phaedo*, David Lewis's natural properties, and Ted Sider's notions of structure, fundamentality, and the privileged quantifier. I take these to be defenders of the standard ontological picture *par excellence.* For the most recent, extensive development of such a view, see (Sider, 2011).

[72] The term "object" is here used in a general sense to mean any object of reference. In particular for the current discussion, this includes properties.

[73] The phrase "privileged ontology" is from (Sidelle, 1992, 1995). He does not, however, defend or reject a privileged ontology (at least, not directly). Rather, he argues that the standard interpretation of Kripkean rigid designation requires a privileged ontology. In opposition, he offers a semantics of rigid designation that can remain neutral between these different ontologies. In this way, there is no direct argument from the semantic notion of rigid designation to the metaphysical conclusion of privileged ontology. For my part, I think the allure of the standard interpretation of rigid designation is because many people presuppose the privileged ontology in the same way that we see in the mind-body debate. As I say in the main text, it's not at all clear that this view of ontology is the correct view (in fact, I think it is incorrect, though I'm not arguing for that here, at least not directly), and as such it is worth exploring what we get with other ontological views.

cutting it out is something further that we had to go and do. For real or natural things in the world, the work of dividing things up is done by the world, not by us.

With this notion of a privileged ontology brought into the light, we can recast the debate between materialism and dualism. Materialism claims that either phenomenal terms cut across the real properties in the world or both phenomenal and physical terms pop out real properties in the world, but for various reasons, the physical terms do so at a more fundamental level of description.[74] Nonetheless, the idea is that the physical language is a complete language. It is able to indicate all of the real properties that the world has given us and to do so at the fundamental level of description. Dualism, on the other hand, argues that physical terms miss something given to us by the world, namely the phenomenal properties. So, the physical language is not complete. And since physical terms do not describe the phenomenal properties at all, obviously they do not do so at the most fundamental level. So, the only complete language for dualism is one containing both physical and phenomenal terms and the only fundamental language also contains both physical and phenomenal terms.[75] In all of this, it is assumed that what our language is doing is to merely pick out (successfully or unsuccessfully) properties that have their identity conditions given by the world rather than by us.

This privileged view of ontology is not obviously true, though, and there are two alternatives. Both of them reject the idea that some of our terms, but not all of them, pick out real objects in the world. The more permissive view says that all of our terms pick out real objects in the world. On this view, the world comes with basically all of the perforations that might be possible. Thus, there is no way to illicitly cut boundaries that don't really exist in the world. Any way of dividing up the world will yield a real object,

---

[74] One might think that emergentist dualists also claim that physical terms describe the world at a more fundamental level of description, so this description of materialism is not correct. But this depends on how the emergence is supposed to work. If it is the kind of emergence that leads to truly ontologically distinct properties, then what emerges (the phenomenal properties) are simply new fundamental properties. And so, the physical terms won't always describe the world at the most fundamental level. If it is another kind of emergence, then the view is, in fact, materialist.

[75] There is a further disagreement here between materialists and dualists. Polger (2004) is fairly explicit about the fact that he thinks that identity conditions are determined *a posteriori*. Others, like (Goff & Papineau, 2014) seem to also think that this is the case, though they are less explicit. On the other hand, see (Chalmers, 2014) for an explicit endorsement of the idea that identity conditions are *a priori* knowable. Nonetheless, the objects picked out by these terms either correspond to the perforations in the world and thus yield real objects or they do not. So, the disagreement between materialism and dualism ends up being about whether or not we know the identity conditions of the physical and phenomenal yet. Chalmers claims that we do and further that we know that they are different and thus the physical terms will not make a complete language. Furthermore, the phenomenal and physical terms are among the privileged class that correctly pick out objects in the world that correspond to the world's perforations. Again, the basic assumption on both sides is that these terms pick out properties wholly given by the world itself.

assuming that thing is in the world.[76] So, while there may be properties or kinds like unicorns or mermaids, they don't exist because nothing falls in the extension of the terms. However, there are quite strange objects like the object consisting of each President during the times they are in office. This object is studied by historians and has a name, The President, which is different from the person who happens to occupy the office of the Presidency at any given time. On the privileged ontological picture, such an object is not a *real* object since it is not in any way given by the world. Historians do study the presidency but that does not mean that there is an object given by the world that they are studying. They are simply studying the effects of the various people during the times in which they held that office. On the permissive view, any way that is available to divide up the world yields a real object, even if it is bizarre.

The more restrictive view says that none of our terms pick out real objects in the world because there simply are no perforations in the world. If what is required to be an object is that its identity conditions are built into the world, then the restrictive view says that there are no objects (properties, kinds, etc.) at all. Or, at least, according to the privileged view of ontology, there are no *real* objects. Thus, this view is sometimes called ontological nihilism.[77] I think this is misleading, however, because this view can be subdivided further. One might insist that objects must have their identity conditions given by the world and yet think that this never happens. On this view, there simply aren't any objects and we should be nihilists about them. This is a strange view, though. It's not clear why we ought to think of objects this way. By the nihilists' own lights, there aren't identity conditions given by the world, so it's more than a little strange to think that the only way to think of an object is one where the term is useless because nothing falls in its extension. On the other hand, one might think that there is a better way to understand what it means to be an object. Objects require that some identity conditions be met and that's it. There are no further requirements about where these identity conditions come from. This view of objects can be combined with the view that identity conditions are never given *by the world*. Thus, all objects require us to invest some

[76] My favorite name for this ontological view is "plenitudinous bazillion-thingism", coined by Karen Bennett (2004). I stick to the less fun and less descriptive "permissive" ontology simply for brevity.

[77] It is actually very rarely called ontological nihilism. More often, the term nihilism is reserved for the view that says that there are simple objects, but no composite objects. So, ontological nihilism is more often a nihilism about composition rather than objects, and is an example of a privileged ontology. However, see (O'Leary-Hawthorne & Cortens, 1995; Sidelle, 1998) for a few examples of using the term ontological nihilism to discuss the view that there are no objects, even simples.

conceptual labor in order to yield the identity conditions that, on the privileged view, are sometimes given by the world. There are objects, but they are carved out by identity conditions that we come up with. This would be a form of ontological constructivism. On the restrictive view, the world simply never does this work whereas on the permissive view it does this work for whatever identity conditions there might be. So, where on the privileged view, our terms only sometimes pick out real objects, on the permissive and both kinds of restrictive views, our terms can never fail to pick out objects that are as real as any object could be.[78] Of course, they can never fail for different reasons, but for the current purposes, this difference will amount to much the same thing. As such, I'll speak of these alternatives as if they are same and refer to them as a conventionalist ontology.[79]

As interesting as this ontology may be, it is much too radical for our purposes. Notice that this version of a conventionalist ontology entails that our terms can *never* incorrectly pick out objects of reference. We can only misunderstand which objects they pick out. For the current discussion, we need only concern ourselves with a conventionalist ontology restricted to the domain of phenomenal and physical terms. This restricted conventionalist ontology can remain silent about whether there are world-given "perforations" or identity conditions for the referents of other kinds of terms (e.g. moral, aesthetic, dispositional, categorical). Of course, it may be plausible that if the referents of phenomenal and physical terms do not have world-given identity conditions, then no objects of reference for *any* terms will have such identity conditions given by the world. But this need not be presupposed for the current discussion.[80] Thus, we can follow Iris Einheuser's conventionalist framework where "the world provides some material, the *substratum* (or stuff)" (2006, p. 461) which does not contain identity conditions of the objects of reference

[78] This is not to say that all terms will have a referent. So, for instance, the term 'unicorn' does not refer to anything and so there are no such objects as unicorns. However, if the term *did* refer, there would be a further question under a privileged ontology about whether or how far unicorns were real or fundamental. For permissive and restrictive ontologies, no such further question will arise because the only way for an object to exist is for some identity conditions to obtain in the world. There is no distinction between identity conditions given by the world and identity conditions due to human conventions as there is in a privileged ontology.

[79] I realize that this may be somewhat confusing, but it is not meant to imply that they are the exact same view. Although, I should say that despite the fact that I understand what the distinction is supposed to be, I don't see what this distinction amounts to. This may be some latent pragmatism coming through, but on the face of it, it seems like the distinction between the views is a distinction without a difference, and any dispute between their proponents is something of a verbal dispute. With that being said, this may be the result of my thinking more about the current topic and it is a distinction without a difference in *this* context. Thus, combining them should really be viewed as more of a simplification than anything substantive.

[80] As such, following (Einheuser, 2006), this should be seen as merely a methodological assumption in order to simplify the discussion. Nothing metaphysically important hangs on this restriction.

of phenomenal and physical terms, namely phenomenal and physical properties. Onto this stuff, our conventional practices impose the identity conditions for phenomenal and physical properties such that those properties are cut away from the rest of the stuff. There are no perforations in this stuff which might yield anything like phenomenal or physical properties. Notice that on this view, the world provides just about everything that we normally think of the world providing, minus the identity conditions for phenomenal and physical properties. This is not a view where anything we imagine is true. If we imagine that unicorns exist, for example, even if we imagine them as existing conventionally, we would be wrong because although we may give the correct identity conditions for what it would be for something to be a unicorn – a white, mythical horse with one long horn – the substratum does not contain the features that match those conditions. There are no unicorns. Similarly, if the substratum of the world contains nothing that matches the identity conditions for one or the other of phenomenal or physical terms, then either materialism or idealism would be true. However, I think that on any plausible set of identity conditions for phenomenal and physical properties, it will turn out that both are supported by the substratum in our world. Both exist, and few deny this.

What is disputed is whether one set of properties or both exist in a way that cuts nature at its proper joints. Materialists argue, using the standard ontological assumption, that insofar as the world is concerned, the identity conditions of physical properties are given by the world while the identity conditions of phenomenal properties are not. Thus, the physical properties can be "popped out" of the world while the phenomenal properties either cut across these natural divisions or simply come along for the ride once we have the physical properties. On the other hand, dualists argue that both physical and phenomenal properties have identity conditions given by the world and that these identity conditions are different. Both sets of properties are real. Thus, we can see that a neutral monism has no logical space if it takes on board the privileged ontology that is standard in the mind-body debate. Neutral monism also claims that the physical and phenomenal are equally fundamental. So, if they are both in the privileged ontology, then it is a dualism, but if they are both outside of the privileged ontology, then there are serious questions about what exactly falls into the privileged ontology.[81] However, once the assumption of a privileged ontology is

[81] I take it that this can only work with some sort of reductionist neutral monism, but see Chapter 4 for why that is not a promising route to take.

dropped, then merely having phenomenal and physical properties supported by the substratum will entail that both exist in the same way that the other does, but it will not yield an ontological dualism.

4. Not a Dualism

The basic difference between dualism and conventionalist neutral monism turns on whether we think the metaphysical difference between phenomenal and physical properties "turns on features of our concepts rather than substantive metaphysical features of the world" (Chalmers, 2014, p. 792). Dualism takes the difference to turn on substantive metaphysical features of the world whereas conventionalist neutral monism takes the difference to be one principally derived from features of our phenomenal and physical concepts. In this section, I'll further expand upon this notion of conventional properties while explaining how this avoids becoming a form of dualism.

From the perspective of a conventionalist ontology that accepts the dualist Conceivability Argument, it might at first appear that we merely have a different kind of dualism. On the conventionalist ontology, it is true that ontologically speaking *both* phenomenal and physical terms pick out properties at an equally fundamental level of description. So, it appears that such a view is dualistic given that we have already agreed that the phenomenal and the physical are ontologically distinct. In fact, there's even some understanding of what it means to be dualist such that this is correct. However, this is certainly not a *standard* type of dualism, and it's so far removed from that standard type of dualism that I don't think the label properly applies here. This is why I call this view a conventionalist neutral monism. The basic difference is that while both dualism and conventionalist neutral monism agree that the phenomenal and physical are distinct and equally fundamental, and even agree that this is proven by their possibly being distinct (rather than some actual distinctness in the world), and further agree that this possibility is proven by a corresponding conceivability, they disagree about *why* the conceivability proves the possibility.

The standard dualist view accepts a privileged ontology where both the phenomenal and physical properties have identity conditions built into the world. Since our phenomenal and physical terms are tracking these properties, and these properties are different, and neither entails the other, we will be able to conceive of scenarios where they come apart even if they never in fact come apart. So, the fact that these properties are given different identity conditions by the actual world leads to their possibly coming apart. When this is combined with the fact that our terms are tracking these properties correctly, it becomes true

that we can conceive of them coming apart. The conceivability, then, proves the possibility because if it weren't possible for them to exist without one another, then we couldn't conceive of them as different. It is the ontology that comes first and since our terms properly pick out the properties, our epistemic notion of conceiving properly tracks the ontological status of the properties. We basically use the notion of conceivability to backtrack to the possibility, so despite the fact that we use our conceiving the properties as different to prove their possibly being different (and therefore actually being different), what makes it the case that conceivability is a guide to possibility is that a conceivable scenario is conceivable *because of* the existence of a corresponding possibility.

On the other hand, for conventionalist neutral monism, the story that dualism tells gets things back to front because of the difference in the basic view of ontology. Where dualists merely think that conceivability is a *guide* to possibility, conventionalist neutral monists claim that the same thing that makes some scenario possible also makes it conceivable. Rather than conceivability merely tracking the existence of a possibility, our willingness to describe a situation as possible is what makes it possible. The idea is that the identity conditions for things in the world, in this case phenomenal and physical properties, are not given by the world. But it seems perfectly reasonable to suppose that there can be no object without identity conditions for it because without such identity conditions we would have no basis for the truth of whether or not the entity exists. In order to know whether there are phenomenal properties in the world, we need to first know what would count as a phenomenal property existing in the world, and then we'd have to check the world to see if these conditions are met. So, since there are no such identity conditions for phenomenal and physical properties in the world, as far as the substratum of the world is concerned, there are no phenomenal or physical properties without some conventions that carve them from the rest of the substratum and say *this* is a phenomenal property and *this* is a physical property. It is only once these identity conditions are given that we can say whether the substratum contains two types of things or only one. Since there are two different sorts of identity conditions, we have two types of properties. This is where conventionalist neutral monism agrees with dualism. The disagreement with dualism is that such a difference in the world is not a result merely of the world itself being a certain way but rather a result of the world being a certain way in combination with our use of certain conventions which give us two different sorts of identity conditions.

Saying that our willingness to describe a situation as possible is what makes it possible is *not* to say that it wasn't possible before we came along and said it was so. Nonetheless, it is the conventions that govern our use of language that make it the case that it is possible, and these conventions are not a part of the mind-independent world. These conventions, or rules, can be applied to situations when we are not around; they can even be applied to worlds where no entities with complex thought exist. In other words, although the possibility is a result of our conventions, this is not a temporal or causal relation. Rather, it is a logical one.

All of this is especially important in the mind-body case. Since the phenomenal perfectly covaries with (certain portions of) the physical in *our* world, it might seem like they are then necessarily identical. However, despite the fact that (some subset of) their identity and application conditions carve at the exact same substratum, they do it in different ways. And because they carve the substratum in different ways, it is possible that they could carve in such a way that the substratum itself comes apart. This would lead to causally verifiable differences in the physical and the phenomenal, disproving the causal closure of the physical and proving some form of interactionist dualism. However, as far as our evidence is concerned in our world, it just so happens that these two ways of carving the world carve at the same spatio-temporal portion of the substratum such that there is no causal variation between the phenomenal and the physical (at least, when we restrict the physical language to that concerning brains and nervous systems and such). So, where dualism states that the physical language is incomplete because it fails to talk about everything that there is, on the conventionalist neutral monist account, the physical language is incomplete because it fails to talk about everything in *all the ways that we find important*.

To make this a little more concrete, I'll offer a pair of analogies before going into some more depth. The first is the classic philosophical puzzle of the lump and the statue, which I like to call "Lumpl and Goliath".[82] There are several ways to build the puzzle, but here is one of them. Suppose a sculptor makes a bronze statue by combining liquid copper and tin into a mold in the shape of Goliath. Then, upon removing the mold, the sculptor is unsatisfied with the result. So, looking to start again but not waste the original copper or tin, the sculptor puts the bronze Goliath into a solution of nitric acid (along with several other things), resulting in the tin and copper being separated for reuse. Now, the puzzle arises. During the time after combining the copper and tin until it is placed into nitric acid, is it a lump of bronze – we can call this

[82] These names come from (Gibbard, 1975), though the example he uses is one with a lump of clay and a statue.

"Lumpl" – or is it a statue, "Goliath"? It is tempting to say both – Lumpl is the same as Goliath. When the sculptor makes or destroys one, the other is made or destroyed. When you look at, lift, or purchase one, you simultaneously do the same to another. They are, in actual fact, inseparable. However, there are serious problems with saying that Lumpl is identical to Goliath. The main issue is that although Lumpl and Goliath in fact perfectly covary, they *could have* come apart. For instance, the sculptor could have destroyed Goliath by simply melting the alloy in order to re-mold the statue or hammered it into a different shape out of frustration. This would have destroyed Goliath while leaving Lumpl intact. Or the sculptor could have removed some portions of the bronze in order to refine Goliath into the desired shape thereby destroying Lumpl while keeping Goliath intact. These possibilities show that Lumpl cannot be the same as Goliath and yet they perfectly covary during their actual existence. We are left wondering what it is that is the same about them such that they perfectly covary and yet they are nonetheless different.

There are various explanations for what is going on here, but the conventionalist answer is one that will be instructive. For a conventionalist, there is nothing about the world itself that is different. The substratum of which the bronze statue is made is both lump-like and statue-like. The difference is only one in the conventions used to identify Lumpl and Goliath. Lumpl is identified as "a lump of a certain amount of bronze" while Goliath is identified as "a statue in the shape of Goliath". The stuff supports Lumpl and Goliath, though it doesn't, by itself, provide the distinction between this particular lump and this particular statue. The distinction arises only in the way we describe what is there. In the actual world, they perfectly covary but we are prevented from saying that they have any necessary connection by what *could* happen in other, counterfactual situations. With some quite important differences that I'll get to shortly, this is the same situation as phenomenal and physical properties. Conventionalist neutral monism gives a conventionalist answer akin to the conventionalist answer given above about Lumpl and Goliath. Conventionalist neutral monism claims that the perfect covariation in our world is explained by the same features of the substratum supporting both phenomenal and physical properties while the difference in the properties is explained by a difference in the conventions used to distinguish them.

There are major disanalogies between the Lumpl-Goliath and mind-body cases, though. One of them is that Lumpl and Goliath are both *individuals* rather than properties or types. So, a second analogy will draw things closer. We return again to water and H_2O, but with a focus on the kind that the term 'water' supposedly picks out. This kind has a collection of properties – being liquid from 0-100 C in normal

pressure, being clear, colorless, tasteless, being in lakes and rivers on Earth, etc. I'll refer to this collection of properties as 'wateriness'. There is a question about how we should treat this collection of properties and whether it is also necessarily connected to H_2O the same way that water is. Now, a common conventionalist story about the term 'water' says that it is semantically built into the term that it necessarily picks out the deep explanatory structure of wateriness in the actual world, whatever that may turn out to be.[83] Of course, the current underlying explanatory structure of modern chemistry says that this is H_2O, and so this is how water is necessarily H_2O. And the fact that H_2O is the deep explanatory structure of wateriness is contingent on how the actual world turned out to be. So, 'water' refers to H_2O via contingent features whereas 'H_2O' refers to H_2O via essential features. On the other hand, the term 'wateriness' does not work like this. I just identified (some of) its essential features above. Wateriness necessarily picks out those features and is only contingently related to the deep explanatory structure of those features. So, wateriness and H_2O are not identical and, further, have no necessary connection. This is analogous to the fact that Lumpl and Goliath are not identical despite the fact that they are always found together during their existences.

But what is it that distinguishes wateriness and H_2O? There are many ways to answer this question, but a conventionalist answer can again be instructive here. As far as features of the substratum are concerned, wateriness is not itself any different from H_2O; the illusion of contingency arises because there are two sets of identity and application conditions which are supported by the same features in the substratum. In the same way that Goliath is a lump as much as Lumpl is and that there is nothing about either the lump or the statue that does not apply to the other, there are no features in the substratum that apply to H_2O that do not apply to wateriness. Wateriness and H_2O perfectly covary. What *could* be different (though it in fact is not) is whether any or all of these features are necessary for wateriness or for H_2O. As it turns out, H_2O is the necessary feature for both 'water' and 'H_2O', though it may not have been necessary for either.[84] But for wateriness, the necessary features are only the set of dispositional properties (listed

[83] This story is most clearly articulated in (Chalmers, 2014; Sidelle, 1989), though of course Chalmers is not a conventionalist. He has expressed his views as either dualist or Russellian monist. It should not be surprising someone who thinks that this is the correct story to tell about the term 'water' is between these two positions. What I am at pains to show in the main text is that the main difference between these two positions seems to be that one takes there to be a privileged ontology with respect to the physical and phenomenal while the other does not.

[84] How could H_2O not be a necessary feature of 'H_2O'? The same way that wateriness is not a necessary feature of 'water' but rather the underlying explanatory structure of wateriness is necessary for 'water'. So, it could be (though it in facts seems to not be the case) that 'H_2O' necessarily refers to whatever the deep explanatory structure of H_2O. Of course, if some advanced chemistry is developed, the meaning of the term 'H_2O' may change to be just this (or maybe it will be revealed that this is what we have meant all along).

above) that we associate with water. Nonetheless, the features of the substratum that support wateriness equally support H_2O. Similarly, according to conventionalist neutral monism, the features of the substratum that support pain equally support c-fibers firing (which I'll remind you is just a shorthand for the physical events in the brain and central nervous system that correspond with pain). The fact that we discuss features (plural) rather than a feature (singular) is not due to anything built into the world but rather due to the conventions we use to describe the world.

So, the substratum is neutral with respect to physical and phenomenal properties because such properties are not given by the world. Conventions need to be applied to the substratum in order to give rise to physical and phenomenal properties, but it's natural to wonder how such conventions are applied to the substratum. One might think that if we were to say what features of the substratum can provide such application conditions, we'd have to do so in neutral terms. Otherwise, we would somehow be sneaking the phenomenal and physical properties into the substratum. The idea is that it seems like we can only say what features the substratum has by using phenomenal and physical terms themselves. As such, to say what features the substratum has, we'd have to say something in terms of phenomenal and physical properties. And this would be building into the substratum something that we've said is not there. In other words, what are the features that the substratum has that can provide the application conditions, or the rules for applying terms, for things like 'is in pain' or 'has c-fibers firing'? Do such features have to be given in neutral terms in order for the neutral monism to work? If so, it seems like we were too quick to give up on reductionist neutral monism. If not, then it seems like the substratum does in fact have physical and phenomenal features in it without the application of conventions, otherwise it seems mysterious how we could apply these conventions to it.

However, although this thought seems natural enough, it is misguided and so the line of reasoning that follows from it is faulty. We do not need some neutral language to describe the substratum in order to specify what features are in the substratum. However, even though our description of the substratum must be in terms of phenomenal or physical terms, this does not mean that the substratum itself must have these features. Rather, what we are saying is that we cannot *describe* these features without using the terms we've developed to individuate them. But our being able to individuate some feature via the use of a set of identity conditions does not mean that the identity conditions are somehow contained in the world independent of them. It says more about us than it does the world. *We* cannot describe the world without using semantic

rules that provide features that we can then use to distinguish them from the rest of the world. So, the way to answer this question is quite simple – the substratum contains the features that something is in pain and that c-fibers are firing and this provides the application conditions for the rules for 'is in pain' and 'has c-fibers firing'. Of course, we could try to give a more enlightening answer by using phenomenal and physical terms that we are not already using. For instance, we could say that the substratum having the features that something is experiencing an ache, pinch, or stabbing sensation provides the application conditions of 'is in pain'. These are largely metaphorical ways of describing what a pain feels like, but these are precisely the kinds of phenomenal terms that we should expect of the features of the substratum that will *necessarily* allow for 'is in pain' to be applied.

Similarly, we could say that the substratum having a brain and nervous system with such-and-such a structure undergoing this-and-that stimulus provides the application conditions for 'has c-fibers firing'. Again, though, this merely redescribes the situation in different physical terms, but again this is what will *necessarily* allow for 'has c-fibers firing' to be applied. More generally, we can see from this that the substratum needs to have an experiencing subject in order for any phenomenal terms to be applied correctly. This is not the case for physical terms. For physical terms, the substratum needs to have an objective state of affairs. This might *sound* like we are describing a dualistic ontology, but again the difference between such a dualistic ontology and a conventionalist ontology is in whether the identity conditions of the phenomenal and physical are baked into the substratum itself. On both ontologies, the ability to identify that there is a subjective experience or objective state of affairs comes only if we can distinguish between these two, but on the conventionalist ontology, this comes only from the conventions that we apply. So, in the temporal and causal sense, these two features are already there. Of course, the features have to be there in order for us to apply the terms to them. However, the *distinction* between these features and our ability to recognize two features rather than merely one (or something else altogether) is due to the conventions that we use to identify the phenomenal experience and the physical property. This is similar to the case of Lumpl and Goliath, in the causal and temporal sense, the features that make Lumpl a lump and Goliath a statue are there regardless of whether we apply the conventions to them such that there is a distinction between them. Nonetheless, these are not features of the world that show us that these things are distinct. They are features of the different identity conditions. So, although the subjective and objective features do need to obtain in order to apply the conventions and the conventions are what give us

the identity of the terms, this isn't as circular as it seems. It is just that in order to describe what needs to obtain, we need to use the terms of phenomenal and physical language. This is a methodological point about how we describe which features are needed; it isn't a metaphysical point about what is in the substratum independent of the conventions used to describe it.

If there were some third, neutral way of describing the features of the substratum, then we could more straightforwardly reduce both phenomenal and physical properties to a neutral set of properties that could be described in this neutral language. No such language is available, but even if there were, I suspect that similar problems of whether or not such features are truly fundamental would arise. What we have instead are two languages that correctly refer to the world by providing criteria of identity and criteria of counterfactual application for properties, some of which are physical and some of which are phenomenal. As such, any description we give of the world's properties will make use of these languages.

Remember that the neutral monist is seeking an explanation that can simultaneously explain the perfect covariation we see between physical and phenomenal properties *and* the (at least apparent) difference between them such that we can conceive of them coming apart. Many times our ability to conceive of things that in fact occur simultaneously, but might not have, can be explained by our understanding of causation. However, in the mind-body problem, this isn't the case. Our understanding of the causal structure of the universe shows us that physical and phenomenal properties *perfectly* covary; they never could come apart given the causal structure of the universe as we know it. As such, if the difference is a genuine one, the covariation must be explained non-causally. The conventionalist neutral monist states that this difference is a matter of the conventions we use to explain the world around us, using one set to explain things in terms of our experiences and another set to explain things in terms of their fundamental causal interactions. The substratum supports both phenomenal and physical properties. The same feature(s) of the substratum supports two different carvings, yielding two different property types. The substratum itself does not provide the difference, and this fact explains why there is a perfect covariation with the physical and the phenomenal. But this type of answer naturally leads to a question about whether we should think of this substratum as physical. That is, we must now discuss whether this neutral monist answer is really a type of physicalist answer to the mind-body problem.

5. Not a Materialism nor an Idealism

Physicalists are *right* to say that the physical language can correctly describe everything in the world, but they are wrong to say that there is nothing more to be said. The phenomenal terms may not carve out any part of the substratum missed by physical terms. Nonetheless, just as a world described entirely in terms of lumps of bronze would not leave any (relevant) part of the substratum out doesn't mean that there is no work to be done by describing the world in terms of statues, having all of the substratum describable in physical terms does not mean that phenomenal properties are somehow redundant (as the Causal Argument would prove if a privileged ontology were assumed). Since we are interested in dividing things by the kinds of identity conditions given by the phenomenal language, those properties that are individuated in this way can't simply go away. Moreover, even if they did, that wouldn't change anything about the substratum in our world. Again, the phenomenal and physical languages carve out the same spatio-temporal portion of the substratum. However, this does not mean that the conventions cause any change in it. Rather what they do is give us boundaries with which we can then recognize a relevantly similar portion of stuff in various different situations. Since they carve the substratum in two different ways, they *could* come apart if it were to turn out that they do not, in fact, track the same portion of the substratum. However, all evidence to date points to the fact that the phenomenal and physical perfectly covary in our world and so we should conclude that the phenomenal and physical carve the same portion of the substratum. If this turned out false, then Premise 1 of the Causal Argument presented in Chapter 2 would be false, and it would turn out that the physical sciences have some serious flaws in their ability to track causes in the world. On the other hand, even if it became overwhelmingly obvious that we will never see any evidence of the phenomenal and the physical coming apart in our world, this will not (and cannot) disprove the possibility that they could be different in some world that is not ours *even if* the physical properties of that world were exact duplicates of those in ours.

Take, for instance, the features that we pick out by the term 'water', the watery features of the world. These features are not necessarily connected to the deep explanatory features of wateriness that we find in our world, H_2O. This doesn't mean that water isn't necessarily H_2O. On the contrary, water *is* necessarily H_2O. But this is true because water necessarily is whatever we find in our world that is the deep explanation of the wateriness, and that deep explanatory feature has turned out to be H_2O. But wateriness itself could

very well have been explained by some other deep features. In some other world they could be XYZ. This would be a world where wateriness is explained by XYZ, but because water is necessarily what we have found to explain wateriness *in our world*, there would be no water in an XYZ world. The main difference between 'water' and 'pain', on the conventionalist neutral monist account, is just that 'water' refers necessarily to whatever deep explanatory feature of wateriness is in our world whereas 'pain' refers necessarily only to painfulness and not to whatever deep explanatory feature of painfulness is in our world. So, just as wateriness is not necessarily connected to H_2O, painfulness is not necessarily connected to c-fibers firing even if there could never be any empirical evidence to the contrary. In other words, that which explains wateriness in the actual world may not explain wateriness in every world (we could change all sorts of laws of the universe such that H_2O does not produce wateriness). And that which explains painfulness in the actual world need not explain painfulness in every world. This is what sets conventionalist neutral monism apart from materialism. However, the question remains why the neutral substratum should not be taken to be the same as the physical. After all, the physical does correctly describe all that there is. And if the neutral substratum isn't the same as the physical stuff of the world, then how should we think of this substratum?

This pair of questions will best be answered by simultaneously thinking about why conventionalist neutral monism should also not be thought of as a version of idealism. I don't think this version of neutral monism is under a big threat of being (mis)understood as some type of idealism. However, since so many previous versions of neutral monism (particularly Russellian monism) have been accused of being thinly veiled versions of idealism, it's worth noting some of the differences between the view offered here and why it avoids an accusation that has plagued previous iterations of the view. First, the phenomenal, just as much as the physical, depends on the substratum. The nature of this dependence is difficult to specify because the phenomenal and physical are really two different descriptions of the substratum, but they are descriptions with real ontological importance. So, the temptation is to say that they are *merely* descriptions but what is real, what is fundamental, is just the substratum. However, this is not a correct understanding of the substratum and its relationship to ontological entities (both physical and phenomenal).

To explain this relationship, let's first explore just what this substratum, or neutral stuff, is. As I've stated previously, the substratum lacks principles of individuation. But then it seems like this stuff does

have a principle of individuation in a way. So, what is it that a conventionalist is committed to when they say that the world is just so much stuff? Thomas Blackson gives an illuminating answer:

> Instead of being a commitment to the existence of a certain kind of object that has no principle of individuation, the commitment to stuff seems better understood as a commitment to the possibility of alternative principles of individuation. The English words, e.g., 'infant,' 'toddler,' 'teenager,' etc., can each be properly said at different times of one and the same individual. The change from infant to toddler, in Aristotelian terms, is an alteration in one individual; it is not the generation of a new individual and the corruption of an old individual (Generation and Corruption 1.4). The conventions for 'infant,' 'toddler,' 'teenager,' etc., do not require language users, if they are to use these words to speak truths, to refrain from uttering statements such as 'Max is no longer an infant; he now is a toddler.' Because alternative semantic conventions are possible, principles of individuation that would make this statement necessarily false are also possible. English does not contain such principles of individuation, but, according to conventionalists, the reasons for this fact are pragmatic ones. One should not conclude that from some "correct" viewpoint the world does not contain such substances. All that is out there, to use their metaphor, is stuff. The truth is not constrained by [objects with essential properties] that have principles of individuation that are not given by the semantic conventions of language users. A world of such [objects with essential properties] is a realist world, and conventionalists claim that this world does not exist (1992, p. 76).

So, the substratum is simply the stuff in the world. And that stuff can be carved up by different principles of individuation. For the view on offer here, one set of individuation principles leads to phenomenal properties while the other leads to physical properties. So, we can see that a central difference between conventionalist neutral monism and the other two types of monism is that where materialism and idealism both accept a privileged ontology of the world, conventionalist neutral monism denies this. Again, this is a similarity that the traditional monist views have with dualism. However, in the previous section, we say that the difference between dualism and conventionalist neutral monism is explained by a difference in how the two views understand modality. As we'll see here, this denial of a privileged ontology also has implications for what the views say about the type(s) of stuff that exists in the world.

First, as we've already made clear, the neutral substratum lacks a unique correct set of principles of individuation. On the other hand, the other three standard views all agree that the basic, fundamental stuff of the world has one privileged set of individuation conditions. They disagree on what they are. For the dualist, they are phenomenal and physical. Both sets of individuation conditions are given by the fundamental stuff of the world. For the idealist and the materialist, just one of these is given while the other just comes along for the ride, so to speak. On the other hand, for the conventionalist neutral monist, neither sets of individuation conditions are given by the basic stuff of the world. Individuation conditions for both

phenomenal and physical properties are matters of convention such that when there is stuff in the world that meets them, the phenomenal or physical properties are carved out from the rest of the stuff.

Second, as far as conventionalist neutral monism is concerned, the existence of the substratum does not depend on anything else whereas phenomenal and physical properties depend, in part, on conceptual practices. One might, therefore, think that the substratum is fundamental. However, another way of thinking about what is fundamental is simply what is the basis for our understanding of the world. In this sense, the substratum is not fundamental at all because it "contains" *all* of the potential ways of carving it up that we could think up. The basis for our understanding of the world is the phenomenal and physical properties that we use to describe it (at the most fundamental level, of course). These two natural ways of thinking about what is fundamental come apart from one another when talking about the conventionalist's neutral substratum. But it can also help explain the difference between the neutral substratum and the physicalist's understanding of the fundamental stuff of the world. For the physicalist, the basis for our understanding of the world are the independently existent properties of the physical world. So, physical properties are fundamental in both senses. On the other hand, for the idealist, the basic stuff of the world is decidedly *not* independently existent and so what is fundamental need only be what is the basis for our understanding. The conventionalist's neutral stuff is exists independently but is not a basis for our understanding.

Third, unlike certain kinds of materialism (and, I suppose, certain kinds of idealism), for conventionalist neutral monism, there is no reduction from the physical or phenomenal to any other, more fundamental type of stuff. As I mentioned in Chapter 4, I don't think a reductionist neutral monism can work. Conventionalism is meant to be a way of doing the work that reductionism is supposed to be doing without generating any of the same problems. On the conventionalist picture, there is no need to posit some third, neutral type of property to which the physical and phenomenal can be reduced. If one thinks that multiple realizability is a concern for materialism, then this kind of solution can't work, anyhow. And by accepting a substratum that is neutral with respect to physical and phenomenal properties in the way described above, conventionalist neutral monism avoids such issues concerning multiple realizability.

In this chapter, I've presented a view that I call conventionalist neutral monism. I've attempted to show that this version of epistemic neutral monism avoids the many variations of the central dilemma that other versions of neutral monism fail to avoid. It is both a coherent view and it is substantially different

from other, more standard solutions to the mind-body problem. However, showing that a view is merely coherent and different from others on offer is not showing very much. In the next and final chapter, I'll be at pains to show how conventionalist neutral monism gives us a satisfying solution to the mind-body problem by showing that it successfully meets the demands from Chapter 1 while answering the question that began our inquiry.

Chapter 6: The Answer

> "Unless the person who answered a question were still going on asking it while he formulated the answer, he would have 'lost interest in the subject', and the 'answer' would not have been an answer at all. It would have been a meaningless form of words. By being answered a question does not cease to be a question. It only ceases to be an unanswered question."
>
> - R.G. Collingwood, *An Essay on Metaphysics* (1940, pp. 24–25)

1. Conventionalist Neutral…Monism?

We began this investigation into the mind-body problem with a simple question that is familiar to almost any inquisitive person in modern society. Is the mind the same as the brain? We then revised and refined that question in light of several developments in the philosophical investigation of this question and came to the following question at the center of Anglo-American philosophy's current mind-body problem: Are phenomenal properties – i.e. the properties of experience that can be confirmed and communicated by only one subject – fundamentally the same or different from physical properties – i.e. properties of experience that can be confirmed by multiple subjects and/or from multiple perspectives? This question is easily answered by dualism; they are fundamentally different and, as such, there are two fundamentally different kinds of properties in the world. Materialism and idealism also easily answer the question; these properties are the same and, as such, there is one fundamental kind of property in the world. The problem for these views is not in answering this simple question, but rather in meeting the demands which a *satisfying* answer to this question needs to meet. It would seem that conventionalist neutral monism ought to also have an easy answer to this question. Insofar as it is a monism, one might expect conventionalist neutral monism to say that phenomenal and physical properties are the same, but as we saw in the previous chapter, there is an important way in which conventionalist neutral monism treats phenomenal and physical properties as fundamentally different. This makes a direct answer to our question somewhat tricky. In a way, I take the question to already have been answered in the previous chapter, but I would like to give a more direct answer than that. However, giving a direct answer to this question is how we've gotten the unsatisfactory answers of materialism, dualism, and idealism. As such, I'll answer the question by focusing more explicitly on how conventionalist neutral monism meets the demands for a satisfying answer. After doing this, I will attempt to give a more direct answer to the question in the hopes of clarifying the position further. Then I'll use this answer to re-evaluate the Causal and Epistemic Arguments. And finally, I'll end

the dissertation with some remarks on conventionalism intended to assuage readers who may be disinclined toward any solution to the mind-body problem that requires such a radical view of the ontology of the world.

As it happens, I think conventionalist neutral monism is committed to saying that phenomenal and physical properties are fundamentally the same – after all, it is a *monism*. How they are fundamentally the same is a difficult account to give. However, saying that they are fundamentally the same is something that I think any satisfactory solution to the mind-body problem must say in order to meet the Ontological Demand. As a reminder, this demand states that the world is a unified whole which is something that needs to be reflected in a satisfactory solution to the mind-body problem. The idea behind this demand, as explained in Chapter 1, is that any explanation of something in the world needs to be able to relate it to other things in the world, either by a causal relation or some other relation that allows us to make sense of it. The problems with interactionist dualism show that things that are fundamentally different from one another either have no relation between them at all or else require some laws that link them – laws that aren't forthcoming in cases where the phenomenal and physical are taken to be fundamentally different. As such, any solution to the mind-body problem that denies a fundamental sameness between the phenomenal and physical will simply be stating that, in an important way, questions about how beings with our phenomenal experiences relate to the rest world have no answers. Materialism can give a simple account here. The phenomenal simply is (or is reducible to) the physical. So, the relation is a necessary one of constitution or identity or reduction or supervenience or one of the many other metaphysical relations that have been proposed to further clarify the materialist position. Idealism has a similarly simple account. It is simply the reverse of the materialist one – that is, the physical simply is (or is reducible to) the phenomenal.

Neither of these simple stories are available to conventionalist neutral monism. As we saw in the previous chapter, the conventionalism of the position entails that conventions, in part, determine the ontology of phenomenal and physical properties. Thus, differences in conventions governing the phenomenal and the physical is what makes the difference between them metaphysical. Therefore, the fundamental sameness of the phenomenal and the physical cannot be explained in terms of metaphysical relations. Rather, the Ontological Demand must be met by explaining the sameness in terms that relate the conventions governing our talk of phenomenal properties to the conventions governing our talk of physical properties. The relation, then, that shows us the fundamental sameness of the phenomenal and the physical that we find in the world is metalinguistic rather than metaphysical.

I don't have a name for this metalinguistic relation, so the best I can do is explain it by an analogy. Homer's *Iliad* is a story told to us from the third-person perspective. Joe Goodkin's *Blues of Achilles* is a story told to us from the first-person perspective. Both of these works tell the same story – a story that takes place in the ninth year of the Trojan War centering on the anger of the hero Achilles and the tragic consequences of the Greeks' commitment to honor above all else. We know that these are the same story because the author of the latter tells us this. But even without being told, we can tell that they are the same stories. The events are simply too similar to not be relating the same events from different storytelling perspectives. Of course, there are significant differences. Joe Goodkin's first-person perspective gives us greater insight into many of the female characters as well as greater insight into the emotional outlook of all the characters than we get in Homer's third-person perspective. On the other hand, Homer's version gives us more details about various battles and deaths of minor characters that are often not mentioned in Goodkin's version. We also get a narrative about the war as a whole which is not present from each individual first-person perspective. There may be other ways to tell this story, but none of them are "the correct" telling of the story.

In the same way, phenomenal and physical properties relate to us the same "story" of the world from two different perspectives – the phenomenal gives us a first-person perspective while the physical gives us a third-person perspective. We know that they tell the same story not because there are no differences between the two. Rather, we know this because the perfect covariation between the phenomenal and the physical that we see in the world indicates that they are the same story. It is the story of how certain types of beings relate to, and interact with, the rest of the world including their previous selves. This is, however, not to say that the properties share any sort of metaphysically necessary relation. After all, the differing perspectives yield *different* properties that we recognize as phenomenal and physical. Just as the same story told from two differing perspectives is neither a "translation" nor a one-to-one correspondence nor is one "the correct" story, so phenomenal and physical properties can tell the same story of the world and yet there will not be a "translation" nor a one-to-one correspondence nor some further vocabulary or story that they can be reduced to that might be "the correct" one. The basic fact is that the phenomenal and physical tell us the same story of a certain portion of stuff in the world from two different perspectives, the first-person and the third-person. Of course, the story that is shared between the phenomenal and the

physical is a story of a small portion of the world. Most of the world is unexperienced after all. But then, most of the world doesn't involve brains and central nervous systems either.

Given the inherently social nature of language, it makes sense that these are the two language paradigms that we've landed on. One for discussing our inner lives and experiences which we might, on occasion, decide are important enough that we'd like to share them with others who are relevantly similar to us but for whom access to such individual experiences is cut off. And another for discussing the external goings-on which we might all have access to if we were in the same locations at the same time, but which we may have reasons to refer to so that others might understand what is going on in our location. Indeed, this is in some ways an extension of the first need. In both cases, we have a need to communicate things that others don't have access to. In the first case, it is something that others lack the access to in principle and the second is something that we don't have access to merely as an accident of history. Of course, when we describe events that no one has any personal relation to, using the third-person, physical vocabulary makes the most sense. There is no first-personal experience to convey. But when we discuss our own inner lives, it does seem mistaken to say that what goes for the events for which there is no first-person perspective should go for events that have such an experience. Of course, we *can* describe it from such a third-person perspective, but that does not mean that it is the correct one any more than our description from the first-person perspective is the correct one. Either may be correct given some occasion or purpose for conveying the story from that perspective, but neither tells "the story" in its full complexity any more than the other does.

Talking of phenomenal and physical properties as telling the same story from two different perspectives also illuminates how conventionalist neutral monism answers the Explanatory Demand. This demand states that the appearance of a fundamental difference between phenomenal and physical properties must be explained by any satisfactory solution to the mind-body problem. This appearance of fundamental difference is a mistake in a way, but it is an understandable mistake. Phenomenal and physical properties are metaphysically different. On the standard view that there is some privileged ontology which will be captured by our best metaphysical account of the world, the fact that phenomenal and physical properties bear no necessary metaphysical relation to one another just means that they are fundamentally different. However, this is not the conventionalist view under discussion. A conventionalist view drops the assumption that there is a privileged ontology in the world. As such, when we recognize a metaphysical

difference between two kinds of purportedly fundamental entities – like phenomenal and physical properties – we need not conclude that they are *fundamentally* different despite the fact that they are *metaphysically* different.[85] As discussed in the previous chapter, fundamentality for a conventionalist is separated into two important notions. So, if we take fundamentality to be "that which is exists mind-independently", then this metaphysical difference which is derived from a difference in conventions will not be a fundamental difference. This is the way in which the appearance of fundamental difference is merely an appearance. On the other hand, if we take fundamentality to be "that which is the basis for our understanding of the world", then the metaphysical difference between the phenomenal and physical is a fundamental difference and the appearance is respected. Either way, conventionalist neutral monism meets the Explanatory Demand by showing how this apparent fundamental difference is in a way genuine just so long as we understand what a fundamental difference entails.

So, back to the question: are phenomenal properties fundamentally the same or different from physical properties? The conventionalist neutral monist answer is that they are the same…kind of. But also different…kind of. They are the same in the way that two tellings of the same story are fundamentally the same. This is the monism of the view. But they are different in that telling the same story in two different ways often means that the two different ways the story is told share almost nothing in common. Moreover, there is no reason to suppose that either phenomenal or physical ways of story-telling are telling the story "correctly". This is the neutrality of the view. Finally, there is no possible story that could be telling the story "correctly". This is the conventionalism of the view. So, where the more standard answers to this question are straightforward and simple, the conventionalist neutral monist answer is neither straightforward nor simple. However, where those standard answers fail to meet the demands of a satisfying answer, conventionalist neutral monism is capable of meeting those demands.

2. Re-examining the Epistemic and Causal Arguments

In Chapter 2, I presented the impasse in current Anglo-American philosophical debates which is created by (a) dualists defending various forms of the Epistemic Argument that concludes that materialism

[85] Some conventionalists would say that these differences aren't metaphysical because there isn't a difference in the substratum, or stuff, in the world. This may track our normal usage more closely, if (Einheuser, 2006) is correct about how we often think about metaphysical differences. However, in the conventionalist picture under discussion, these distinctions between phenomenal and physical properties still concern their natures even though the differences are due to conventions. Thus, I prefer the usage of 'metaphysical' that treats differences due to conventions as genuine metaphysical differences even when there is no difference in substratum.

is false and (b) materialists defending the Causal Argument that concludes that dualism is false. When these are the only two possibilities, we are left wondering what to do. At this point, I think I've done enough to show that there is at least one more possible solution to the mind-body problem. In fact, it is one that can more easily meet the demands of a satisfactory solution than any previously given standard solution. Here I'd like to detail how conventionalist neutral monism can accept both of these arguments and their conclusion with ease while some other non-standard solutions to the mind-body problem are left struggling to adequately escape the arguments.

Because of the way that conventionalist neutral monism meets the Ontological Demand, it can freely accept that the physical is causally closed and that there are no further causal properties than the physical ones. Furthermore, conventionalist neutral monism agrees with materialism that causation isn't the only way in which two events or properties might be related. The difference is in whether there is a reason to suppose that the relation is one of ontological dependence. Where materialism insists that it must be, the conventionalism of the view on offer allows that there is no metaphysically necessary relation at all between the phenomenal and physical. The story of the world told to us by physics is simply the most complete and fundamental way of telling the *causal* story of the world. But there is no reason to suppose that there isn't some other *non-causal* way of telling the same story. This is not a competing understanding of the world but rather one that complements our causal understanding by filling in the first-personal perspective that would otherwise be only hinted at (or completely ignored) by the third-personal, causal understanding given by the brain sciences.

The problem that the various forms of dualism have with this argument is that they all state that the difference between the properties is one that is found in the world and thus the question of whether there are causal relations between the two types of properties end up being very important. And if there are no causal relations between them, then the phenomenal properties seem somehow redundant. At least, it seems perfectly legitimate to wonder what it is that they are doing in the world. However, if conventionalism is right, then the difference between phenomenal and physical properties is not *in the world* but rather *in our understanding* of it. Thus, we needn't try to understand what it is that the phenomenal properties are doing in the world. Rather, we need to understand what it is that phenomenal properties do for us in just the same way that we need to understand what physical properties do for us. Physical properties provide us with the best causal understanding of the world that we have come up with to date while phenomenal

properties provide us with the best experiential understanding of the same world. This isn't to say that the phenomenal can never figure into a good causal explanation, but rather, especially at the fundamental level, phenomenal properties will most often not play a role in our most detailed causal explanation of what is going on.

Similarly, conventionalist neutral monism can also freely accept as sound the Epistemic Argument without any of materialism's problems with the argument. Of course, the argument is specifically designed to prove that materialism is false, so materialism cannot simply accept the argument as sound – some premise or other must be false. The problem with this is that materialism ends up leaving the Explanatory Demand unmet. On the other hand, accepting the Epistemic Argument as sound is a perfectly fine way to meet the Explanatory Demand. However, the dualist interpretation of the argument makes it so that dualism is unable to fulfill the Ontological Demand. The way that dualism interprets the Epistemic Argument is to say that there is a genuine difference in the world between the phenomenal and physical properties, a difference given to us by the way the world is. This gives us an explanation for why they appear to be different kinds of properties. However, it makes it highly difficult to see how they fit together into the same world. Conventionalist neutral monism interprets the Epistemic Argument not as a demonstration of ontological differences given to us by the world but rather as ontological differences that result from our conventions. So, although there is an agreement that what is conceivable is possible, as mentioned in the previous chapter, there is a disagreement about why this is the case. Again, it is the conventionalism that makes the difference here. It is what allows conventionalist neutral monism to accept that conceivability implies possibility while still meeting the Ontological Demand. The phenomenal and physical properties are different, in part, because of the difference in conventions governing what makes some property physical or phenomenal. Meanwhile, dualism accepts that the difference in their possibly coming apart is what makes it true that we can conceive of them coming apart. This is simply the difference between insisting on a privileged ontology or accepting, as the conventionalist does, that there may be no such privileged "correct" ontology of the world.

It is here that I'll state what I think should be abundantly clear by now. The main difference between the standard solutions to the mind-body problem and the only kind of neutral monism that I think offers a genuine alternative is a denial of the privileged ontology view of properties (at least, with respect to phenomenal and physical properties). The acceptance of a non-privileged ontology is what separates

conventionalist neutral monism from other solutions to the mind-body problem that are "non-standard" as well. Daniel Stoljar has been a major proponent of what he calls "non-standard materialism". And while his view is certainly non-standard in the sense that there are major departures from the traditional views that are canvassed in an undergraduate introduction to the mind-body problem, it fits very much in the standard category of materialist views that (a) accept a privileged ontology, (b) fail to avoid materialism's problems with the Epistemic Argument, and (c) thereby fail to meet the Explanatory Demand.

In his "Panpsychism and Non-standard Materialism: Some Comparative Remarks" (2020), Stoljar argues convincingly that panpsychism, in various forms, falls victim to something very similar to the conceivability argument leveled against materialism (this is one variation of the more general Epistemic Argument). What he fails to realize is that similar reasoning to that which he uses against panpsychism applies to his favored non-standard materialism. Central to Stoljar's argument against panpsychism is the idea that the phenomenal[86] properties that the view is committed to are in an important sense "non-standard". He clarifies this notion by stating that they are non-standard in two ways. First, the bearers of these phenomenal properties are not the standard bearers of phenomenal properties. Standard bearers of such properties are biological lifeforms that are of a certain level of complexity. To ignore some difficult debates about non-human consciousness, we can simply say that the standard bearer of phenomenal properties is an adult human being. But panpsychism (in its most common form) is the view that all fundamental objects have phenomenal properties. Adult human beings are quite unlike fundamental objects like quarks and electrons. So, these are certainly non-standard bearers of phenomenal properties. The second way in which these properties are non-standard, and by far the more important way, is that these phenomenal properties of fundamental objects are *unimaginable* by us. This is plausible given that fundamental objects are so vastly and obviously different from us that whatever phenomenal properties that they may have, we are not in a position to imagine what they are.

This unimaginability of the phenomenal properties of fundamental objects is supposed to offer the panpsychist a reply to the Combination Problem.[87] Stoljar's presentation is meant to draw out the similarities between this problem and the conceivability argument against materialism. The first premise is

[86] Stoljar uses the term 'psychological' here, but it is clear that he means the same thing as I mean by 'phenomenal'. I've substituted phenomenal to stay consistent with the rest of the dissertation.

[87] This name was coined by (Seager, 1995).

that it is conceivable that the physical properties and the non-standard phenomenal properties could be exactly the same as they are in the actual world and yet the phenomenal properties of standard objects (like us humans) could be very different. That is, the combination of non-standard phenomenal properties and physical properties can conceivably come apart from standard phenomenal properties. The second premise is simply that since such a state of affairs is conceivable, it is possible. Therefore, there is no necessary connection between the phenomenal properties of fundamental objects and the phenomenal properties of humans. As such, these non-standard phenomenal properties fail to explain why we have any phenomenal properties at all. Thus, panpsychism falls to the Combination Problem in the same way that standard materialism falls to the conceivability argument. However, Stoljar shows that panpsychism has a response to this argument. If these non-standard phenomenal properties are unimaginable, then we are not able to conceive of the relevant situation wherein the phenomenal properties of fundamental objects are exactly as they are now but the standard phenomenal properties are different. A property that we cannot even imagine is not a property that we can conceive of in the relevant way, and so the first premise is false.

The problem with this reply, Stoljar argues, is that even though we might not be able to imagine what these properties are, we certainly know at least *some* essential property that they have. He mentions a few candidates, but the one which seems most promising is that "[phenomenal] properties essentially consist in (or at least partially consist in) awareness of properties" (2020, p. 223). If we know this about the phenomenal properties of fundamental objects, then we know enough to conclude that no amount of, or arrangement of, such properties could ever rule out the possibility that the standard phenomenal properties that we experience might not exist. In other words, we know enough about non-standard phenomenal properties to conceive of the relevant situation. So, premise one is true after all. The natural response for the panpsychist is to say that we actually do not even know this about these phenomenal properties. In fact, we know nothing at all about them! This, Stoljar is quick to point out, is not a helpful line of response. He says, "if we literally know nothing about non-standard [phenomenal] properties, it is hard to see why they should be called '[phenomenal] properties' in the first place. And if they are not [phenomenal] properties, then this 'panpsychist' position stands revealed as no different to non-standard materialism" (2020, p. 224).

Importantly, Stoljar claims that the same thing cannot be said about his favored non-standard materialism. He argues that where the non-standard phenomenal properties do not help panpsychism avoid it's version of a conceivability argument, the non-standard physical properties, specifically the fact

that they are unimaginable, allow non-standard materialism to sidestep the conceivability argument. They are able to do this because, again, what is unimaginable cannot be conceived (in the relevant sense). Herein lies the (supposed) difference. Where we know an essential feature of a phenomenal property, Stoljar claims that we have a much more "open-ended" understanding of physical properties. He states that "according to our contemporary understanding of a physical property, a physical property (very roughly) is either a property distinctive of ordinary physical objects or else is a property that explains the properties distinctive of ordinary physical objects" (2020, p. 224). He seems to think that this second clause, in particular, is what opens the door for us having no way of knowing what kinds of physical properties some future completed physics will uncover[88], and thus that we have no idea what essential properties belong to physical properties. As such, we cannot imagine these non-standard physical properties even in outline. And so, we cannot conceive of various situations wherein these properties are instantiated while phenomenal properties are not. Thus, Stoljar argues that such non-standard physical properties allow non-standard materialism to avoid the conclusion of the Epistemic Argument, that materialism is false.

However, he seems to miss the mark here. Although he is right that we do not know what kind of property will be needed to explain the properties distinctive of ordinary physical objects, he is wrong that we therefore know nothing about the essential features of such properties. For the same reason that we must know *something* about a non-standard phenomenal property in order to count it as a phenomenal, rather than physical, property, we must know something about these non-standard physical properties that will make it such that we will include them in our expanded supply of physical properties. In particular, I think it's fairly obvious that the *type* of essential feature of these properties must be something along the lines of "being intersubjectively confirmable" and/or "being causally efficacious" and/or "being necessarily related to the causally efficacious properties" and *not* "consisting of an individual awareness of properties". The list of physical properties is not as open-ended as Stoljar makes it out to be. When he says that we cannot imagine non-standard physical properties, he is thinking of first-order properties that these objects

[88] This makes a direct reference to Stoljar's preferred non-standard materialism, "Nagelian" materialism, which he states as the non-Russellian way of making a non-standard materialism. Nagelian materialism states that the non-standard physical properties are those discovered by some future, completed physics. Russellian physicalism, as one would imagine, states that the non-standard physical properties are the categorical basis for standard physical dispositional properties. Rusellian materialism was one of the views mentioned in Chapter 3; it is a fine view, but fails to differentiate itself from materialism in the relevant way of failing to meet the Explanatory Demand. Here, although I put it in more Nagelian terms, we see the more detailed reason why this view fails to meet the demand; it falls to the Conceivability Argument.

have. This is very open-ended. Nevertheless, the *type* of properties that count as physical is *not* that open for discussion. We have a very clear understanding of the essential feature of physical properties in the same way that we have for phenomenal properties. Stoljar states that "the notion of 'conceivability' that is in operation in these arguments is epistemically demanding: you cannot conceive a situation in which various properties are instantiated unless you know, at least, in outline what those properties are" (2020, p. 225). But the relevant difference between some property consisting in awareness and some property consisting in causal efficacy is enough of an outline to know what those properties are in the relevant sense. The issue of the conceivability argument is that causal efficacy is simply not enough to produce awareness. A non-standard materialist would need to reply that their non-standard physical properties do not essentially consist (wholly or partly) in being causally efficacious. But besides stretching the notion of "physical property" beyond all recognition, this would then run straight into the standard materialist's Causal Argument (appropriately modified) and starts looking startlingly like epiphenomenalist dualism.

The result of all this is that, despite Stoljar's best efforts to prove the contrary, non-standard materialism is quite standard in its lack of a satisfactory reply to the Epistemic Argument. My diagnosis of this issue is that Stoljar's materialism accepts a privileged ontology. In his own words, "what we need are unknown elements, the non-standard materialist says, not non-standard arrangements of known ones [as the panpsychist says we need]" (2020, p. 227). The assumption that there is one, and only one, way to completely and fundamentally specify the way that the world is leads to the impasse that makes the mind-body problem so intractable, both for standard and "non-standard" solutions. Letting go of that assumption allows us to see how to accept both the Causal Argument and the Epistemic Argument without running afoul of the other. Conventionalist neutral monism explains the soundness of these arguments such that they present one coherent worldview rather than being opposed to one another. So, what we've needed all along, the conventionalist says, is neither unknown elements, nor non-standard arrangements of known elements but rather a better understanding of our known elements and their known arrangements.

3. Concluding Remarks

To conclude the dissertation, I'd like to say a few words to convince you that conventionalism isn't as ridiculous as it is sometimes made out to be. I'll begin by recounting a criticism that Philip Goff levels at the dual carving strategy:

> If our world admits of Dual Carving, no amount of reasoning can reveal to us that the numerous ways of carving up the world "hang together" (i.e. are just different ways of understanding the same reality). An omnipotent and infinitely rational being could bring into existence a wholly pure physical reality and be surprised that that same reality can also be described in experiential terms….if Dual Carving is true, there are unintelligible necessary connections between distinct carvings. (2017, p. 128)

I begin with this criticism of the dual carving strategy because it certainly sounds like something one might bring against conventionalism, but I've been at pains to distance conventionalism from this view in at least one very important way. The main point of departure should be made clear from the discussion of strong necessities in the last chapter. The dual carving strategy is committed to strong necessities where conventionalism is not. And the commitment to strong necessities just is a commitment to the "unintelligible necessary connections" for which Goff is (rightly) ridiculing the dual carving strategy.

Nevertheless, one might be suspicious that despite this, the crux of the criticism is that "no amount of reasoning can reveal to us that the numerous ways of carving up the world 'hang together'", and *this* applies equally well to conventionalism as it does the dual carving strategy. After all, one of the main ways I've been describing the conventionalist break from the standard metaphysical view is that, on a conventionalist picture of the world, there is no privileged ontological description of what is in the world at a fundamental level, whereas on the standard view there is one such privileged description. Moreover, given the denial of a strong necessity between them, conventionalism accepts that these different, non-privileged descriptions of fundamental reality do not *necessarily* "hang together" so we might wonder how they are supposed to hang together at all. However, because of the source of these carvings on the conventionalist view of ontology, the fact that no amount of reasoning can prove that they hang together is not really much of a problem. These different carvings of the world come from our conventions governing what is essential for some property to be phenomenal or physical. So, an omnipotent and perfectly rational being would not be surprised at all to find an experiential description of a 'wholly pure physical reality', assuming such a reality contains some states that are aware of other states from a first-person perspective. Such a description, though not necessary for the perfectly rational being interested only in the physical state of that world, would nevertheless not be surprising. That our reasoning cannot connect the phenomenal and physical is no mark against the idea that they *might be* connected. Our reasoning cannot connect them simply because they are not *necessarily* connected. All one needs to do to see that they are connected is to look and notice the unmistakable fact that the phenomenal and physical perfectly covary in our world. It is

the *difference* between them that comes from our use of different carvings; their (contingently) hanging together comes from the stuff in the world.

All of this has been to say, in short, that conventionalism (about the phenomenal and physical) is not obviously false, and so we cannot abandon conventionalist neutral monism on those grounds. But one might press on and argue that by committing ourselves to conventionalism in order to solve the mind-body problem, we've given up on the whole project of metaphysics. If the project of metaphysics is restricted to the project of giving the privileged description of reality, then this is right. However, this is not the only area of metaphysics that such a constraint on the discipline has been called into question. Questioning the constraint of a privileged ontology has been made prominent by nominalism (and later conventionalism) about essence, conventionalism about standard physical objects, pragmatism (about it all), late Wittgenstein, and these are just the list of areas of interest to current Anglo-American philosophers. Others of significance include Nietzsche, Hegel, and Madhyamika Buddhism. If giving up on finding the privileged ontology of the world is giving up on metaphysics, it is less like giving up on a game in the first half of play and more like Russell and Whitehead giving up on providing the world with a complete logical system after Gödel proved it to be impossible. Nonetheless, I don't think that accepting conventionalism should convince us that trying to solve the mind-body problem (or other metaphysical problems) should be viewed as a waste of time. Many of those who argue for some version of conventionalism (including many engaged in debates on the above figures and topics) agree with Richard Rorty when he says that the battle between the privileged view of ontology and the non-privileged view "is about whether the traditional problems of modern philosophy are to be taken seriously or set aside" (2016, p. 37), where if you are on the side of the non-privileged view of ontology, you think that the traditional problems should be set aside. I disagree. At the end of the day, if conventionalism gives us a good metaphysical understanding of the world, then it can help us understand how to solve the mind-body problem. This is a problem that does not simply go away once we accept conventionalism, rather it moves from an unsolvable problem to a problem with a satisfactory solution. So, instead of setting the problem aside, we ought to explore the details of this solution in order to understand what conventionalism does for us.

References

Alter, T. A., & Nagasawa, Y. (Eds.). (2015). *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.

Armstrong, D. M. (1996). *A World of States of Affairs* (Issue 3). Cambridge University Press.

Averill, E. W., & Keating, B. (1981). Does interactionism violate a law of classical physics? *Mind*, *90*(January), 102–107.

Ayer, A. J. (1971). *Russell and Moore: The Analytical Heritage*. Harvard University Press.

Banks, E. C. (2014). *The Realistic Empiricism of Mach, James, and Russell: Neutral Monism Reconceived*. Cambridge University Press.

Bennett, K. (2004). Spatio-temporal coincidence and the grounding problem. *Philosophical Studies*, *118*(3), 339–371.

Bergson, H. (1907). L'évolution créatrice. *Revue de Métaphysique et de Morale*, *15*(5), 620–670.

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford University Press.

Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, *108*(1), 1–46.

Broad, C. D. (1925). *The Mind and its Place in Nature*. Routledge.

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

Chalmers, D. J. (2002). Consciousness and Its Place in Nature. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 247–272). Oxford University Press.

Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford University Press.

Chalmers, D. J. (2013). Panpsychism and Panprotopsychism. *Amherst Lecture in Philosophy*, *8*.

Chalmers, D. J. (2014). Strong necessities and the mind–body problem: A reply. *Philosophical Studies*, *167*(3), 785–800.

Chalmers, D. J. (2016). The Combination Problem for Panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism*. Oxford University Press.

Churchland, P. M. (1981). Eliminative Materialism and Propositional Attitudes. *Journal of Philosophy*, *78*(2), 67–90.

Churchland, P. M., & Churchland, P. S. (1981). Functionalism, qualia and intentionality. *Philosophical Topics*, *12*(1), 121–145.

Churchland, P. S. (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.

Coates, A. (2020). Making sense of powerful qualities. *Synthese*, *198*(9), 8347–8363.

Collingwood, R. G. (1940). *An Essay on Metaphysics*. Oxford University Press UK.

Cucu, A. C. (2020). Does Consciousness-Collapse Quantum Mechanics Facilitate Dualistic Mental Causation? *Journal of Cognitive Science*, *21*(3), 429–473.

Cucu, A. C., & Pitts, J. B. (2019). How Dualists Should (Not) Respond to the Objection from Energy Conservation. *Mind and Matter*, *17*(1), 95–121.

Cutter, B. (2019). Against the Middle Ground: Why Russellian Monism is Unstable. *Analytic Philosophy*, *60*(2), 109–129.

Demircioglu, E. (2013). Physicalism and Phenomenal Concepts. *Philosophical Studies*, *165*(1), 257–277.

Dennett, D. C. (1978). *Brainstorms*. MIT Press.

Dennett, D. C. (1988). Quining qualia. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science*. Oxford University Press.

Einheuser, I. (2006). Counterconventional Conditionals. *Philosophical Studies*, *127*(3), 459–482. https://doi.org/10.1007/s11098-004-7790-5

Einstein, A. (1998). *The collected papers of Albert Einstein: The Berlin Years: Correspondence, 1914-1918* (R. Schulmann, A. J. Kox, M. Janssen, & J. Illy, Eds.; Vol. 8). Princeton Univ. Pr.

Ellis, B. (2001). *Scientific Essentialism* (Issue 450). Cambridge University Press.

Ellis, B. (2002). *The Philosophy of Nature: A Guide to the New Essentialism*. Routledge.

Ellis, B. (2005). Universals, the essential problem and categorical properties. *Ratio*, *18*(4), 462–472.

Farris, J. R., & Göcke, B. P. (2021). *The Routledge Handbook of Idealism and Immaterialism*. Routledge.

Feigl, H. (1958). The "mental" and the "physical." *Minnesota Studies in the Philosophy of Science*, *2*, 370–497.

Feigl, H. (1967). *The "Mental" and the "Physical" the Essay and a Postscript* (Issue 2, pp. 214–215). University of Minnesota Press.

Feyerabend, P. K. (1963). Mental Events and the Brain. *Journal of Philosophy*, *60*(11), 295–296.

Foster, J. (1993). The Succinct Case for Idealism. In H. Robinson (Ed.), *Objections to Physicalism* (pp. 293–313). Clarendon Press.

Foster, J. A. (1982). *The Case for Idealism* (Issue 3, pp. 465–468). Routledge.

Gibbard, A. (1975). Contingent Identity. *Journal of Philosophical Logic*, *4*(2), 187–221. JSTOR.

Goff, P. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy*, *89*(2), 191–209.

Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford University Press.

Goff, P., & Papineau, D. (2014). What's wrong with strong necessities. *Philosophical Studies*, *167*(3), 749–762.

Griffin, D. R. (1998). *Unsnarling the World Knot: Consciousness, Freedom, and the Mind-Body Problem*. (Issue 3, pp. 353–367). University of California Press.

Heil, J. (2003). *From an Ontological Point of View*. Oxford University Press.

Heil, J. (2010). Powerful qualities. In A. Marmodoro (Ed.), *The Metaphysics of Powers: Their Grounding and Their Manifestations*. Routledge.

Heil, J. (2020). *Philosophy of mind: A contemporary introduction* (Fourth edition). Routledge, Taylor & Francis Group.

Hiddleston, E. (2019). Dispositional and categorical properties, and Russellian Monism. *Philosophical Studies*, *176*(1), 65–92.

Hildebrand, T. (2014). Can bare dispositions explain categorical regularities? *Philosophical Studies*, *167*(3), 569–584.

Holman, E. L. (2013). Phenomenal concepts as bare recognitional concepts: Harder to debunk than you thought, …but still possible. *Philosophical Studies*, *164*(3), 807–827.

Horgan, T. E., & Tienson, J. L. (2001). Deconstructing new wave materialism. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents* (pp. 307--318). Cambridge University Press.

Ingthorsson, R. (2013). Properties: Qualities, Powers, or Both? *Dialectica*, *67*(1), 55–80.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, *32*(April), 127–136.

Jacobs, J. D. (2011). Powerful Qualities, Not Pure Powers. *The Monist*, *94*(1), 81–102.

James, W. (1879). Are we automata? *Mind*, *4*(13), 1–22.

James, W. (1890). *The Principles of Psychology* (Issue n/a, p. 284). Dover Publications.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind/Body Problem and Mental Causation*. MIT Press.

Kind, A. (2015). Pessimism About Russellian Monism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism* (pp. 401–421).

Kirk, R. (1974). Sentience and behaviour. *Mind*, *83*(January), 43–60.

Kirk, R., & Squires, J. E. R. (1974). Zombies v. Materialists. *Aristotelian Society Supplementary Volume*, *48*(1), 135–164.

Kripke, S. A. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium* (Issue 217, pp. 431–433). Harvard University Press.

Levin, J. (2007). Nagel vs. Nagel on the nature of phenomenal concepts. *Ratio*, *20*(3), 293–307.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(October), 354–361.

Lewis, D. (1986). Postscripts to `causation'. In D. Lewis (Ed.), *Philosophical Papers Vol. II*. Oxford University Press.

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, *4*, 81–108.

Loar, B. (1997). Phenomenal states II. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press.

Lockwood, M. (1981). What was Russell's neutral monism? *Midwest Studes in Philosophy*, *6*(1), 143–158.

Lockwood, M. (1989). *Mind, Brain and the Quantum: The Compound "I"* (Issue 396, pp. 650–652). Oxford University Press.

Lockwood, M. (1993). The grain problem. In H. M. Robinson (Ed.), *Objections to Physicalism* (pp. 271–291). Oxford University Press.

Lycan, W. G. (2009). Giving dualism its due. *Australasian Journal of Philosophy*, *87*(4), 551–563.

Martin, C. B. (1993). The Need for Ontology: Some Choices: C. B. Martin. *Philosophy*, *68*(266), 505–522.

Martin, C. B., & Heil, J. (1999). The ontological turn. *Midwest Studies in Philosophy*, *23*(1), 34–60.

Maxwell, G. (1979). Rigid designators and mind-brain identity. *Minnesota Studies in the Philosophy of Science*, *9*.

McKitrick, J. (2003). The bare metaphysical possibility of bare dispositions. *Philosophy and Phenomenological Research*, *66*(2), 349–369.

McLaughlin, B. P. (2001). In defense of new wave materialism: A response to Horgan and Tienson. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents*. Cambridge University Press.

Mellor, D. H. (1974). In defense of dispositions. *Philosophical Review*, *83*(2), 157–181.

Mills, E. O. (1996). Interactionism and overdetermination. *American Philosophical Quarterly*, *33*(1), 105–115.

Molnar, G. (1999). Are dispositions reducible? *Philosophical Quarterly*, *49*(194), 1–17.

Molnar, G. (2003). *Powers: A Study in Metaphysics*. Oxford University Press.

Mumford, S. (1998). *Dispositions* (Issue 2). Clarendon Press.

Mumford, S. (2006). The Ungrounded Argument. *Synthese*, *149*(3), 471–489.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation* (Issue 1, p. 618). Harcourt, Brace & World.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*(October), 435–450.

Okon, E., & Sebastián, M. Á. (2020). A consciousness-based quantum objective collapse model. *Synthese*, *197*(9), 3947–3967.

O'Leary-Hawthorne, J., & Cortens, A. (1995). Towards ontological nihilism. *Philosophical Studies*, *79*(2), 143–165.

Papineau, D. (1993). *Philosophical Naturalism*. *Philosophical Naturalism* (Issue 4, pp. 1070–1077). Blackwell.

Pautz, A. (2015). *A Dilemma for Russellian Monists About Consciousness*. Metaphysics at the Ranch, City University of New York and University of Cambridge.

Pelczar, M. (2019). Defending Phenomenalism. *Philosophical Quarterly*, *69*(276), 574–597.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*(1), 3–21.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness* (Issue 1, pp. 457–461). MIT Press.

Pincock, C. (2005). History of Philosophical Analysis [review of Scott Soames, _Philosophical Analysis in the Twentieth Century_]. *Russell: The Journal of Bertrand Russell Studies*, *25*(2), 167–171.

Pitts, J. B. (2019). Conservation Laws and the Philosophy of Mind: Opening the Black Box, Finding a Mirror. *Philosophia*, *48*(2), 673–707.

Polger, T. W. (2004). *Natural Minds*. Bradford.

Prior, E. W. (1985). *Dispositions* (Issue 146, p. 109). Humanities Press.

Quine, W. V. O. (1960). *Word and Object*. MIT Press.

Ramm, B. (2021). Panpsychism and the First-Person Perspective: The Case for Panpsychist Idealism. *Mind and Matter*, *19*(1), 75–106.

Rey, G. (1982). A reason for doubting the existence of consciousness. In R. J. Davidson, G. E. Schwartz, & D. H. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–39). Plenum.

Rey, G. (1986). A question about consciousness. In H. R. Otto & J. A. Tuedio (Eds.), *Perspectives on Mind*. Kluwer Academic Publishers.

Rorty, R. (1965). Mind-body identity, privacy, and categories. *Review of Metaphysics*, *19*(1), 24–54.

Rorty, R. (2016). *Philosophy as Poetry*. University of Virginia Press.

Russell, B. (1921). *The Analysis of Mind*. Allen & Unwin.

Russell, B. (1927). *The Analysis of Matter* (Issue 4, pp. 382–385). Kegan Paul.

Seager, W. (1995). Consciousness, information, and panpsychism. *Journal of Consciousness Studies*, *2*(3), 272–288.

Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Cornell University Press.

Sidelle, A. (1992). Rigidity, ontology, and semantic structure. *Journal of Philosophy*, *89*(8), 410–430.

Sidelle, A. (1995). A semantic account of rigidity. *Philosophical Studies*, *80*(1), 69–105.

Sidelle, A. (1998). A sweater unraveled: Following one thread of thought for avoiding coincident entities. *Noûs*, *32*(4), 423–448.

Sider, T. (2011). *Writing the Book of the World*. Oxford University Press.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, *68*(April), 141–156.

Soames, S. (2003). *Philosophical Analysis in the Twentieth Century, Volume 1: The Dawn of Analysis*. Princeton University Press.

Stace, W. T. (1944). Russell's Neutral Monism. In P. A. Schilpp (Ed.), *The Philosophy of Bertrand Russell* (Issue 22). Harper & Row.

Stewart, D. (1833). *Elements of the Philosophy of the Human Mind* (Vol. 1). James Munroe and Company.

Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press.

Stoljar, D. (2001). Two Conceptions of the Physical. *Philosophical and Phenomenological Research*, *62*(2), 253–281.

Stoljar, D. (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. Oxford University Press USA.

Stoljar, D. (2014). Four Kinds of Russellian Monism. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind*. Routledge.

Stoljar, D. (2015). Russellian Monism or Nagelian Monism? In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*.

Stoljar, D. (2020). Panpsychism and Non-standard Materialism: Some Comparative Remarks. In W. Seager (Ed.), *The Routledge Handbook of Panpsychism*.

Strawson, G. (1994). *Mental Reality* (Issue 184, p. 414). MIT Press.

Strawson, G. (2003). Real Materialism 2003. *Chomsky and His Critics Ed. Anthony and Hornstein*. https://www.academia.edu/397808/Real_Materialism_2003

Stubenberg, L. (2016). *Neutral Monism > Reducing Mind and Matter to Neutral Entities (Stanford Encyclopedia of Philosophy)*. https://plato.stanford.edu/entries/neutral-monism/reducing-to-neutral.html

Stubenberg, L. (2018). Neutral Monism. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/

Taylor, H. (2018). Powerful qualities and pure powers. *Philosophical Studies*, *175*(6), 1423–1440.

Taylor, J. H. (2013). Physicalism and Phenomenal Concepts: Bringing Ontology and Philosophy of Mind Together. *Philosophia*, *41*(4), 1283–1297.

Thomas A. Blackson. (1992). The Stuff of Conventionalism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *68*(1), 65–81.

Tse, C. Y. P. (2022). Metaphysical idealism revisited. *Philosophy Compass*, *17*, 1–21.

Wilkes, K. V. (1988). Yishi, duh, um and consciousness. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science*. Oxford University Press.

Wilkes, K. V. (1995). Losing consciousness. In T. Metzinger (Ed.), *Conscious Experience*. Ferdinand Schoningh/Imprint Academic.

Wishon, D. (2015). Russell on Russellian Monism. In T. A. Yujin Nagasawa (Ed.), *Consciousness in the Physical World: Perspectives on Russellian Monism* (pp. 91–118).

Woodward, P. (2018). A Posteriori Physicalism and the Discrimination of Properties. *Acta Analytica*, *33*(1), 121–143.